

Abstract
Resonances in Nonintegrable Open Systems
Jens Uwe Nöckel
1997

Resonances arising in elastic scattering or emission problems are investigated as a probe of the Kolmogorov-Arnol'd-Moser (KAM) transition to chaos and its wave manifestations. The breaking of symmetries that leads to this transition affects all the intrinsic properties of a resonance, which suggests applications where these properties can be controlled and predicted in parameter ranges beyond the reach of perturbation theory.

Convex dielectric optical microcavities are studied which support long-lived “whispering-gallery” (WG) modes that classically correspond to rays trapped by total internal reflection in orbits close to the interface with the outside lower-index medium. These resonators with substantial but always convex deformation are termed asymmetric resonant cavities (ARCs). The connection between individual resonances and ray ensembles in an asymmetric billiard is established via a novel application of Einstein-Brillouin-Keller (EBK) quantization, based on the adiabatic approximation of Robnik and Berry which describes WG trajectories even when the deformation exceeds the threshold at which Lazutkin’s caustics cease to exist in the relevant regions of phase space. At such strong distortions, resonance lifetimes are determined not by the wavelength as in symmetric cavities, but by the classical diffusion time from the EBK initial condition in phase space to an escape window corresponding to classical violation of the total internal reflection condition.

Instead of the isotropic emission from rotationally invariant objects, highly asymmetric resonators exhibit strongly peaked intensity in directions which can be predicted from the phase space structure near the classical escape window. This creates unambiguous fingerprints of the KAM transition in the emission anisotropy of ARCs, which are universal for all classically chaotic WG modes. Ray calculations are compared to numerical wave solutions as well as to experiments, and good agreement is found especially for the directionality. Ray predictions for the lifetimes fail when wave mechanical corrections such as chaos-assisted tunneling and dynamical localization are important.

Resonances in Nonintegrable Open Systems

A Dissertation
Presented to the Faculty of the Graduate School of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Jens Uwe Nöckel

Dissertation Director: Professor A. Douglas Stone

May 1997

Acknowledgements

It would require a historical account of my thesis progress to do justice to all persons whose ideas, know-how and comments had an impact on the work presented here. During the four years I spent at Yale, Professor Stone allowed me to pursue research on a variety of topics, not all of which found a place in this thesis. Knowing that my own ideas were welcomed and appreciated, the experience of entering uncharted territory offered great satisfaction. My advisor's ability to seek out such new fields was very beneficial to me. In particular my work on Asymmetric Resonant Cavities was initiated by Doug Stone in close involvement with Richard Chang and his experimental group.

It was very valuable to receive experimental feedback and results from Gang Chen. In solving the problem of generating numerical solutions of the wave equation for open dielectrics, I had crucial discussions with Henrik Bruus, who applied his in-depth experience with wave solutions for the closed Robnik billiard to the problem at hand, making contributions that are still waiting to be tied in to the work being done at present. His involvement goes far beyond the specific physics questions, however, because he was always willing and (more importantly) able to help with the hang-ups that a beginning graduate student invariably faces; in particular he taught me much about the workings of the computing equipment. On this subject, I also received help from Shanhui Xiong and Subir Sachdev.

The concepts at the heart of this thesis were introduced to me at Yale in lectures held by Dan Lathrop, Martin Gutzwiller and Dima Shepelyansky. Professor Shepelyansky was very involved in the effort to characterize the billiard map by analyzing the effective kick strength and exploiting similarities to the standard map. He also proposed a first strategy for obtaining wave results, which Henrik Bruus employed to obtain accurate data on the real parts of the WG resonance pole positions. Professor Gutzwiller pointed out connections to other phenomena that have not yet been fully explored, like surface waves and ray splitting.

Important progress was made after discussions with Marko Robnik about the adiabatic approximation to Lazutkin's invariant curves. On the experimental optics side, I also enjoyed discussions with A. Serpengüzel, S. Arnold and R. Slusher.

Several newer members of Professor Stone's group were and are working actively on ARCs. The study of classical escape from billiards was conducted in great detail by Attila Mekis. Ying Wu made important comments that led me to clarify the discussion of escape directionality in the thesis. Gregor Hackenbroich brought his experience in semiclassical quantization, scattering theory and the supersymmetry method to bear on the problem of resonance widths in optical resonators, and in particular pointed out to me the relevance of chaos-assisted tunneling. An important test of my own understanding has always been a discussion with Evgueni Narimanov, whom I trust to find any potential flaw in almost any argument.

In the course of my earlier thesis research, I had very motivating discussions early on with Mark Reed, Bob Wheeler, Markus Büttiker, Heinrich Heyszenau and Bernhard Kramer. With respect to scattering theory and its history in nuclear physics, I profited from discussions with Robert Adair and Frank Firk.

Finally, it was a pleasure to have been sharing offices, terminals and thoughts with my fellow graduate students in Becton Center: Harsh Mathur, Serdar Orgut, Umesh Waghmare, Milica Milovanovic, Alex Elliot, Mandar Desphande, Kedar Damle, Senthil Todadri, Ilya Gruzberg, Mehmet Gokcedag, Mohiuddin Mazumder, Anurag Mittal, Kenneth Segall and Stephan Friedrich.

This work was supported by U. S. Army Research Office Grant DAAH04-94-G-0031 and NSF Grant DMR-9215065.

Contents

1	Introduction	1
2	Synopsis of the classical ideas	9
2.1	Transition to chaos in closed billiards	9
2.2	Poincaré section and classical ray escape	12
3	Hamiltonian systems	15
3.1	The flow in phase space	16
3.2	The first Poincaré invariant	17
3.3	Transformations with a generating function	19
3.3.1	Lagrange bracket	19
3.3.1.1	Theorem	20
3.3.2	Canonical transformations	21
3.3.2.1	Theorem	21
3.3.3	Eigenvalues and determinant of symplectic matrices	22
3.4	Integrable systems	22
3.5	Action-angle variables	23
3.5.1	Path-independence on the torus	24
3.5.2	The action function	26
3.6	The Hamilton-Jacobi differential equation	27
3.6.1	Winding numbers	28
3.7	Example: Classical mechanics of the circular billiard	28
3.7.1	Polar coordinates	28
3.7.2	Actions from the caustic	32
3.8	Example: Elliptical billiard	34
3.8.1	Integrability of the ellipse	34
3.8.2	An expression for the angle of incidence	36
3.8.3	Actions from the curvature	37
3.8.4	How exceptional is the ellipse billiard ?	39
3.9	The Poincaré map	40
3.9.1	Area-preservation	41
4	Billiard maps and the transition to chaos	43
4.1	Birkhoff coordinates	43
4.2	SOS of the circle and the ellipse	44
4.3	Fixed points and tangent map	46
4.4	Symmetries of the billiard map	48
4.5	Poincaré-Birkhoff fixed point theorem	48
4.6	Stable and unstable manifolds and their intersections	50
4.7	Chaotic motion	51

4.8	The KAM theorem	52
4.9	Resonance overlap and global chaos	53
4.9.1	The most irrational numbers	53
4.9.2	The Chirikov standard map	54
4.10	Lazutkin's theorem	55
5	Effective map for planar convex billiards	58
5.1	Non-analyticity of the kick strength	58
5.2	Geometric derivation of the momentum mapping equation	60
5.3	The position mapping equation	64
6	Phase space structure and transport theory	67
6.1	Some model deformations	67
6.1.1	Quadrupole	68
6.1.2	Robnik billiard	68
6.1.3	Ellipse	69
6.2	Poincaré sections of the model billiards	70
6.3	Local nonlinearity and its relation to Lazutkin's theorem	72
6.4	Diffusion in the SOS	75
6.5	Adiabatic invariants from the effective map	77
6.6	Transport in the presence of phase space structure	80
6.6.1	Cantori	80
6.6.2	Stable and unstable manifolds as barriers	81
6.6.3	Stable islands	83
7	Semiclassical quantization	84
7.1	Einstein-Brillouin-Keller quantization on fuzzy caustics	84
7.2	Quantization conditions	87
7.3	Reliability considerations	90
8	The wave equation for symmetric resonators	91
8.1	Electromagnetic waves in the circular cylinder	91
8.1.1	Metastable well in the effective potential	92
8.1.2	Boundary conditions	94
8.1.3	Comparison to quantum mechanics	94
8.1.4	Total internal reflection	95
8.2	Quasibound states	95
8.3	Above-barrier resonances	97
8.4	Below-barrier resonances	99
8.5	Resonance widths from WKB approximation	102
8.6	Global accuracy of the approximations	103
9	Wave equation for asymmetric resonant cavities (ARCs)	106
9.1	Wave solutions for deformed cylinders	106
9.1.1	Rayleigh Hypothesis	107
9.1.2	Wavefunction Matching	108
9.1.3	Solving the matching equations	109
9.1.4	Quasibound states at complex k	110
9.1.5	Symmetry considerations	111
9.1.6	Incrementing the deformation from the circle	112
9.2	Emission directionality of quasibound states	113

10 Ray-optic model for resonances	116
10.1 Semiclassical quantization for the resonance position	116
10.1.1 Applicability of EBK quantization	116
10.1.2 Dynamical localization	119
10.2 Quasibound states and ray dynamics	120
11 Comparison of ray model and exact solutions	122
11.1 Universal widths at large deformation, localization correction	122
11.2 Chaos-assisted tunneling	123
11.2.1 The annular billiard	124
11.2.2 Chaos-assisted tunneling in open systems	124
11.2.3 Tunneling without chaos in the ellipse	125
11.3 Emission directionality	127
11.3.1 Ergodic model	127
11.3.2 Diffusion between adiabatic curves	129
11.3.3 Dynamical eclipsing	131
11.3.4 Universal directionality	132
11.4 Directional lasing emission from dye jets	137
11.4.1 Multimode lasing	137
11.4.2 Observations on lasing jets	138
12 Thresholds and Intensity Distribution in Lasing Droplets	140
12.1 Experimental configuration	140
12.2 The centrifugal billiard	140
12.2.1 Justification for a classical treatment	140
12.2.2 Classical ray dynamics	142
12.3 The prolate-oblate difference	145
13 Conclusion	147
A Einstein-Brillouin-Keller quantization in closed systems	150
A.1 WKB approximation for eigenstates on a torus	150
A.2 Semiclassical quantization in the circle	153
A.3 Eikonal theory	154
B Periodic-orbit theory for the density of states	156
B.1 Berry-Tabor formula for integrable systems	157
B.2 Gutzwiller trace formula for chaotic systems	159

List of Figures

1.1	Integrable shapes and non-integrable deformed counterparts.	3
1.2	Ray trajectories for circle and quadrupole.	4
2.1	Poincaré surface of section of a quadrupole at eccentricity $e = 0.55$ with corresponding real-space images.	10
2.2	Left: A single trajectory in the quadrupole SOS, encircling islands. Right: The same trajectory in real space.	11
2.3	Phase-space flow at intermediate number of reflection, escape condition.	13
3.1	(a) A typical quasiperiodic trajectory in the circular billiard. (b) Five-bounce periodic orbits.	29
3.2	The integration loops for obtaining the azimuthal (a) and radial (b) action in the circle.	33
3.3	The normal bisects the angle between incoming and outgoing ray in the ellipse.	35
3.4	The two types of trajectories in the ellipse are whispering-gallery orbits (left) and bouncing-ball orbits (right).	36
3.5	Integration contours for the radial and azimuthal action in the ellipse.	38
4.1	Poincaré SOS of the circle.	45
4.2	Surface of section for the ellipse at eccentricity $e = 0.4$	45
4.3	Origin of hyperbolic and elliptic points in the Poincaré-Birkhoff fixed point theorem.	50
4.4	Stable and unstable manifolds emanating from hyperbolic fixed points.	51
4.5	Poincaré surfaces of section for quadrupolar deformations.	55
5.1	Kick strength in the momentum mapping equation, as a function of the initial $\sin \chi$ at eccentricity $e = 0.56$ in the quadrupole. The solid line is obtained from the true map by averaging the squared kick amplitude over all final ϕ and taking the square root. The circles are a fit with $(1 - \sin^2 \chi)^{3/2}$	59
5.2	A ray intersects the convex boundary at arc lengths s_1 and s_2	60
6.1	Poincaré sections for the quadrupole billiard. Chaos is seen to spread beginning from the low- $\sin \chi$ region where the stable and unstable two-bounce diametral orbits dominate the phase portrait. The next largest islands belong to the four-bounce orbit, and they survive to large deformation. The whispering gallery orbits near $\sin \chi = 1$ are largely unaffected by the deformation.	70
6.2	Poincaré sections for the dipole. Surviving WG orbits and their break-up at large e are shown on an expanded scale.	71

6.3	Poincaré sections for the ellipse billiard.	71
6.4	Poincaré sections for the effective map of the quadrupole.	72
6.5	Poincaré sections for the effective map of the dipole.	72
6.6	Poincaré sections of the quadrupole, showing 10 iterates of a cloud of trajectories launched close to each unstable fixed point of winding number 4.	82
7.1	Trajectory following the adiabatic curve for 200 reflections. Bottom: same trajectory followed for 2000 bounces.	84
7.2	Two possible counterclockwise trajectories through an arbitrary point in the annulus between caustic (C) and boundary (B).	86
7.3	The two fields defined by the CB and BC rays, respectively.	87
7.4	The integration path for the radial action.	88
8.1	Effective potential picture for whispering gallery resonances; $k_{min} = m/(nR)$ and $k_{max} = m/R$	93
8.2	Intensity scattered off a circular cylinder at 50° with respect to the direction of a plane wave.	96
8.3	Resonance positions in the circle of refractive index $n = 2$ with TM polarization for the lowest 44 angular momentum numbers m	97
8.4	Exact resonance widths calculated from Eq.(8.27) for selected resonances whose real parts fall into the interval $44 < kR < 45$ at refractive index $n = 2$, compared to Eq. (8.88).	104
8.5	Reflection probability p_0 at each collision with a circular interface in comparison to the Fresnel formula for a plane interface.	105
9.1	Wavenumber of an eigenstate in the closed quadratic Robnik billiard as a function of deformation, from an exact diagonalization and from the wavefunction matching approach.	113
9.2	False color representation of the squared electric field in the TM mode for the $m = 68$, $kR = 45.15$ resonance of the quadrupole-deformed cylinder at eccentricity $e = 0.66$. The intensity is higher for redder colors, and vanishes in the dark blue regions. High intensity regions in the near-field occur just outside the surface at the highest curvature points $\phi = 0, \pi$, and high emission intensity lines (green) emanate from these points in the tangent directions. The high intensity inside the cavity reflects the quasibound nature of the state, which is seen to be still WG-like with no intensity in the center.	114
10.1	Shift in real part of kR , denoted here by ΔkR , versus deformation for a resonance with $kR = 28.6$, $m = 46$ (in the circle), under an elliptical deformation. The refractive index is $n = 2$. The green curve shows the semiclassical result.	117
10.2	Shift in real part of kR versus deformation for three resonances of the quadrupole with refractive index $n = 2$	118
10.3	The starting condition for the ray escape simulation.	120
11.1	Logarithmic plot of imaginary part of resonance positions vs. deformation for quadrupole with refractive index $n = 2$	123

11.2	Logarithmic plot of the imaginary part (γR) of the resonance position kR as a function of deformation for the resonance $kR = 12.1$, $m = 20$ at refractive index $n = 2$ in the ellipse (circles), quadrupole (stars) and dipole (crosses). The arrows indicate the classical threshold deformation for the onset of diffusion from starting to escape condition, in the dipole (D) and quadrupole (Q).	125
11.3	Classical escape directionality starting from $\tilde{p} \approx 0.85$ and $e = 0.58$	127
11.4	Illustration of the areas A and \bar{A} which end up below or come from above $\sin \chi_c$, respectively, in one mapping step.	128
11.5	Far-field intensity distribution for the quadrupole (a,b) and the ellipse (c) at eccentricity $e = 0.6$, for a resonance with $m = 68$, showing wave results and the pseudoclassical predictions.	130
11.6	Far-field intensity distribution for the quadrupole at the same parameters as in Fig. 11.5 (a,b), neglecting above- barrier reflection and tunneling.	131
11.7	Emission directionality in the far-field of the quadrupole at eccentricity $e = 0.65$ for 5 different resonances with various kR and $\sin \chi$ (numbers given for the circle). The refractive index is $n = 2$	133
11.8	Far-field directionality for 5 different resonances of the quadrupole at eccentricity $e = 0.65$ and refractive index $n = 1.54$, displaying the peak splitting due to dynamical eclipsing.	134
11.9	Far-field directionality in the quadrupole with increasing eccentricity e at $n = 2$ for the resonance with $m = 45$, $kR = 27.8$	135
11.10	Far-field directionality in the quadrupole with increasing eccentricity e at $n = 1.54$ for the same resonance as in Fig. 11.5.	136
11.11	Total lasing intensity images of vertically flowing jets of Rhodamine B dye in ethanol. The jet with deformed cross section (left) is observed perpendicular and parallel to the slit from which the liquid originates, giving rise to the two intensity traces. On the righthand side is shown the same observation for a liquid column of circular cross section.	139
12.1	Shadow graphs (left) and simultaneous total-energy images (right) of three lasing droplets.	141
12.2	Poincaré surfaces of section for prolate droplets with $\epsilon = 0.2$ for $L_z = 0.735$ (a), $L_z = 0.6$ (b), $L_z = 0.45$ (c) and $L_z = 0$ (d).	143
12.3	Prolate (a) and oblate (b) droplet SOS's at $\epsilon = 0.2$ and $L_z/\rho_b(0) = 0.3$	145
A.1	A bundle of trajectories evolving in time.	151

Chapter 1

Introduction

This thesis attempts a synthesis of ideas from two active research fields: The study of dielectric optical microcavities, and the theory of dynamical systems whose classical phase space is partially chaotic. In the process, the theory has been extended to capture novel phenomena which are relevant to applications in optics as well as other fields, like the physics of solid state microstructures. The pivotal objects of the present work are long-lived resonant states whose properties can be predicted and tailored with great flexibility, thus making them amenable to application development.

As a historical example for such an application, we mention the Fabry - Pérot interferometer of 1899, which has proven to be one of the most valuable tools in optical spectroscopy. The resonant tunneling diode is an example for electronic applications of the same ideas which have become feasible since semiconductor fabrication techniques make the wave nature of the electron accessible.

Generally, resonances are long-lived quasi-bound states in an open system that arise due to interference, and they give rise to sharp variation in scattering phase shifts, cross sections, transmission coefficients, etc., as the incident wavelength is varied. An open system is characterized by the existence of propagating waves at large distance from the region where the quasi-bound states are formed. The long *lifetime* of photons in a laser resonator is what makes it possible to obtain coherent stimulated emission. The *sharply peaked* wavelength-dependent transmission of a Fabry - Pérot interferometer is the basis of high resolution spectroscopy. The characteristic wavelengths λ at which resonances occur, as well as their lifetimes τ , are device-specific. In optics, one uses the Q factor as a figure of merit for the resonator, where $Q \equiv \omega\tau$, $\omega \equiv 2\pi c/\lambda$.

The intrinsic properties of a resonant state, like ω and τ , can usually be determined straightforwardly when the wave equation is separable in some coordinate system due to the existence of conservation laws. When there exist as many conservation laws as there are degrees of freedom, a system is called integrable. The complexity of the problem is greatly increased in *non-integrable* systems, where it becomes impossible to reduce the wave equation to a collection of separate first-order differential equations. In optical device applications, one therefore prefers to work with symmetric geometries where (approximate) rotational or translational invariance provides the desired simplification of the wave equation.

This luxury is not always available to physicists studying electronic transport in microstructures, simply due to the nature of the lithographic techniques used in manufacturing these devices.¹ While it has become possible to fabricate structures in which electrons move ballistically, i.e. are scattered only by the boundaries of

the device and not by impurities or phonons, the shape of the boundaries generally leads to non-integrability even in the simplest non-interacting electron description. Transport experiments on open ballistic microstructures^{2,3} have, beginning with the theoretical work of Jalabert, Baranger and Stone in 1990,⁴ become both a catalyst and a testing ground for important new developments in the field of *quantum chaos*. For an overview, see for example the chapters by Stone and by Smilansky in Ref.⁵. Quantum chaos is the study of nonseparable Schrödinger equations based on a knowledge of the underlying classical mechanics, which can be chaotic when the system is non-integrable, as will be explained below.

Meanwhile in optics, a theory of nonintegrable resonators only existed in the form of perturbation approaches^{6,7} where the breaking of symmetries was treated in the limit where it causes only a small correction to the symmetric solutions. Technological development did not seem to require a more general theory, although there existed experimental evidence for novel phenomena that occur in strongly asymmetric resonator geometries beyond the reach of perturbation theory^{8,9}. When such situations are encountered, it is in principle always possible to solve the wave equation numerically, albeit with significant effort due to computational difficulties that arise especially for quasibound states (as opposed to off-resonant scattering), and are compounded when wavelengths are short compared to the cavity dimensions.

It has recently been realized¹⁰ that nonperturbative effects may in fact be useful in device applications, and one therefore desires models that could make predictions and provide explanations for phenomena observed in strongly non-separable wave equations. It is at this juncture that we propose to build a bridge to the approach that was successful in ballistic microstructures, the main building blocks being the similarity between optical and quantum mechanical wave equations and the resulting classical limits, i.e. mechanics and ray optics. We shall in this way arrive at phenomena and questions that do not arise in microstructures.

This work grew out of earlier studies concerning aspects of ballistic electronic transport in non-integrable microstructures where progress can be made using general methods of quantum scattering theory. The author first became acquainted with non-integrable systems during this earlier part of the thesis research, which will however not be included here. For details on this microstructure work, the reader is referred to Refs.¹¹⁻¹⁵ Although the open systems studied there are non-integrable, we were able to derive statements about the limits between which the conductance (or transmission in the language of scattering theory) can vary when resonances are encountered, and about the possible lineshapes. In this context we turned also to some model systems which have as their distinguishing feature the fact that their resonance lifetimes can be tuned externally by applying a magnetic field in various ways. In fact, these systems can in principle support infinitely long-lived states in the continuum, due to the existence of selection rules that prevent decay. The magnetic field is introduced as a symmetry-breaking perturbation that destroys these selection rules. What was learned from these systems is that the loss of the selection rule does not cause the long-lived state to vanish abruptly, but instead to smoothly decrease in lifetime. External *in situ* control over resonance lifetimes is desirable from the point of view of potential applications, and this has led us to investigate related mechanisms in optics. It should, however, be stressed that many of the ideas put forward here could prove useful in microstructure and even nuclear physics as well. This is why the more general title was chosen for this thesis.

The prototypical optical system that spurred our interest is the *microcavity laser*, which has been realized experimentally both in liquid droplets with a lasing

dye,¹⁶ and as a disk containing an InGaAsP-InGaAs quantum well laser¹⁷. Typical diameters are $10\mu\text{m}$ for the disks, and $30\mu\text{m}$ for the droplets. These novel geometries also point the way to other optical devices^{18,19}, as well as to novel probes of non-linear optical and cavity-quantum-electrodynamical effects²⁰⁻²². In particular the disks which can be pumped electrically²³ could enter into competition with the currently widespread vertical-cavity surface-emitting lasers, due to the fact that microcavities provide a confinement of the modes in all spatial dimensions, leading to an enhancement η of the spontaneous emission coefficient of the lasing medium in the cavity as compared to its vacuum value. One finds^{17,24} that the figure of merit β , which measures how much of the total spontaneous emission goes into each cavity mode, is given by $\beta = (1 + \eta^{-1})^{-1}$, so that β can be increased toward its maximum, $\beta = 1$, if η is made large. Now the enhancement factor η is given by the ratio between the frequency spacing of the resonances and their width²⁵. Therefore, one desires a large mode spacing, and consequently a small quantization volume, in conjunction with a small resonance width, in order to increase β . The latter then lowers the pump threshold for lasing.

Conventionally, one uses Bragg reflectors to provide Fabry-Pérot type mode confinement, but this does not lead to quantization of all degrees of freedom, and one faces limitations in the feasibility of fabricating small devices. The reason is that the bragg reflectors then become large in relation to the actual cavity. To achieve sufficient reflectivity, these mirrors have to consist of at least 40 quarter-wave layers¹⁷. The microcavity lasers mentioned above, on the other hand, do not require bragg reflectors at all. They make use of modes propagating inside a dielectric close to the interface with the air outside. These modes correspond to rays traveling around the perimeter, confined to the dielectric by *total internal reflection* at the interface. If the interface can be made clean and smooth, the only leakage out of such a cavity stems from the fact that the surface has a finite curvature so that total internal reflection is violated, allowing a small fraction of the internal intensity to escape. This mechanism is closely related to quantum mechanical tunneling, and the escape rates are correspondingly small. The most recent record for the highest quality confinement was achieved in fused-silica microspheres of $500 - 1000\mu\text{m}$ diameter, where $Q \approx 0.8 \times 10^{10}$ was measured at optical wavelength²⁷. These high- Q modes

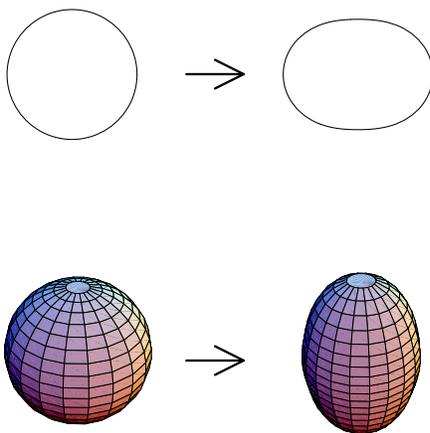


Figure 1.1: Integrable shapes (left) in two and three dimensions and their non-integrable deformed counterparts.

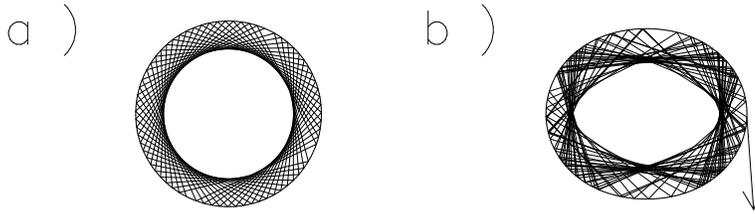


Figure 1.2: Ray trajectories for circle (a), and quadrupole-deformed circle (b) parametrized by $r(\phi) = 1 + \epsilon \cos 2\phi$ in polar coordinates for $\epsilon = 0.08$ corresponding to an 8% fractional deformation. Rays are launched from the boundary at the same ϕ and angle of incidence $\sin \chi_0 = 0.7$ in both cases; ray escape by refraction occurs in case (b).

are also referred to as whispering-gallery (WG) modes, after Lord Rayleigh, who explained that analogous acoustic modes cause whispers to propagate unattenuated along the walls of St. Paul’s Cathedral while being inaudible in the center of the hall²⁸. The WG modes of a dielectric sphere or disk are confined by total internal reflection because the angle of incidence χ between the rays and the normal to the boundary is conserved at each reflection. This conservation law implies that a ray which satisfies the total internal reflection condition,

$$\sin \chi > \frac{1}{n}, \quad (1.1)$$

n being the refractive index of the cavity, will never be able to violate it.

The question we want to address is, what happens to such a WG orbit when the rotational symmetry that is the reason for the conservation of $\sin \chi$ is broken, as sketched in Fig. 1.1. This purely classical problem was rigorously solved for rays moving in a plane bounded by perfectly reflecting walls only in 1973 by Lazutkin²⁹. The difficulty is how to guarantee that $\sin \chi$ will not fall below $1/n$ after some number of reflections, given that there is no symmetry any more forcing χ to be constant. The situation is described in Fig. 1.2. Chaos theory already lurks behind this seemingly innocuous problem, if we for the moment define chaos to be associated with irregular trajectories of the type shown in Fig. 1.2(b). One might argue that there is no need to build strongly asymmetric cavities beyond the reach of wave-mechanical perturbation theory. However, symmetric microcavities have shortcomings that are precisely due to their high symmetry. One of these is their isotropic emission caused by the absence of a preferred direction. To couple the lasing emission into an adjacent optical component like a fiber, one would wish for a strongly peaked intensity in as small a solid angle as possible. Disk lasers that achieve a certain degree of directionality have been realized using gratings of varying spacing around the disk edge, or a tab-like extension on one side of the disk³⁰.

Based on Lazutkin’s theorem, we will see that simpler types of deformation lead to potentially stronger emission anisotropy while preserving a higher Q -factor if desired. The theorem asserts that for sufficiently smooth convex deformations of the circle, the situation in Fig. 1.2(b) does *not* occur for certain families of WG trajectories while it may occur for others. This separation of the ray dynamics into two distinct classes is a consequence of the Kolmogorov-Arnol’d-Moser (KAM) theorem of Hamiltonian classical mechanics (to be discussed later), applied here to ray optics. It implies, among other things, that a shape perturbation does not abruptly cause all trajectories to become chaotic. Consequently, we will in this thesis study the particular class of resonators characterized by a (not necessarily small) deformation

which however preserves convexity everywhere along the boundary. We call this class **asymmetric resonant cavities** (ARCs). Asymmetric resonant cavities hold great promise as experimental systems in which to test more detailed predictions of KAM theory.

Despite the mathematical complexity of these underlying theorems, our basic ideas and predictions are very tangible and pertinent to the applied problems of controlling frequency, width and directionality of quasibound states. The results that are obtained for whispering gallery modes in simple convex but strongly asymmetric resonant cavities can be summarized as follows:

Red shift: The resonance frequency always shifts to lower values with increasing deformation when constant area is maintained. This can be explained using an adiabatic approximation based on the proximity to the boundary and hence to Lazutkin’s caustics.

Broadening: The resonance lifetime, τ , always decreases with deformation. For each resonance, there is a classical threshold deformation beyond which its lifetime is dominated by classical ray escape as opposed to tunneling (i.e. the small violation of total internal reflection present even in the circle). At such large deformations, τ becomes independent of frequency provided ω is large enough. The *universal* resonance broadening depends only on the index of refraction and the angle of incidence characterizing the whispering gallery orbits (in a sense to be defined below).

Directionality: Emission from a quasibound state is highly anisotropic at strong deformations, with intensity peaks in directions that are determined to high accuracy by the phase space structure of the classical ray dynamics inside the cavity. At deformations high enough for classical escape to dominate over tunneling, the directionality is furthermore *universal* for all whispering gallery resonances, and the only parameter that affects it is the refractive index.

To arrive at these conclusions, we have to establish a link between the problem of solving the wave equation for the resonator on one hand, and the classical ray optics picture to which we can apply the theory of nonlinear dynamics. After all, resonances are manifestly a wave phenomenon.

Contact between waves and rays is made using semiclassical approximations. Together with statistical approaches originally devised in nuclear physics and the study of random systems^{31,32}, these methods are at the heart of quantum chaos, and they have been applied successfully to open systems^{33–35}. A comparison between the stochastic and semiclassical methods can be found in Ref.³⁶. Both in open and closed systems, semiclassical methods can provide information even on the level of individual states. For example, in the scattering off three hard disks in the plane one can express the complex resonance positions as well as the differential cross sections in terms of a sum over periodic orbits which exist in infinite number despite the openness of the system³⁷. This is an application of Gutzwiller’s trace formula,³⁸ one of the central semiclassical approaches to locating poles of the Green function. The trace formula can even become exact, as it does for the scattering phase shift of a particle entering a “box” in a space of constant negative curvature through an attached horn extending to infinity³³. However, in many other applications this approach faces severe convergence problems³⁹, which can sometimes be overcome by truncating the problematic series on the grounds that a physical cutoff is provided

by some dephasing mechanism⁴⁰. These arguments, however, cannot be made in our optical systems, and consequently the Gutzwiller approach is not used here.

Instead, a more direct version of the well-known WKB (or eikonal quantization) method is employed, in spite of the fact that the latter requires integrability to be rigorously valid. This approach consists in quantizing the actions associated with the conserved quantities. In the presence of chaos, this is in general not possible, but for WG modes in convex billiards the chaotic diffusion of the relevant action is found to be sufficiently slow, so that reasonable results are obtained. The reason for this is the slow departure of the ray trajectory from the annular region near the boundary to which it is initially confined, cf. Fig. 1.2(b) where one observes that the trajectory does not explore the whole billiard area but looks reminiscent of the path in the circle. This phase-space transport property is characteristic of convex billiards where, according to Lazutkin’s theorem, chaos and regular motion coexist, but it has not been explicitly utilized in previous quantization approaches for billiards. The quantization of systems with such mixed phase spaces is currently the subject of intense study, because the only cases that are generally well understood are the extremes of fully integrable and completely chaotic dynamics. A method like ours which, unlike the period-orbit approach, incorporates phase space transport effects *a priori*, may prove advantageous for more general applications than considered in this thesis.

Whereas the semiclassical part of our method is based on the fact that even chaotic WG trajectories do not show large variation in χ over many reflections, it is precisely this slow change in χ that is responsible for the possibility of classical escape. Therefore, our calculation of the resonance lifetimes must take into account the associated phase space diffusion. This is done by extracting an initial condition for the diffusion process from the result of the semiclassical approximation (which neglects diffusion), and then determining the diffusion time as the average time needed to escape.

In calculating the positions and widths of WG resonances, we thus apply novel methods to the determination of familiar quantities. In fact, the spectrum (which in our case is complex) has been the focus of by far the most quantum-chaos studies³⁹. Only more recently has the wavefunction itself attracted increased interest,⁴¹ in particular after the discovery of wavefunctions “scarred” by classical periodic orbits⁴². In the present work, properties of the wave function, and their relation to classical phase space, play a crucial role. Quantitative comparison between our predictions for the resonance widths based on classical diffusion and exact numerical solutions of the wave equation gives insight into the size of the neglected quantum effects. There are two such effects that we observe to be important.

The first is *dynamical localization*, which has been studied first in periodically driven systems^{43,44}. An experimental realization exists in the form of the Hydrogen atom in a microwave field of nonperturbative intensity^{45,46}. There, the observed ionization thresholds as a function of the driving field strength can be predicted on a classical basis, with the ionization mechanism being a classical diffusion process in phase space. Quantitative discrepancies to the experimental thresholds arise because the chaotic diffusion occurring in that system is suppressed when one makes the transition to quantum mechanics. This is in close analogy to what we find for the escape rates of the leaky billiard. The problem of dynamical localization in billiards has in itself attracted appreciable attention recently^{47–50}, and optical resonators have been recognized as actual applications of the theoretical developments in this field⁵⁰. The consequence of dynamical localization for the WG wave function is, loosely

speaking, that it contains fewer low-angular momentum (hence low-sin χ) admixtures than would be expected from the chaotic spreading of the classical trajectory.

A second quantum effect is *chaos-assisted tunneling*, which was introduced even more recently as an explanation for anomalously large tunnel splittings between non-chaotic states in a mixed system, mediated by the mere presence of chaos in classically inaccessible parts of phase space⁵¹. Whereas such enhanced tunnel splittings are still exponentially small and thus experimentally almost impossible to access, we propose a similar mechanism for the observed *enhancement of resonance widths* in dielectric resonators at small but non-perturbative deformations. The numerical data suggest that resonances can be found for which the widths in question should be experimentally resolvable. The manifestation of chaos-assisted tunneling in wave functions has been demonstrated using the Husimi distribution and a modification thereof,^{54–56} and one finds qualitatively that the chaotic enhancement results from the fact that the tunneling distance between two points in phase space can be significantly reduced if part of the way is navigable by means of classically allowed diffusion.

The relationship between quantum decay and properties of the wave function has been previously investigated for strongly chaotic systems and with an emphasis on statistical statements about collections of resonances^{52,53}. It should be stressed that we are interested in the decay of modes in a KAM (i.e. only partially chaotic) system where narrow resonances exist not by virtue of physical tunnel barriers but due to the phase space structure itself (i.e. local conservation laws). Furthermore, we will be able to make statements about *individual* resonances in an open system, as well as to identify universal properties.

Motivated by the microcavity laser applications, we are concerned with a pure emission problem in the absence of an incoming wave. The predictions for resonance positions and widths do not depend upon this assumption, but our theoretical approach mirrors this quasibound-state point of view. A quantity that does depend on the choice of the incoming wave (or its presence in the first place) is the partial scattering cross section. In the emission problem we assume the photon has been created inside the cavity in one of its quasibound states, and therefore the resulting directionality of escape is entirely an intrinsic property of that state. There is no free parameter such as the incident wavevector to vary, in contrast to the partial cross section. This type of emission directionality from highly deformed leaky billiards will be of fundamental importance in various applications, but it has not been investigated before. One reason for this may be that in microstructure physics, which has provided much of the motivation for recent theoretical efforts, leads are attached to the sample, thus prescribing where the decay has to occur. Clearly the escape directions are not prescribed in this way in a free-falling microdroplet, and the observed emission anisotropy is all the more interesting.

The necessary concepts from classical mechanics and semiclassical theory are expounded in some detail in chapters 3 to 6. These chapters contain novel elements as well as known results, and each builds on the previous one to form a logical presentation of the methods. However, the central ideas and results can be understood at a lower level of detail, and for this purpose an introductory primer is provided in chapter 2. This synopsis of the subsequent chapters 3 to 6 will enable the reader to proceed directly to chapter 7 and the following, where it is shown in detail how the classical and semiclassical concepts are applied to open resonators. To better appreciate the results presented in the later chapters, the reader may eventually want to return to the earlier chapters for further details on

classical Hamiltonian dynamics, in particular on the consequences of integrability versus non-integrability for phase space transport, and for a discussion of the specific model systems that are used repeatedly.

The outline of the classical-mechanics chapters is as follows: After making the transition from continuous-time dynamical systems to the discrete mappings induced by them, the appearance of chaos is described with the help of the Poincaré surface of section. This is a tool which proves valuable to our understanding of global phase space structure, which in turn is crucial to the predictions listed above. The ray escape from an asymmetric cavity is a transport process in phase space, so we devote some discussion to the phenomena that occur there. An important result is that diffusion in the variable whose value is critical for the possibility to escape (namely the angle of incidence) can be sufficiently slow to allow a semiclassical treatment in the spirit of that used for integrable systems.

This is explained after the general ideas of the semiclassical approach have been introduced. Before the semiclassical and ray dynamics theories are combined to make contact with the wave properties of asymmetric resonant cavities, we describe the numerical techniques developed to solve the quasi-bound emission problem exactly. The final chapters contain the results of our theory and comparisons with numerics as well as experiments.

Chapter 2

Synopsis of the classical ideas

As mentioned in the introductory chapter 1, we want to make predictions about frequency shifts, broadening and directional emission from whispering gallery resonances in dielectric cavities. This requires a connection between wave and ray optics using a semiclassical approximation, and further an application of results from KAM theory to the ray optics. In chapters 3 to 7, these ideas are developed for closed systems, and their relevance to the open resonant cavities is explored afterwards.

In this synopsis we explain the importance of classical phase space structure both for understanding the transition to chaos and for predicting the intrinsic properties of quasibound states in dielectric resonators. The reader should thus be enabled to enter the more detailed treatment of our semiclassical method in chapter 7 immediately, referring to the intervening chapters on the classical mechanics whenever the need for expanded discussion arises.

2.1 Transition to chaos in closed billiards

Although we will devote some discussion to resonant cavities in the shape of deformed spheroids, most of our theory is developed for the dynamically simpler case of dielectric cylinders with a convex but asymmetric cross section. The classical ray optics in a closed cavity of this shape is equivalent to a point particle undergoing specular reflections at the walls and moving in a straight line segment otherwise. Because of the translational symmetry along the cylinder axis, the only nontrivial motion takes place in the projection on a two-dimensional cross section. This is the well-studied plane billiard problem. Any billiard trajectory can be completely specified by the sequence of points on the boundary where successive collisions occur. These points simply have to be connected by straight lines. When the boundary is everywhere convex, the bounce positions are uniquely identified by their polar angle along the boundary, i.e. one only needs the sequence ϕ_ν with ν numbering the consecutive reflections. This level of solution to the problem is referred to as *ray tracing*, and it allows us, e.g., to identify periodic orbits as infinite repetitions of some finite sequence ϕ_ν , $\nu = 1, \dots, N$.

The majority of trajectories are not periodic, and one can ask if it is possible to classify these orbits further. Indeed, it is known in ray optics⁵⁷ that non-periodic rays can under certain circumstances form *caustics*. These are curves or surfaces in real (configuration) space to which rays are tangent between any two successive reflections during their motion inside the cavity. Their existence can be identified by ray tracing, see Fig. 3.1(a). However, ray tracing quickly loses its usefulness when attempting a meaningful description of trajectories that are neither periodic

nor constrained by a caustic. This situation is encountered in deformed cavities.

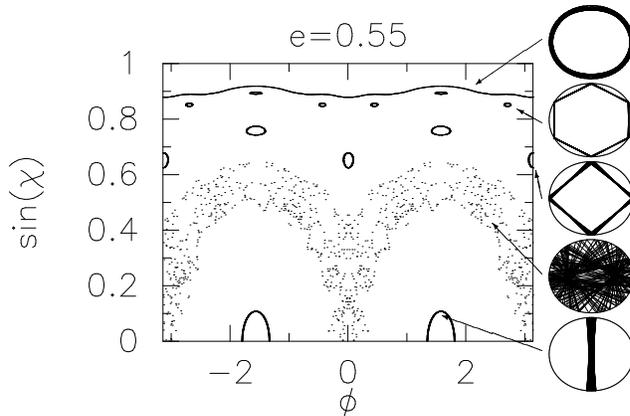


Figure 2.1: Poincaré surface of section of a quadrupole-deformed billiard, parametrized in polar coordinates by $r(\phi) \propto 1 + (e^2/4) \cos 2\phi$, at eccentricity $e = 0.55$, containing five trajectories. The corresponding real-space images are indicated to the right with arrows. The direction $\phi = 0$ coincides with the horizontal in the ray plots.

As was shown (to our knowledge for the first time) in Ref.⁵⁸, the study of ray optics in asymmetric resonators can profit greatly from the *Poincaré section* method that is ubiquitous in nonlinear dynamics. In contrast to ray tracing, the Poincaré surface of section (SOS) allows one to describe the *ray dynamics*, because information is retained about the motion in phase space, not just real space. This is realized by recording not only the sequence of bounce positions ϕ_ν along the boundary, but also the successive values of $\sin \chi_\nu$, where χ is the angle of incidence with respect to the normal to the billiard wall, as defined in the introduction. A trajectory is then plotted as a sequence of points $(\phi_\nu, \sin \chi_\nu)$ as shown in Fig. 2.1. The Poincaré section is obtained by repeating this for a (relatively small) number of orbits launched with different values of initial ϕ and $\sin \chi$, following each for typically 500 reflections. This plot represents a section through the classical mechanical phase space, because $\sin \chi$ is proportional to the tangential component of the momentum at the point of collision with the boundary, and this momentum component is canonically conjugate to the position along the boundary where the reflection occurred, which is in turn measured by ϕ . The trajectory shown at the top exhibits a caustic very close to the boundary, which is typical of whispering gallery orbits. In the SOS, this manifests itself in the existence of a one-dimensional curve $\sin \chi(\phi)$ on which all points generated by the orbit must lie. That the existence of $\sin \chi(\phi)$ and the caustic are equivalent is easy to see. Given any bounce position ϕ , the value of $\sin \chi$ is completely determined (up to a sign) by the requirement that the trajectory be tangent to the given caustic curve. The figure assumes a convex deformation of quadrupolar shape. It has in fact been shown by Lazutkin that there exist orbits with caustics for any convex billiard as long as it is sufficiently smooth (see section 4.10).

Note that the existence of a curve $\sin \chi(\phi)$ to which an orbit is confined and which we therefore call *invariant curve*, is highly non-trivial in view of the fact that angular momentum conservation is destroyed in the deformed billiard. This conservation law is the reason why in the circle $\sin \chi$ is constant for successive reflections. In the absence of the additional constraint equation due to angular momentum conservation, there is at first sight nothing that prevents a given trajectory from filling out a whole area in the SOS instead of merely a line. The figure does indeed show

a single trajectory which fills a finite domain of the SOS. The corresponding real-space image (the second from below) clearly does not show a caustic. Trajectories exploring a finite area of the SOS are called chaotic. The coexistence of trajectories with a caustic and chaotic orbits is typical of arbitrary convex billiards (except for circles and ellipses), and such billiards can be counted as members of the larger class of KAM systems. The scenario named after Kolmogorov, Arnol'd and Moser is encountered when a Hamiltonian mechanical system is perturbed away from a state where the number of conservation laws equals the number of degrees of freedom. In the transition to chaos that ensues, a crucial role is played by the periodic orbits. In their vicinity, the perturbation leads to the break-up of invariant curves, and hence to the disappearance of caustics. In the SOS, the simple invariant curves are then replaced by island chains. Three examples are shown in the figure. The island centers correspond to periodic orbits which are stable against deviations in their initial conditions. This is shown in the real-space plots, where trajectories close to the six-, four- and two-bounce periodic orbits are seen to oscillate around the periodic orbit without ever deviating from it beyond some limit. Each chain of stable islands is accompanied by an equal number of unstable periodic orbits, in whose vicinity trajectories become chaotic.

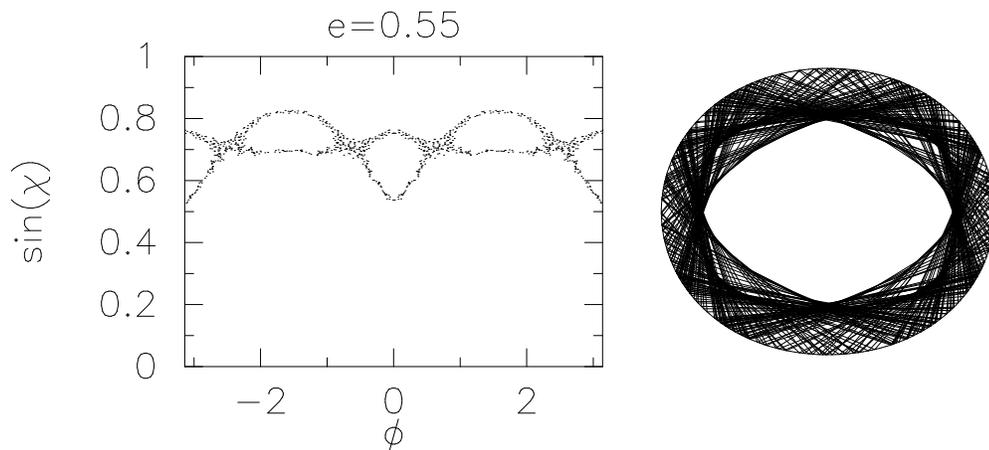


Figure 2.2: Left: A single trajectory in the quadrupole billiard at eccentricity $e = 0.55$ as in the previous figure. Right: The same trajectory in real space. This trajectory is seen in the SOS to encircle the islands shown in the previous figure. This is not apparent from the real-space plot.

To illustrate how the SOS method reveals information about trajectories that is obscured in the real space motion, consider Fig. 2.2, which shows a single chaotic trajectory in the SOS representation (left) and in real space (right). The real-space plot indicates that there is no caustic, because not all segments of the orbit are tangent to one and the same smooth curve between any two reflections. Other than this and the observation that small values of $\sin \chi$ do not seem to occur, one would be hard-pressed to draw any further conclusions from the ray tracing picture. However, the SOS shown to the left reveals additional structure in phase space: Although collisions occur at all possible positions ϕ , there are some *combinations* of $(\phi, \sin \chi)$ that are conspicuously absent, forming blank islands in the region of the SOS explored by the trajectory. Comparison with Fig. 2.1 shows, of course, that these are precisely the four-bounce islands whose corresponding stable orbits we already encountered. The fact that stable and chaotic motion form non-communicating, disjoint sets in the SOS, is of great importance for phase space

transport theory, and it follows simply from the limited amplitude of the island motion around stable periodic orbits. Any orbit started inside an island remains there forever due to the stability of the motion, implying that conversely no chaotic trajectory can ever enter an island. In a similar way, invariant curves that have not yet been destroyed by the perturbation form impenetrable barriers to chaotic trajectories, preventing any phase space flow from one side to the other.

2.2 Poincaré section and classical ray escape

The surface-of-section method is not only a tool allowing us to visualize the global phase space structure during the transition to chaos, but it is also ideally suited for a discussion of the whispering-gallery modes in open dielectric resonators, because the canonical variable $\sin \chi$ is exactly the quantity the value of which decides whether or not total internal reflection can take place. Classically, we can draw a straight horizontal line in the SOS at the critical value $\sin \chi_c = 1/n$, where n is the refractive index. If a trajectory generates points below this line, total internal reflection is violated and classical escape can occur. At this point we can already make the important statement that a convex asymmetric resonant cavity will certainly support long-lived whispering-gallery resonances, due to Lazutkin’s theorem which guarantees the existence of families of orbits with caustics. For these, the allowed fluctuations in $\sin \chi$ are bounded from below, so that orbits near $\sin \chi = 1$ will remain trapped classically. This existence theorem is not widely known in the optics community, but provides us with a justification to push on to higher deformations, knowing that long-lived states will not vanish immediately.

The KAM theorem allows us to make an even stronger statement which previously has neither been expected nor observed: There will be a *threshold deformation* above which a given resonance of interest ceases to be supported by an invariant curve of the KAM type. This is because the regions of chaos in the SOS grow with deformation, encroaching on the WG region near $\sin \chi = 1$. A qualitative change in the dependence of lifetimes on both frequency and deformation parameter should be anticipated when this threshold is exceeded, because the ray picture indicates a transition from motion on an invariant curve to chaotic motion with much weaker limitations on $\sin \chi$. In fact, the evolution of $\sin \chi$ for a chaotic trajectory can approximately be described as a diffusion process which is *biased* to lower values of $\sin \chi$ (cf. chapter 6).

A further problem which can be attacked efficiently with the help of the SOS is that of determining the emission directionality of a quasibound state. Since the diffusion which constitutes our classical escape mechanism is influenced by the structure of the SOS, we can make inferences about allowed and forbidden escape directions for the rays. Not only must the chaotic trajectory bypass stable islands, but the diffusion to lower $\sin \chi$ also becomes slower the higher is the initial $\sin \chi$. This relates directly to the existence of Lazutkin’s invariant curves at even larger $\sin \chi$. The slowness of the diffusion means, in a quantitative example, that it can take several hundred reflections for a trajectory started near a point $(\phi_0, \sin \chi_0)$ with $\sin \chi_0 \approx 0.8$ to reach a point in the SOS that has roughly the same ϕ_0 but a significantly lower value of $\sin \chi = 0.7$. Another way of saying this is that for a few hundred reflections $\sin \chi$ is roughly the same when returning to the same ϕ_0 , i.e. (since ϕ_0 is arbitrary) there exists an *almost* unique function $\sin \chi(\phi)$ on which the trajectory moves for intermediate times. As illustrated for two chaotic orbits in Fig. 2.2, this “almost” invariant curve guides the flow of trajectories in phase space even

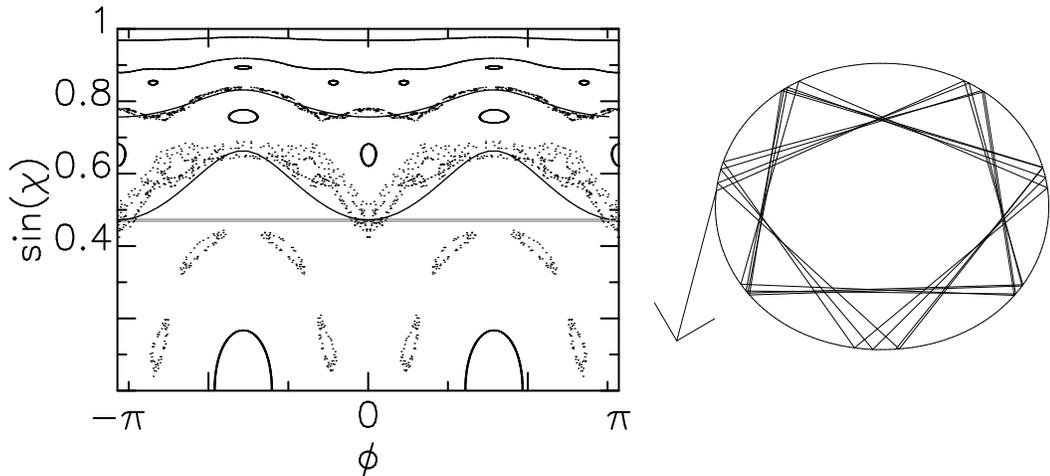


Figure 2.3: Fluctuations in $\sin \chi$ become stronger for orbits moving in the lower- $\sin \chi$ regions of the SOS. Still, the motion follows a well-defined pattern for an intermediate number of reflections, see text. The escape condition $\sin \chi_c$ is drawn as the grey line, Eq. (2.1) as solid lines. The phase-space flow near $\sin \chi_c$ determines the positions and orientations with which rays can escape.

though the motion is chaotic so that a finite area of the SOS will be filled by the orbit after a large number of reflections. The functional form of these curves can be derived in an *adiabatic approximation* due to Robnik and Berry⁵⁹, which assumes that the curvature κ of the billiard varies slowly on the length scale set by the distance between successive reflections, i.e. the small parameter is $d\kappa/d\phi \cos \chi \ll 1$. The result is (cf. section 6.5)

$$\sin \chi(\phi) = \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}}, \quad (2.1)$$

where \tilde{p} is the typical value around which $\sin \chi$ oscillates due to the oscillation of $\kappa(\phi)$. If this formula constitutes an acceptable description of the ray dynamics even near the critical line for escape, $\sin \chi_c = 1/n$, then we can immediately predict the escape directionality. Classical escape is possible only after the rays have diffused from their initial condition down to the adiabatic curve which is just tangent to $\sin \chi_c$ from above. Then, the trajectory will reach the minima of this curve within a few hundred bounces (or less). Since these are also the points of tangency to $\sin \chi_c$, escape will occur in the narrowly defined ϕ -intervals near the points of tangency. From Eq. (2.1), we see that these are just the points of highest curvature. Upon escape, the rays are refracted according to Snell's law. The slowness of the diffusion has an additional consequence here in that the distribution of $\sin \chi$ at escape is sharply peaked at $\sin \chi_c$. That is to say, there is no time for significant diffusion further down below $\sin \chi_c$ once the tangent invariant curve has been reached. The combination of narrow ϕ - and $\sin \chi$ - intervals for escape gives rise to a highly anisotropic emission pattern even in the far-field.

The emission directionality can be seen from the previous arguments to depend only on the phase space flow near $\sin \chi_c$ and not on the starting condition for the rays, as long as the latter is above the tangent adiabatic curve. This holds true even when island structure invalidates the adiabatic approximation near $\sin \chi_c$, resulting in a significantly different emission directionality, as we will discuss in subsection 11.3.3.

To make quantitative predictions for the resonance *lifetimes*, however, one needs to determine the *initial condition* for the ray diffusion, which is specific to

each individual resonance. That topic will be addressed in chapter 7.

Chapter 3

Hamiltonian systems

It is a surprising fact for nonspecialists that such an ancient discipline as classical mechanics turns out to be incomplete and makes its first steps nowadays whereas nonrelativistic quantum mechanics which only started in the twentieth century has acquired its mathematical apparatus entirely and now is a very elaborate and transparent branch of mathematical physics.

– V. F. Lazutkin, *KAM Theory and Semiclassical Approximations to Eigenfunctions*, 1993.

The second part of this statement must provoke objections from anyone studying the quantum mechanics of more than two interacting particles, but it sets the stage for the analysis we shall attempt in this thesis. We will be dealing with the single-particle Schrödinger equation and more importantly a close relative, the scalar wave equation of optics, where Lazutkin’s remark about the mathematical apparatus is not in dispute. Nevertheless, the problem of actually calculating solutions to non-separable wave equations can be a daunting task plagued with numerical difficulties. It is a current field of investigation to understand the observed connection between numerical problems and the occurrence of chaos in the classical counterpart of the physical system under consideration⁶⁰. The main thrust of our work will be in a different direction, however: Taking for granted the existence of numerical and experimental data about real wave phenomena, significant new insights into the physics behind the wave solutions will be gained by making use of classical ideas. The mathematical apparatus of classical mechanics therefore will provide many of the concepts that are needed in subsequent chapters. We shall introduce this subject in a way that is tailored to our purposes. We will primarily be interested in two-dimensional billiard systems, whose classical mechanics can be discussed conveniently from the point of view of their Poincaré map – a two-dimensional, discrete, “stroboscopic” sampling of the trajectories as they evolve in phase space. However, for a semiclassical quantization of the billiard, one needs to interpret this map in the light of the underlying continuous-time dynamics, which is that of a conservative Hamiltonian system. Therefore, we first discuss the full phase space of these systems, leading us from the fundamental role of symplectic matrices and generating functions to conservation laws and action-angle variables in integrable systems. The examples that are presented will serve as the basis for an understanding of all other specific systems discussed in the later chapters.

Billiards are closed domains in which a point particle is confined by infinitely hard walls. The latter has the effect that some derivatives of the Hamiltonian be-

come singular when momenta are reversed abruptly upon reflection at the boundary. However, we will include billiards under the species of Hamiltonian systems, and use them as illustrative examples. One might think of the hard wall condition as being the limiting case of some family of smooth potentials which do generate differentiable Hamiltonian dynamics. In that way, the discussion requires less formalism as compared to the more rigorous theory based on the geometry of symplectic manifolds, of which we only use a bare minimum. The interested reader is referred to the book by Arnol'd⁶⁴ for a more formal treatment of Hamiltonian mechanics using the calculus of differential forms, and to Lazutkin's book, Ref.⁶¹, for an advanced study that includes the KAM transition in billiard systems. The treatment that we present tries to capture the essentials of the geometry of phase space, and thus departs from usual textbook approaches^{62,63}. For example, we do not go through the variational principles that underly all of classical mechanics, and which can also be made the basis of a study of billiard dynamics, cf. Ref.⁶⁵.

It is common^{82,81} to introduce the notion of chaos in Hamiltonian systems by discussing classical perturbation theory and its breakdown, and this is in fact the central point of the famous Kolmogorov-Arnol'd-Moser (KAM) theorem. However, we do not follow this route, because the important concepts can be introduced more directly and with less formalism after the billiard motion has been cast in the form of a map on the Poincaré surface of section. The pivotal role is then played by the Poincaré-Birkhoff fixed point theorem.

3.1 The flow in phase space

We shall take as our *defining equation* of Hamiltonian mechanics the equation of motion in phase space for a system with N degrees of freedom,

$$\frac{df}{dt} = \{f, H\} + \frac{\partial f}{\partial t}, \quad (3.1)$$

$$\{f, H\} = \sum_{i=1}^N \left[\frac{\partial f}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q_i} \frac{\partial f}{\partial p_i} \right], \quad (3.2)$$

where f is some function of coordinates q_i , momenta p_i and continuous time t ; H is the Hamiltonian function, and curly brackets indicate the Poisson bracket. This can be considered for now as merely a mathematician's way of defining the subject, and it is not meant to explain classical mechanics, because a certain familiarity with the subject on the part of the reader is assumed. This starting point is consistent with the role classical mechanics plays in our work: none of the systems of interest to us are genuinely classical, but we know that in quantum mechanics the correct wave equation is obtained by the usual quantization of the classical Hamiltonian (replacing momenta by $-i\hbar\nabla$ etc.); and that is precisely the way in which Eq. (3.1) leads back to the Heisenberg equation of motion⁶⁶. In subsequent chapters, a prominent question will be how to use the results of classical mechanics in order to find *approximate* solutions to the corresponding wave equation.

Equation (3.1) with $f = q_i$ or $f = p_i$ immediately implies Hamilton's equations,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad (3.3)$$

which hold even if H is time-dependent. The Poisson bracket with H can be thought

of as a differential operator,

$$\{\cdot, H\} = \sum_{i=1}^N \left[\frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i} \right], \quad (3.4)$$

which in turn can be written as the following bilinear product of two $2N$ -dimensional vectors:

$$\{\cdot, H\} = - \left(\vec{\nabla}_x H \right)^t \Gamma \vec{\nabla}_x. \quad (3.5)$$

Here we have defined the $2N \times 2N$ matrix

$$\Gamma = \begin{pmatrix} \emptyset_N & -\mathbf{1}_N \\ \mathbf{1}_N & \emptyset_N \end{pmatrix}, \quad (3.6)$$

and the phase space gradient

$$\vec{\nabla}_x \equiv \begin{pmatrix} \partial/\partial \mathbf{p} \\ \partial/\partial \mathbf{q} \end{pmatrix}. \quad (3.7)$$

Unlike a scalar product, this is a skew-symmetric bilinear (also called *symplectic*) form, as follows already from the definition Eq. (3.1). Using this vectorial form and the corresponding definition

$$\vec{x} \equiv \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}, \quad (3.8)$$

Eq. (3.3) can be written as

$$\frac{d}{dt} \vec{x} = \Gamma \vec{\nabla}_x H. \quad (3.9)$$

Clearly, H is a constant of motion if it has no explicit time dependence; but even if H is not conserved we still have

$$0 = -\{H, H\} = \left(\vec{\nabla}_x H \right)^T \Gamma \vec{\nabla}_x H = \left(\vec{\nabla}_x H \right)^T \frac{d}{dt} \vec{x}. \quad (3.10)$$

This means that the phase space trajectories are perpendicular to the gradient of H at all times. We now specialize to time-independent Hamiltonians, where $H = E$. For such Hamiltonians, the trajectory moves on a hypersurface $H(\vec{x}) = E$.

3.2 The first Poincaré invariant

What distinguishes the flow defined by Eq. (3.9) from a general flow that might be encountered in fluid dynamics, is the existence of additional quantities that remain invariant over time; one of which is the phase space volume itself (Liouville's theorem). Of particular importance in the discussion below is the first of Poincaré's integral invariants,

$$\oint_C \mathbf{p} \cdot d\mathbf{q}, \quad (3.11)$$

where C is the \mathbf{q} -space projection of any closed curve in phase space, and all p_i and q_i are evaluated at a the same time t . The quantity in Eq. (3.11) is independent of t , even though C will change with t according to the equations of motion. To show the invariance of Eq. (3.11), consider each term $p_i dq_i$ in the scalar product $\mathbf{p} \cdot d\mathbf{q}$ separately and apply the integration [of course each such integral is just a special

case of Eq. (3.11) where only one q_i varies along C . The result is clearly just the area $|A_i|$ in the p_i - q_i plane enclosed by the loop, which can also be expressed as

$$|A_i| = \oint p_i dq_i = \int_{A_i} dp_i dq_i. \quad (3.12)$$

Now after some time Δt , the point (q_i, p_i) is mapped onto a new point (\bar{q}_i, \bar{p}_i) so that the new area $|\bar{A}_i|$ is

$$|\bar{A}_i| = \int \bar{p}_i d\bar{q}_i = \oint \bar{p}_i d\bar{q}_i. \quad (3.13)$$

The last contour integral is equal to the original one, $\oint p_i dq_i$ if one can show $|\bar{A}_i| = |A_i|$. In terms of the old coordinates, we have

$$|\bar{A}_i| = \int_{A_i} \left| \frac{\partial(\bar{q}_i, \bar{p}_i)}{\partial(q_i, p_i)} \right| dp_i dq_i. \quad (3.14)$$

In order to find the Jacobian in this integral, we let $\Delta t \rightarrow dt$ become infinitesimal, so that $\bar{q}_i \rightarrow q_i + \delta q_i$ and $\bar{p}_i \rightarrow p_i + \delta p_i$, with $\delta q_i = \dot{q}_i dt$ and $\delta p_i = \dot{p}_i dt$. Then to leading order one can equate

$$\frac{\partial H}{\partial p_i} \approx \frac{\partial H}{\partial \bar{p}_i}, \quad (3.15)$$

and Hamilton's equations of motion yield

$$\bar{q}_i = q_i + \dot{q}_i dt = q_i + \frac{\partial H}{\partial \bar{p}_i} dt, \quad (3.16)$$

$$\bar{p}_i = p_i + \dot{p}_i dt = p_i - \frac{\partial H}{\partial q_i} dt. \quad (3.17)$$

Here we take as the independent variables the old coordinate q_i and the new momentum \bar{p}_i , considering the remaining quantities H , \bar{q}_i and p_i as functions of the former. Integrating the first equation over \bar{p}_i or the second over q_i , we obtain a *generating function*

$$F(q_i, \bar{p}_i) \equiv q_i \bar{p}_i + H(q_i, \bar{p}_i) dt, \quad (3.18)$$

from which both Eqs. (3.16) can be recovered as derivatives,

$$p_i = \frac{\partial F}{\partial q_i} \quad (3.19)$$

$$\bar{q}_i = \frac{\partial F}{\partial \bar{p}_i}. \quad (3.20)$$

From the existence of this generating function it can be shown that the Jacobian of the infinitesimal transformation is unity,

$$\det \left[\frac{\partial(\bar{q}_i, \bar{p}_i)}{\partial(q_i, p_i)} \right] = 1. \quad (3.21)$$

The proof is given as theorem 3.3.1.1 in the next section, because we will refer to it in various different forms and contexts later on. We devote a whole section to that and related issues because they are in fact central to an understanding of Hamiltonian systems and maps.

To complete the discussion of the Poincaré invariant, note that since the Jacobian for any finite Δt is the limit of a product of infinitesimal transformations, that Jacobian is also unity. Therefore, Eq. (3.14) reduces to Eq. (3.12), which proves that Eq. (3.11) is indeed an invariant.

A geometric interpretation of this Poincaré invariant can be obtained by rewriting the Jacobian in Eq. (3.21) as a cross product

$$\det\left[\frac{\partial(\bar{q}_i, \bar{p}_i)}{\partial(q_i, p_i)}\right] = \begin{pmatrix} \partial\bar{q}_i/\partial q_i \\ \partial\bar{p}_i/\partial q_i \end{pmatrix} \times \begin{pmatrix} \partial\bar{q}_i/\partial p_i \\ \partial\bar{p}_i/\partial p_i \end{pmatrix}. \quad (3.22)$$

The two vectors appearing here point in the direction of the coordinate lines that are obtained by applying the time evolution to the canonical basis of the p_i - q_i plane. They form a parallelogram whose area, given by the cross product above, remains constant in time. This is not the same as the conservation of phase space volume, because we have been concerned with simple two-dimensional areas in each p_i - q_i plane, and their separate conservation is in fact a stronger statement than the constancy of phase space volume (the latter can be derived from the former⁶⁷, but we will not need it here).

Having derived the first of Poincaré's integral invariants, the curious reader will want to know what the other Poincaré invariants are. We only mention that the phase space volume is in fact the N -th invariant (for N degrees of freedom), and refer to the literature for further details^{82,38}. Only the first invariant will be needed later, when we introduce action-angle variables for integrable systems.

3.3 Transformations with a generating function

3.3.1 Lagrange bracket

In the previous section the crucial step in showing the invariance of $\oint \mathbf{p} d\mathbf{q}$ was to obtain a unit Jacobian for an infinitesimal transformation. It is the existence of a generating function which allows us to make this and some other far-reaching statements about the transformation it applies to. Consider the two pairs (\mathbf{q}, \mathbf{p}) and (\mathbf{u}, \mathbf{v}) , where the vectors are N -dimensional. Denote the corresponding $2N$ -dimensional phase space vectors by

$$\vec{x} \equiv \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}, \quad (3.23)$$

$$\vec{w} \equiv \begin{pmatrix} \mathbf{v} \\ \mathbf{u} \end{pmatrix}. \quad (3.24)$$

Define further the *Lagrange bracket* of the transformation from (\mathbf{q}, \mathbf{p}) to (\mathbf{u}, \mathbf{v}) as

$$[x_j, x_k] \equiv \left(\frac{\partial\vec{w}}{\partial x_k}\right)^T \Gamma \frac{\partial\vec{w}}{\partial x_j}. \quad (3.25)$$

Note that for $N = 1$, this can be used directly to express the Jacobian determinant of the transformation,

$$[q, p] = \det\left[\frac{\partial(u, v)}{\partial(q, p)}\right]. \quad (3.26)$$

The difference to the Poisson bracket is that the summation in Eq. (3.25) is over the new variables in the numerator, and not over the old variables with respect to

which the derivatives are formed. Poisson and Lagrange bracket become equivalent in the case $N = 1$, because

$$[q, p] = \{u, v\}. \quad (3.27)$$

3.3.1.1 Theorem

If there exists a function $F(\mathbf{q}, \mathbf{v})$ such that

$$\mathbf{p} = \left. \frac{\partial F}{\partial \mathbf{q}} \right|_{\mathbf{v}}, \quad \mathbf{u} = \left. \frac{\partial F}{\partial \mathbf{v}} \right|_{\mathbf{q}}, \quad (3.28)$$

then the Lagrange bracket [as defined in Eq. (3.25)] satisfies

$$[x_i, x_j] = \Gamma_{ij} \quad (i = 1 \dots 2N), \quad (3.29)$$

with the matrix Γ as defined in Eq. (3.6), or more explicitly

$$[q_i, q_j] = [p_i, p_j] = 0, \quad [q_i, p_j] = \delta_{ij} \quad (i = 1 \dots N). \quad (3.30)$$

Another useful way of rewriting this statement is obtained by writing out the definition of the Lagrange bracket. Then one can write Eq. (3.29) as an equation for the $2N \times 2N$ matrix of derivatives, $D\vec{w}/D\vec{x}$:

$$\left(\frac{D\vec{w}}{D\vec{x}} \right)^T \Gamma \frac{D\vec{w}}{D\vec{x}} = \Gamma. \quad (3.31)$$

A matrix that satisfies this condition is called *symplectic*. Because $\Gamma \Gamma = -1_N$, a symplectic matrix A with $A^T \Gamma A = \Gamma$ automatically satisfies the reverse relation

$$A \Gamma A^T = \Gamma. \quad (3.32)$$

Since the infinitesimal transformation generating the Hamiltonian flow satisfies the symplectic condition due to Eq. (3.18), the flow is also called symplectic. Hamiltonian systems are in fact only a special case of the more general class of symplectic flows⁶¹.

Proof

Consider F as a function of \vec{x} , then the partial derivatives with respect to these old variables are

$$\left. \frac{\partial F}{\partial q_i} \right|_{\mathbf{p}} = \left. \frac{\partial F}{\partial q_i} \right|_{\mathbf{v}} + \sum_k \left. \frac{\partial F}{\partial v_k} \right|_{\mathbf{q}} \left. \frac{\partial v_k}{\partial q_i} \right|_{\mathbf{p}} \quad (3.33)$$

$$= p_i + \sum_k u_k \left. \frac{\partial v_k}{\partial q_i} \right|_{\mathbf{p}}, \quad (3.34)$$

$$\left. \frac{\partial F}{\partial p_i} \right|_{\mathbf{q}} = \sum_k \left. \frac{\partial F}{\partial v_k} \right|_{\mathbf{q}} \left. \frac{\partial v_k}{\partial p_i} \right|_{\mathbf{q}} \quad (3.35)$$

$$= \sum_k u_k \left. \frac{\partial v_k}{\partial p_i} \right|_{\mathbf{q}}. \quad (3.36)$$

If we now form the second derivatives, the order of differentiation has to be interchangeable because dF is an exact differential. As a result we obtain the following

relations:

$$\begin{aligned}
(A) \quad & \frac{\partial^2 F}{\partial q_i \partial q_j} = \frac{\partial^2 F}{\partial q_j \partial q_i} \\
\Rightarrow & \sum_k \left\{ \frac{\partial u_k}{\partial q_j} \frac{\partial v_k}{\partial q_i} + u_k \frac{\partial^2 v_k}{\partial q_i \partial q_j} \right\} = \sum_k \left\{ \frac{\partial u_k}{\partial q_i} \frac{\partial v_k}{\partial q_j} + u_k \frac{\partial^2 v_k}{\partial q_j \partial q_i} \right\} \\
\Rightarrow & [q_i, q_j] = 0 \quad (\text{after canceling the last terms}) \tag{3.37}
\end{aligned}$$

$$\begin{aligned}
(B) \quad & \frac{\partial^2 F}{\partial q_i \partial p_j} = \frac{\partial^2 F}{\partial p_j \partial q_i} \\
\Rightarrow & \delta_{ij} + \sum_k \left\{ \frac{\partial u_k}{\partial p_j} \frac{\partial v_k}{\partial q_i} + u_k \frac{\partial^2 v_k}{\partial q_i \partial p_j} \right\} = \sum_k \left\{ \frac{\partial u_k}{\partial q_i} \frac{\partial v_k}{\partial p_j} + u_k \frac{\partial^2 v_k}{\partial p_j \partial q_i} \right\} \\
\Rightarrow & [q_i, p_j] = \delta_{ij} \tag{3.38}
\end{aligned}$$

$$\begin{aligned}
(C) \quad & \frac{\partial^2 F}{\partial p_i \partial p_j} = \frac{\partial^2 F}{\partial p_j \partial p_i} \\
\Rightarrow & \sum_k \left\{ \frac{\partial u_k}{\partial p_j} \frac{\partial v_k}{\partial p_i} + u_k \frac{\partial^2 v_k}{\partial p_i \partial p_j} \right\} = \sum_k \left\{ \frac{\partial u_k}{\partial p_i} \frac{\partial v_k}{\partial p_j} + u_k \frac{\partial^2 v_k}{\partial p_j \partial p_i} \right\} \\
\Rightarrow & [p_i, p_j] = 0. \tag{3.39}
\end{aligned}$$

This completes the proof of Eq. (3.30), and Eq. (3.29) follows from the definition of Γ in Eq. (3.6).

3.3.2 Canonical transformations

We now apply Eq. (3.31) to a Hamiltonian flow \vec{x} which hence satisfies Eq. (3.9). Then the following theorem holds:

3.3.2.1 Theorem

Consider the $2N$ dimensional vectors \vec{x} and \vec{w} , where

$$\frac{d}{dt} \vec{x} = \Gamma \vec{\nabla}_x H, \tag{3.40}$$

and the matrix of first derivatives for the transformation from \vec{x} to \vec{w} is symplectic. Then \vec{w} is also a Hamiltonian flow, i.e. satisfies Eq. (3.40), and the transformation from \vec{x} to \vec{w} is called *canonical*.

The symplectic condition is satisfied in particular when there exists a generating function with the properties of Eq. (3.28). Canonical transformations are an essential tool in the solution of many mechanics problems, and we will use them to introduce the action-angle variables later on.

Proof

Use the matrix of first derivatives to perform a transformation of variables in

$$\frac{d}{dt} \vec{w} = \frac{D\vec{w}}{D\vec{x}} \frac{d}{dt} \vec{x} = \frac{D\vec{w}}{D\vec{x}} \Gamma \vec{\nabla}_x H, \tag{3.41}$$

where we used Eq. (3.9). Transforming the gradient to the variable \vec{w} , we get

$$\frac{d}{dt} \vec{w} = \frac{D\vec{w}}{D\vec{x}} \Gamma \left(\frac{D\vec{w}}{D\vec{x}} \right)^T \vec{\nabla}_w H = \Gamma \vec{\nabla}_w H, \tag{3.42}$$

using Eqs. (3.31) and (3.32) in the last equality. This constitutes Hamilton's equations for \vec{w} .

3.3.3 Eigenvalues and determinant of symplectic matrices

If A is a symplectic matrix, its eigenvalues come in pairs $\lambda_i, 1/\lambda_i$, so that its determinant is unity. We do not have to do any work to prove this for $N = 1$ because of Eq. (3.26) which tells us that the Jacobian determinant is unity provided Eq. (3.31) holds. Since the matrix in question is 2×2 , its eigenvalues have to be reciprocals of each other. To prove the statement in the general case, assume A is diagonalized by similarity transformation with a matrix X . First we note that the resulting diagonal matrix is trivially symmetric, so that we can transpose to find that A^T has the same eigenvalues as A . Consider an eigenvalue \vec{y} such that

$$A^T \vec{y} = \lambda \vec{y}, \quad (3.43)$$

then we multiply from the left by $\Gamma (A^T)^{-1}/\lambda$ to get

$$\frac{1}{\lambda} \Gamma \vec{y} = \Gamma (A^T)^{-1} \vec{y}. \quad (3.44)$$

On the righthand side, we want to use the symplectic condition,

$$A^T \Gamma A = \Gamma \quad (3.45)$$

$$\Rightarrow \Gamma A = (A^T)^{-1} \Gamma \quad (3.46)$$

$$\Rightarrow (A^T)^{-1} = -\Gamma A \Gamma, \quad (3.47)$$

where we used $\Gamma \Gamma = -1$. With this, Eq. (3.44) becomes

$$\frac{1}{\lambda} \Gamma \vec{y} = A \Gamma \vec{y}, \quad (3.48)$$

which means that $1/\lambda$ is an eigenvalue of A (with eigenvector $\Gamma \vec{y}$) as we claimed.

3.4 Integrable systems

In most isolated mechanical systems occurring in nature, the only conserved quantity is the total energy. As soon as there is more than one degree of freedom, it becomes very difficult to solve the equations of motion under these circumstances. Practically all textbook examples for motion in more than one dimension therefore belong to a special class of systems which are called *integrable* because the trajectory can be found by a set of quadratures.

Assume that in addition to the Hamiltonian H , we found another function $K(\vec{x})$ of the phase space variables that is also conserved, i.e. $\{K, H\} = 0$. The system is integrable if there are N conserved quantities like H and K which are also pairwise independent and have vanishing Poisson bracket with each other. Let us explore this situation for the special case of $N = 2$. Independence of H and K means that we require their gradients to be linearly independent,

$$\vec{\nabla}_x H \neq \alpha \vec{\nabla}_x K, \quad (3.49)$$

everywhere except at isolated points in the four-dimensional phase space. The trajectory is constrained by the two equations $H(\vec{x}) = E$ and $K(\vec{x}) = \text{const}$, and therefore lies on a 2D surface F . Any point on that surface can serve as the initial condition for a unique trajectory consistent with the conservation laws, and the resulting set of trajectories defines a new vector field: the projection of $\frac{d}{dt} \vec{x}$ onto the

local tangent plane of F . This 2D field covers the whole surface, and it is nonzero everywhere provided $\frac{d}{dt}\vec{x} \neq 0$. The latter condition is assumed to be valid in the cases of interest here, as is done in the textbook by Arnold⁶⁴.

In order for F to be covered by an everywhere nonvanishing vector field, its topology must be that of a torus. One can easily visualize that a sphere does not admit such a field (the problem is equivalent to combing the hair on a coconut without introducing a part or an eddy³⁸). Thus we have to introduce a handle on the sphere, which can then be deformed into a torus. This topology is consistent with our requirement. In fact it is the only such topology, because introducing further handles on the torus will again make it impossible to avoid points of vanishing field.

The generalization of the above arguments to more than two degrees of freedom is that the trajectory will move on an N -dimensional torus in the $2N$ -dimensional phase space. However, the case $N = 2$ is of particular interest to us.

Concerning the assumption that the vector field on F is nonvanishing, it should be pointed out that there are counter-examples, namely the rational polygon billiards⁶⁸, which consist of flat billiards bounded by straight line segments enclosing an angle which is a rational multiple of π . These systems are called *pseudo-integrable*, and the motion of a trajectory is confined to a two-dimensional manifold in phase space. However, the possibility of (classical) “beam splitting” at sharp corners implies that singularities of the above vector field can occur. As a result, the surface F is a multi-handed sphere, which becomes so complicated that the system shows some properties commonly associated with nonintegrable dynamics, as for example level repulsion in the quantum mechanical spectrum.

3.5 Action-angle variables

The torus on which a trajectory in an integrable system moves suggests that we choose coordinates that are adapted to this special topology, namely the action-angle variables. In standard textbooks on classical mechanics, these variables are introduced either for one-dimensional systems only⁶², or it is assumed that a separation of variables in the Hamilton-Jacobi equation (see below) has been performed such that one is left with N effectively one-dimensional problems^{63,69}. However, we want to apply the action-angle formalism to systems that cannot be reduced to these simple cases. Therefore, we want to use only the definition of integrability given above. This will allow us to define actions and angles even for tori that are embedded in a partially chaotic phase space.

For the sake of clarity we will again specialize to $N = 2$ degrees of freedom, so that there are exactly two types of loops on the torus that can not be transformed into one another: If we think of the familiar doughnut, we can generate these different curves by cutting either through the large or the small cross section. Then we can calculate the first Poincaré invariant along two such inequivalent curves and call them J_1 , J_2 , or *action* variables due to their dimension.

An important point here is that the particular choice of loops is immaterial as long as they are topologically inequivalent. We have not shown this so far – all we know is that the J_i are time-independent; but the fact that J_1 and J_2 depend only on the topology of the paths is crucial if we want to use them as a complete parametrization of the torus that could then replace the old parameters H and K .

3.5.1 Path-independence on the torus

The claim is that we obtain the same Poincaré invariant,

$$J_i = \oint_{C_i} \mathbf{p} d\mathbf{q}, \quad (3.50)$$

if the curve C_i is deformed without changing the number of times we wind around either of the torus cross sections described above. Such a deformation can be introduced simply by adding to the original loop integral another closed loop, part of which is antiparallel to the first path and thus cancels the corresponding segment. The added closed loop C_0 must be topologically reducible to a point in order to leave the number of windings unchanged. The projection of \mathbf{p} onto the \mathbf{q} -plane has at least two sheets, corresponding to the different values \mathbf{p} can have on the “top” or “bottom” side of the torus (there may be more sheets if the torus lies oblique to the \mathbf{q} -plane). We want to show that

$$\oint_{C_0} \mathbf{p} d\mathbf{q} \quad (3.51)$$

is zero for a reducible loop on each sheet of \mathbf{p} . An equivalent statement for the path-independence is that each sheet of $\mathbf{p}(\mathbf{q})$ must be irrotational if \mathbf{J} is kept fixed.

All we need for the proof (with $N = 2$) is that there exist two conserved phase-space functions H and K that have vanishing Poisson bracket,

$$\{K, H\} = 0. \quad (3.52)$$

First, let us use the conservation of H to write one of the momenta as a function of the remaining three phase space variables and the energy:

$$\tilde{p}_1 = p_1(q_1, q_2, p_2, H) \quad \text{or} \quad (3.53)$$

$$\tilde{p}_2 = p_2(q_1, q_2, p_1, H), \quad (3.54)$$

where the tilde identifies a function that depends on one momentum and both coordinates. If we use in addition the conservation of K , both momenta can be written as a function of the two coordinates only, and this is how the p_i without tilde should be understood. The explicit momentum dependence in the \tilde{p}_i allows us to take the partial derivative

$$\frac{\partial \tilde{p}_1}{\partial p_2} = \left(\frac{\partial \tilde{p}_2}{\partial p_1} \right)^{-1}. \quad (3.55)$$

where all other quantities are considered fixed. We want to express this in terms of derivatives of the Hamiltonian, which was used to define \tilde{p}_i . A very useful formalism that allows one to deal with transformations of variables in partial derivatives (in particular to keep track of sign changes that are otherwise not so obvious) is based on the Jacobi determinant¹. The above derivative can then be written as

$$\begin{aligned} \left. \frac{\partial \tilde{p}_1}{\partial p_2} \right|_H &= \det \left[\frac{\partial(\tilde{p}_1, H)}{\partial(p_2, H)} \right] = \det \left[\frac{\partial(\tilde{p}_1, H)}{\partial(p_1, p_2)} \frac{\partial(p_1, p_2)}{\partial(p_2, H)} \right] \\ &= - \det \left[\frac{\partial(\tilde{p}_1, H)}{\partial(p_1, p_2)} \frac{\partial(p_1, p_2)}{\partial(H, p_2)} \right] = - \left. \frac{\partial H}{\partial p_2} \right|_{p_1} \left(\left. \frac{\partial H}{\partial p_1} \right|_{p_2} \right)^{-1}. \end{aligned} \quad (3.56)$$

¹to understand the manipulations, the reader can simply write out the determinants.

Note the minus sign that arises when we exchange p_2 and H in the second Jacobian. Similarly, we have for the partial derivative with respect to the corresponding (conjugate) coordinate,

$$\begin{aligned}\left.\frac{\partial \tilde{p}_i}{\partial q_i}\right|_H &= \frac{\partial(\tilde{p}_i, H)}{\partial(q_i, H)} = -\frac{\partial(\tilde{p}_i, H)}{\partial(p_i, q_i)} \frac{\partial(p_i, q_i)}{\partial(H, p_i)} \\ &= -\left.\frac{\partial H}{\partial q_i}\right|_{p_i} \left(\left.\frac{\partial H}{\partial p_i}\right|_{q_i}\right)^{-1}.\end{aligned}\quad (3.57)$$

If we insert \tilde{p}_1 into K , then the only remaining momentum variable that is still unknown is p_2 . We can in principle use the conservation of K to write p_2 as a function of the q_i alone. Our goal is to solve for $\partial p_2/\partial q_1$. Further below, we shall similarly solve for $\partial p_1/\partial q_2$, and then use the results to show that $\partial p_2/\partial q_1 - \partial p_1/\partial q_2 = 0$, i.e. that \mathbf{p} is irrotational, which proves the path-independence of the action integral.

Now take the *total* derivative of the equation $K = \text{const}$ with respect to q_1 , given the substitution just described. The result is

$$0 = \frac{dK}{dq_1} = \frac{\partial K}{\partial q_1} + \frac{\partial K}{\partial p_1} \frac{d\tilde{p}_1}{dq_1} + \frac{\partial K}{\partial p_2} \frac{dp_2}{dq_1}.\quad (3.58)$$

In the last term, p_2 does not carry a tilde because, as mentioned above, it is only a function of the q_i . This actually means that

$$\frac{dp_2}{dq_1} = \frac{\partial p_2}{\partial q_1}.\quad (3.59)$$

Alternatively, we can insert \tilde{p}_2 into K and differentiate with respect to q_2 , obtaining thus the analogous equation

$$0 = \frac{dK}{dq_2} = \frac{\partial K}{\partial q_2} + \frac{\partial K}{\partial p_1} \frac{\partial p_1}{\partial q_2} + \frac{\partial K}{\partial p_1} \frac{d\tilde{p}_2}{dq_2}.\quad (3.60)$$

The remaining total derivatives can be expanded as

$$\frac{d\tilde{p}_i}{dq_i} = \frac{\partial \tilde{p}_i}{\partial q_i} + \frac{\partial \tilde{p}_i}{\partial p_j} \frac{\partial p_j}{\partial q_i},\quad (3.61)$$

where $j \neq i$ labels the momentum that appears explicitly in \tilde{p}_i . This is now used in Eqs. (3.58) and (3.60) to obtain the two equations

$$-\frac{\partial K}{\partial q_1} - \frac{\partial K}{\partial p_1} \frac{\partial \tilde{p}_1}{\partial q_1} = \frac{\partial p_2}{\partial q_1} \left(\frac{\partial K}{\partial p_2} + \frac{\partial K}{\partial p_1} \frac{\partial \tilde{p}_1}{\partial p_2} \right),\quad (3.62)$$

$$-\frac{\partial K}{\partial q_2} - \frac{\partial K}{\partial p_2} \frac{\partial \tilde{p}_2}{\partial q_2} = \frac{\partial p_1}{\partial q_2} \left(\frac{\partial K}{\partial p_1} + \frac{\partial K}{\partial p_2} \frac{\partial \tilde{p}_2}{\partial p_1} \right).\quad (3.63)$$

In the second equation, we multiply both sides by $\partial \tilde{p}_1/\partial p_2$ and use Eq. (3.55). This makes the term in brackets on the righthand side identical to the one in the first equation. Then we divide the first by the second equation to obtain

$$\frac{\partial p_2}{\partial q_1} \left(\frac{\partial p_1}{\partial q_2} \right)^{-1} = \left(\frac{\partial K}{\partial q_1} + \frac{\partial K}{\partial p_1} \frac{\partial \tilde{p}_1}{\partial q_1} \right) \left(\frac{\partial K}{\partial q_2} + \frac{\partial K}{\partial p_2} \frac{\partial \tilde{p}_2}{\partial q_2} \right)^{-1} \left(\frac{\partial \tilde{p}_1}{\partial p_2} \right)^{-1}.\quad (3.64)$$

Here, one can eliminate all derivatives of the dependent momenta (carrying a tilde) through the use of Eqs. (3.56) and (3.57):

$$\frac{\partial p_2}{\partial q_1} \left(\frac{\partial p_1}{\partial q_2} \right)^{-1} = - \left[\frac{\partial K}{\partial q_1} - \frac{\partial K}{\partial p_1} \frac{\partial H}{\partial q_1} \left(\frac{\partial H}{\partial p_1} \right)^{-1} \right] \times \quad (3.65)$$

$$\left[\frac{\partial K}{\partial q_2} - \frac{\partial K}{\partial p_2} \frac{\partial H}{\partial q_2} \left(\frac{\partial H}{\partial p_2} \right)^{-1} \right]^{-1} \frac{\partial H}{\partial p_1} \left[\frac{\partial H}{\partial p_2} \right]^{-1} \quad (3.66)$$

$$= - \left[\frac{\partial K}{\partial q_1} \frac{\partial H}{\partial p_1} - \frac{\partial K}{\partial p_1} \frac{\partial H}{\partial q_1} \right] \left[\frac{\partial K}{\partial q_2} \frac{\partial H}{\partial p_2} - \frac{\partial K}{\partial p_2} \frac{\partial H}{\partial q_2} \right]^{-1} \quad (3.67)$$

But the numerator is just the negative of the denominator because of the condition $\{K, H\} = 0$. Thus we finally arrive at

$$\frac{\partial p_2}{\partial q_1} = \frac{\partial p_1}{\partial q_2}. \quad (3.68)$$

This means that the curl of \mathbf{p} is zero on the sheet under consideration. Therefore, the action integral in Eq. (3.50) is path-independent as long as the deformation can be described by attaching a reducible loop to the original path, as described above.

3.5.2 The action function

This fact allows one to define the following function of position, which we will also call the *action function* (not to be confused with the action variables):

$$S(\mathbf{q}, \mathbf{J}) = \int_{\mathbf{q}_0}^{\mathbf{q}} \mathbf{p}(\mathbf{J}, \mathbf{q}') d\mathbf{q}', \quad (3.69)$$

where the starting point \mathbf{q}_0 is arbitrary but fixed. Because of the path-independence of this integral, S is indeed only a function of the arguments \mathbf{q}, \mathbf{J} (dS is exact) under the above condition of unchanged number of windings around the torus. Hence this function can serve as the generating function for a canonical transformation from the variables q_i, p_i to a new set Φ_i, J_i according to the rule

$$\mathbf{p} = \frac{\partial S}{\partial \mathbf{q}} \quad (3.70)$$

$$\Phi = \frac{\partial S}{\partial \mathbf{J}}. \quad (3.71)$$

The second equation defines the *angle variables*. The reason for this definition is that it guarantees that the transformation is canonical, which means Φ_i, J_i satisfy the same equations of motion as q_i, p_i . This follows immediately from the theorem 3.3.2.1. Note also that S has as many different sheets over the \mathbf{q} plane as does \mathbf{p} , due to Eq. (3.70).

As a consequence of the above definition, H can depend only on \mathbf{J} and not on Φ , because the action variables are constants so that

$$0 = \frac{d}{dt} \mathbf{J} = - \frac{\partial H}{\partial \Phi}. \quad (3.72)$$

But since neither H nor the J_i depend on time, we also conclude

$$\frac{d}{dt} \Phi = \frac{\partial H}{\partial \mathbf{J}} = \vec{\omega} = \text{const}. \quad (3.73)$$

This means the angle variables grow linearly in time with angular frequency ω_i . To clarify the physical meaning of the angle variables and their frequencies, note that the generating function in Eq. (3.69) reduces to the Poincaré invariant, Eq. (3.50), when the former is evaluated along a closed loop winding around the torus (C_1 or C_2 in our notation). More generally, if we wind around C_i a number of times μ_i , then the action increases by J_i . Therefore, one concludes for the change ΔS upon completion of the loop

$$\Delta S(\mu_1, \mu_2) = \mu_1 J_1 + \mu_2 J_2. \quad (3.74)$$

For the same loop, the change in the angle variable is calculated from Eq. (3.71) to be

$$\Delta \Phi_i = \Delta \frac{\partial S}{\partial J_i} = \frac{\partial \Delta S}{\partial J_i} = \mu_i. \quad (3.75)$$

Thus, Φ_i changes by exactly unity for each trip around the torus along a loop equivalent to C_i .

3.6 The Hamilton-Jacobi differential equation

The function S is also called Hamilton’s characteristic function, and it satisfies the *Hamilton-Jacobi differential equation*,

$$H(\mathbf{q}; \partial S / \partial \mathbf{q}) = E, \quad (3.76)$$

which is obtained by substituting Eq. (3.70) for the momenta into the Hamiltonian function and equating it to the energy. The name “characteristic function” stems from the fact that the field lines of $\nabla S = \partial S / \partial \mathbf{q}$ are just the characteristics of the elliptic differential equation (3.76). This equation is merely a by-product of our derivation, but it is often a useful tool in solving dynamical problems. Once one has determined S from this differential equation, the dynamical variables can be deduced from the properties of S as a generating function. This situation is quite similar to the Schrödinger approach to quantum mechanics, where one solves a differential equation to find the wave function, from which the actual physical variables can then be deduced. We will see later in the discussion of the WKB approximation that this analogy is not entirely coincidental.

A special situation arises when the Hamilton-Jacobi equation is separable. Then the solution can be written as a sum,

$$S(\mathbf{q}) = \sum_i S_i(q_i), \quad (3.77)$$

which means that all momenta p_i are going to be functions of the respective conjugate coordinate q_i alone, instead of depending on all other q_j as well. Thus the degrees of freedom are *decoupled* from each other, and we only have to deal with a collection of effectively one-dimensional problems.

As mentioned earlier, the action-angle variables are often introduced under the explicit assumption that one has already found a coordinate system in which the Hamilton-Jacobi differential equation is separable. Then the actions are defined for the resulting one-dimensional systems. As noted by Gutzwiller³⁸, this is an unnecessarily restrictive program, because it precludes us from considering systems for which a separation of variables either cannot be achieved or is too complicated. In particular, the KAM theorem to be discussed later shows that there are nonintegrable (and hence surely non-separable) systems for which one can still find tori in

phase space. Our broader definition of action-angle variables still applies to such KAM tori, even though the proof of path-independence has to be modified when there is no globally conserved quantity K .

3.6.1 Winding numbers

Consider now a special deformation of the “primitive” loops C_i , which consists in following the actual *time evolution* of the starting point in phase space. Recall that we have so far chosen our phase space loops without any relation to the trajectories, the only restriction being that the loop be on the surface of the torus. In order for the real trajectory to provide us with a closed loop, the trajectory itself has to be periodic, of course. Let us say the trajectory closes on itself after μ_i revolutions around the respective primitive loops. It is very useful to characterize such periodic trajectories by their *winding number*

$$w = \frac{\mu_1}{\mu_2} = \frac{\Delta\Phi_1}{\Delta\Phi_2}. \quad (3.78)$$

By definition, the time it takes to evolve to $\Delta\Phi_1$ is the same as that for $\Delta\Phi_2$, because we followed the trajectory; it is simply

$$t = \Delta\Phi_i/\omega_i. \quad (3.79)$$

Using this to replace the $\Delta\Phi_i$ above, we arrive at one of the useful properties of action-angle variables:

$$w = \frac{\omega_1}{\omega_2} = \frac{\partial H/\partial J_1}{\partial H/\partial J_2}. \quad (3.80)$$

Hence we know the winding numbers of all trajectories as soon as we have expressed the Hamiltonian in terms of the action variables. This is often much easier to do than solving the equations of motion.

3.7 Example: Classical mechanics of the circular billiard

At this point it behooves us to illustrate the concepts introduced so far. This will be done by first discussing the simple special case of a point particle moving inside a flat billiard with a circular hard wall of radius R . This circular billiard will serve as the prototype systems in many of the later sections, where we deal with noncircular billiards. First, we will look at the problem in polar coordinates; later we shall show how the same answers are obtained without the use of this special coordinate system.

3.7.1 Polar coordinates

The standard way of treating systems with rotational symmetry is to write down the Hamiltonian in polar coordinates, because then the angle that is canonically conjugate to the conserved angular momentum does not appear in H , it is cyclic. In polar coordinates, the Hamiltonian for the free motion of the particle between collisions with the circular boundary is

$$H = \frac{1}{2m} \left(p_r^2 + \frac{p_\phi^2}{r^2} \right), \quad (3.81)$$

where the conjugate momenta are $p_r = m\dot{r}$ and $p_\phi = mr^2\dot{\phi}$. Since ϕ is cyclic, the angular momentum p_ϕ is conserved. The reflections at the boundary reverse p_r , but do not affect p_ϕ . This does not mean, of course, that the angular velocity stays constant, because r varies with time. As can be seen in Fig. 3.1(a), r is in fact a periodic function of time, independent of whether the orbit closes on itself or not. The inner turning point of the radial motion, i.e. the minimum of r , is obtained by

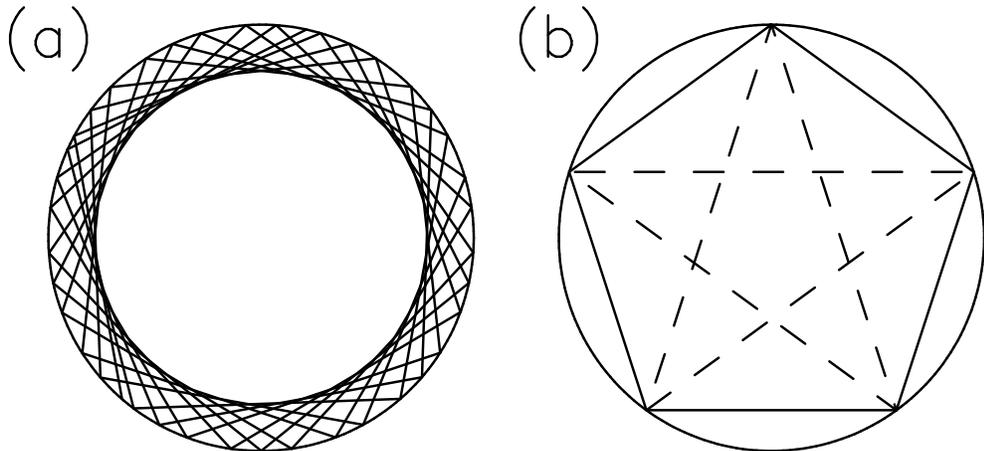


Figure 3.1: (a) A typical quasiperiodic trajectory in the circular billiard. (b) Five-bounce periodic orbits.

equating the Hamiltonian with the total energy E and setting $p_r = 0$:

$$r_0 = \frac{p_\phi}{\sqrt{2mE}}. \quad (3.82)$$

In polar coordinates, all degrees of freedom execute periodic motion in the form of either rotation (ϕ) or libration (r). This would not generally be true in cartesian coordinates, since one could, e.g., start a particle at $x = R$ on an orbit which does not close on itself so that this x never recurs. The Hamilton-Jacobi differential equation, $H(\mathbf{q}; S) = E$, is separable in polar coordinates if one multiplies both sides by r^2 .

Having identified the periodic and decoupled degrees of freedom, it becomes particularly easy to calculate the action variables, because $\oint \mathbf{p} d\mathbf{q}$ decomposes into the sum of two integrals along each coordinate direction. This is possible because the integrand p_i depends only on one coordinate q_i (or none at all, as for p_ϕ); the reason being the separability of the Hamilton-Jacobi equation, cf. Eq. (3.77). We get

$$J_\phi = \int_0^{2\pi} p_\phi d\phi = 2\pi p_\phi \quad (3.83)$$

$$J_r = 2 \int_{r_0}^R p_r dr. \quad (3.84)$$

Both J_ϕ and J_r are manifestly constants of the motion because the dependence on ϕ and r has been integrated out. They could, however, still be functions of energy.

The radial action is obtained by substituting

$$p_r = \sqrt{2mE - \frac{p_\phi^2}{r^2}} \quad (3.85)$$

into Eq. (3.84):

$$J_r = 2 \int_{r_0}^R dr \sqrt{2m \left(E - \frac{p_\phi^2}{2mr^2} \right)}, \quad (3.86)$$

where the factor 2 accounts for the fact that the closed loop runs from r_0 to R and back with opposite sign of p_r . In principle, this can now be solved for E in terms of J_ϕ and J_r . One has thus rewritten the Hamiltonian as a function of the constants of motion alone. If we interpret J_ϕ and J_r as generalized momenta, the conjugate variables Φ_ϕ and Φ_r by definition satisfy Hamilton's equations,

$$\dot{\Phi}_\phi = \frac{\partial H}{\partial J_\phi}, \quad (3.87)$$

$$\dot{\Phi}_r = \frac{\partial H}{\partial J_r}. \quad (3.88)$$

These partial derivatives are just constants, so that the angle variables all have a simple linear time dependence:

$$\Phi_\phi = \omega_\phi t + \Phi_{\phi,0}, \quad (3.89)$$

$$\Phi_r = \omega_r t + \Phi_{r,0}. \quad (3.90)$$

To carry out the solution with these variables, one has to evaluate the radial integral,

$$J_r = 2 \int_{r_0}^R dr \frac{p_\phi}{r} \sqrt{\frac{r^2}{r_0^2} - 1} \quad (3.91)$$

$$= 2p_\phi \left[\sqrt{\frac{R^2}{r_0^2} - 1} + \arcsin \frac{r_0}{R} - \frac{\pi}{2} \right]. \quad (3.92)$$

We cannot solve this directly for E (which is contained in r_0) to obtain the derivatives of the energy with respect to the action variables needed in Eq. (3.87). However, some progress can be made by implicit differentiation² of Eq. (3.92), considering J_r and J_ϕ as the independent variables. Using the definition of r_0 and the abbreviation

$$\beta \equiv 2mER^2, \quad (3.93)$$

we can rewrite Eq. (3.92) as

$$\sqrt{\frac{\beta}{p_\phi^2} - 1} + \arcsin \frac{p_\phi}{\sqrt{\beta}} = \frac{\pi}{2} + \frac{J_r}{2p_\phi}. \quad (3.94)$$

Take the derivative with respect to J_r , which yields

$$\frac{1}{2p_\phi^2 \sqrt{\beta/p_\phi^2 - 1}} \frac{\partial \beta}{\partial J_r} + \frac{1}{\sqrt{1 - p_\phi^2/\beta}} \left[-\frac{p_\phi}{2\beta^{3/2}} \frac{\partial \beta}{\partial J_r} \right] = \frac{1}{2p_\phi} \quad (3.95)$$

$$\Rightarrow \frac{1}{2} \frac{\partial \beta}{\partial J_r} \left[\frac{1}{p_\phi \sqrt{\beta - p_\phi^2}} - \frac{p_\phi}{\beta \sqrt{\beta - p_\phi^2}} \right] = \frac{1}{2p_\phi} \quad (3.96)$$

$$\Rightarrow \frac{\partial \beta}{\partial J_r} = \frac{\beta}{\sqrt{\beta - p_\phi^2}}. \quad (3.97)$$

²This was pointed out to me by Henrik Bruus

Now we similarly differentiate Eq. (3.94) with respect to p_ϕ :

$$\frac{1}{2\sqrt{\beta/p_\phi^2 - 1}} \left[\frac{1}{p_\phi^2} \frac{\partial\beta}{\partial p_\phi} - \frac{2\beta}{p_\phi^3} \right] + \quad (3.98)$$

$$\frac{1}{\sqrt{1 - p_\phi^2/\beta}} \left[\frac{1}{\sqrt{\beta}} - \frac{p_\phi}{2\beta^{3/2}} \frac{\partial\beta}{\partial p_\phi} \right] = -\frac{J_r}{2p_\phi^2} \quad (3.99)$$

$$\Rightarrow \frac{\partial\beta}{\partial p_\phi} = 2p_\phi\beta \left[\frac{1}{p_\phi^2} - \frac{J_r}{2p_\phi^2\sqrt{\beta - p_\phi^2}} \right]. \quad (3.100)$$

This provides us with enough information to determine the winding number, cf. Eq. (3.78), which is given by the frequency ratio

$$w = \frac{\omega_\phi}{\omega_r} = \frac{\partial H/\partial J_\phi}{\partial H/\partial J_r} = \frac{\partial\beta/\partial J_\phi}{\partial\beta/\partial J_r} = \frac{\partial\beta/\partial p_\phi}{2\pi\partial\beta/\partial J_r} \quad (3.101)$$

$$= \frac{1}{\pi} \left[\sqrt{\frac{\beta}{p_\phi^2} - 1} - \frac{J_r}{2p_\phi} \right]. \quad (3.102)$$

Here we still have the problem that $\beta = 2mER^2$ has to be known as a function of J_r and J_ϕ (or p_ϕ) if we want to express w in terms of these actions. Even without this knowledge, we can however illuminate the relation between the winding number and the angle of incidence χ with respect to the normal at the boundary. In the expression for w , we first eliminate J_r by means of Eq. (3.94) to get

$$w = \frac{1}{\pi} \left[\frac{\pi}{2} - \arcsin \frac{p_\phi}{\sqrt{\beta}} \right]. \quad (3.103)$$

If the linear momentum is $p = \sqrt{2mE}$, then the angular momentum at each reflection is

$$p_\phi = |\mathbf{R} \times \mathbf{p}| = Rp \sin \chi \quad (3.104)$$

$$= \sqrt{\beta} \sin \chi. \quad (3.105)$$

We could also derive the last equality purely from geometry, using the definition of the inner turning point, Eq. (3.82), and the fact that the angle of incidence at the boundary is given by

$$\sin \chi = r_0/R = p_\phi/\sqrt{\beta}. \quad (3.106)$$

This allows us to eliminate β from the previous expression, so that we can write

$$\pi w = \frac{\pi}{2} - \chi, \quad \text{or} \quad \cos \pi w = \sin \chi. \quad (3.107)$$

The first equality implies $0 \leq w \leq 1$, because $-\pi/2 \leq \chi \leq \pi/2$. If w is rational, $w = l/k$, then the orbit closes on itself after k reflections, having wound around the azimuthal direction l times. As an example, Fig. 3.1(b) shows closed orbits with $k = 5$. From the figure, it can also be seen that $w = 2/5$ and $w = 3/5$ are simply time-reversed versions of each other, which reflects the general relation implied by Eq. (3.107),

$$(w \rightarrow 1 - w) \Rightarrow (\sin \chi \rightarrow -\sin \chi). \quad (3.108)$$

In the circle, there is an infinite family of closed orbits for each rational w , obtained by rotating all reflection points by an arbitrary angle. The energy E has no effect

on the geometry of the trajectories, it only determines how fast they are traced out. This is a peculiarity of the hard-wall boundary.

Most trajectories, namely the ones with irrational winding number, never close on themselves. As shown in Fig. 3.1(a), they exhibit instead a *caustic*, consisting of the circle of radius r_0 to which all rays are tangent between any two reflections, as follows from the radial equation. The caustic is the line at which the real-space projection of the phase-space torus becomes singular, in the sense that a flat measure on the surface of the torus will give rise to a divergent density on the caustic (an equivalent statement is that the probability per unit time of finding a particle at its classical turning points diverges, if the potential is smooth).

Even when the orbit does not close on itself, it is possible to associate the irrational winding number with the number of reflections per revolution around the billiard. The general relation is

$$w = \frac{1}{2\pi} \lim_{i \rightarrow \infty} \frac{\phi_i - \phi_0}{i}, \quad (3.109)$$

where ϕ_i is the position where the i -th reflection occurs, *not* taken modulo 2π . This definition has the advantage of being applicable even in systems where one cannot determine the action-angle variables.

3.7.2 Actions from the caustic

When we introduced the action-angle variables, it was a deliberate choice not to rely on the separability of the Hamilton-Jacoby equation but instead to use only the torus topology resulting from the existence of N constants of the motion. It is illuminating to show for the circle that the actions can be calculated even if one does not use polar coordinates, just based on the fact that the real space trajectories in general create a caustic of known shape.

This is the procedure followed by Keller and Rubinow⁹⁹. The main subject of their 1960 paper is the semiclassical quantization of various billiard systems (we will return to this question below), and a crucial part of that work is the calculation of the action variables.

For the circle, the caustic is also a circle whose radius is related to the angle of incidence χ by

$$r_0 = R \sin \chi. \quad (3.110)$$

In calculating $\oint \mathbf{p} d\mathbf{q}$, Keller's and Rubinow's approach is to choose all the loops C_i such that \mathbf{p} and $d\mathbf{q}$ are either parallel, antiparallel or perpendicular. I.e., the loop always follows some ray, except for segments where $\mathbf{p} d\mathbf{q} = 0$ which hence make no contribution to the integral. For loop segments where \mathbf{p} and $d\mathbf{q}$ are parallel, the actions reduce to $\sqrt{2mE}$ times the length of the respective loop, since the magnitude $p = \sqrt{2mE}$ is constant. The same holds for antiparallel \mathbf{p} and $d\mathbf{q}$, except for a negative sign in front of their contribution to the integral.

In the integration, one has to keep track of which sheet of \mathbf{p} is being used. To characterize these sheets, note that at each point \mathbf{q} in the annulus between r_0 and R , there are four possible values of \mathbf{p} (always keeping in mind that all rays are tangent to the caustic): one can draw a ray through \mathbf{q} that encircles the caustic in the counterclockwise sense (positive angular momentum), and either is directed away from the caustic or toward it; the analogous two possibilities exist for clockwise sense of rotation. We can discard the clockwise trajectories because of time-reversal invariance, so that only two sheets of \mathbf{p} are left. The corresponding sheets of S

will be denoted by S_{CB} and S_{BC} . The former yields for every point \mathbf{q} the ray that passes through \mathbf{q} in the direction from the caustic to the boundary, and vice versa for S_{BC} , always in the counterclockwise sense. Recall that $\mathbf{p} = \nabla S(\mathbf{q})$, so we have to consider the different sheets of S . The first loop that is considered consists of the caustic itself, taken entirely on a single sheet, e.g. S_{CB} . This can be considered as the limiting case of the contour shown as the sawtooth line in Fig. 3.2 (a), and the resulting action is simply

$$J_\phi = \sqrt{2mE} 2\pi r_0 = 2\pi R p \sin \chi = 2\pi p_\phi, \quad (3.111)$$

in agreement with Eq. (3.83). Thus we have re-derived the azimuthal action.

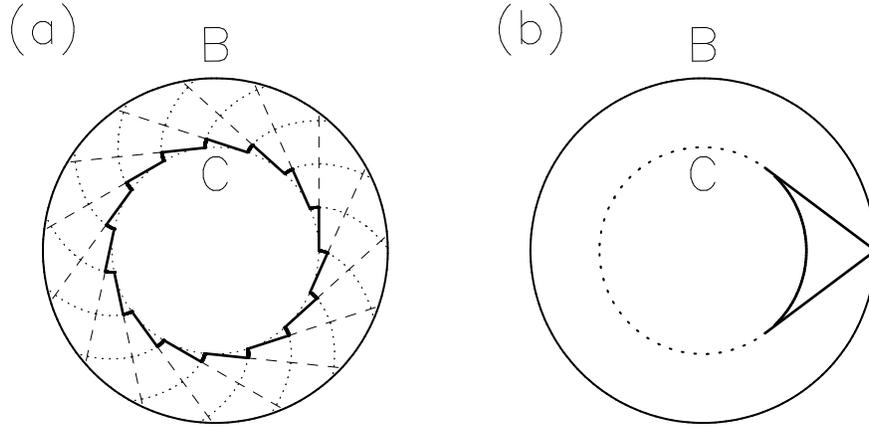


Figure 3.2: Knowing the shape of the boundary (B) and caustic (C) in the circle, one can obtain the azimuthal (a) and radial (b) action. The integration loops (indicated as bold lines) are either parallel to the rays [dashed lines in (a)] or follow the contour lines of S [dotted lines in (a)]. The resulting saw-tooth line in (a) never crosses the caustic, but can be drawn infinitesimally close to it, so that its length becomes equal to that of C . In (b), this limit has already been performed, but one needs to cross (C) and (B) once each.

To calculate the radial action, we choose the loop shown in Fig. 3.2 (b), which has to involve both sheets of \mathbf{p} because clearly two different directions of \mathbf{p} intersect at the boundary point. If we start at this point in the counterclockwise sense, the first segment of the loop uses S_{BC} up to the caustic. The next segment is the shorter of the two arcs between the points of tangency at the caustic. For this we can remain on the sheet S_{BC} , but the loop is antiparallel to the rays so that this contribution has to be subtracted from the first segment. Alternatively, one could use the longer arc with S_{BC} , and the contribution would be positive since the loop runs parallel to the ray directions. However, the results differ by J_r because one would in addition have encircled the azimuthal loop by going along the longer arc. Finally, we have to cross over to the sheet S_{CB} at the second point of tangency to the caustic, in order to return to the starting point on the boundary. The sum of all three contributions is

$$J_r = p \left[2\sqrt{R^2 - r_0^2} - 2 \left(\frac{\pi}{2} - \chi \right) \right] \quad (3.112)$$

$$= 2pr_0 \left[\sqrt{\frac{R^2}{r_0^2} - 1} - \left(\frac{\pi}{2} - \arcsin \frac{r_0}{R} \right) \right], \quad (3.113)$$

where we have used $\sin \chi = r_0/R$. But this is identical to Eq. (3.92) because $p_\phi = pr_0$.

The result of this more geometric calculation is thus the same as that of the previous subsection, and there was no need to use polar coordinates explicitly.

3.8 Example: Elliptical billiard

The billiard with elliptical boundary seems at first sight to be reducible to the circular billiard, because we can transform the boundary given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (3.114)$$

into a circle by the linear substitution $x' = x/a$, $y' = y/b$, which at the same time maps straight line trajectories onto straight lines. However, this transformation does not map the specular reflection law in the ellipse onto specular reflection in the circle. I.e., a real trajectory in the ellipse will, after the above contraction and dilation, not be geometrically identical to a real trajectory of the circle assuming specular reflection. This can be seen for example by taking a trajectory in the circle which reflects at 45° angle of incidence from the point with polar angle $\phi = 45^\circ$. If one contracts the y -axis to create an ellipse, the incoming and outgoing rays remain horizontal and vertical, respectively, whereas the normal no longer bisects the angle between incident and reflected ray.

It is in fact a problem of some recent interest⁷⁰ to discuss conservation laws in the ellipse billiard and its three-dimensional generalization, because of its relation to Poinclet's theorem of projective geometry⁷¹. This theorem states that if a trajectory closes on itself after p reflections, then so do all others that are tangent to the same caustic. From the point of view of our Hamiltonian mechanics, it is evident that all trajectories on the same torus have the same winding number, so Poinclet's theorem comes as no surprise. However, it remains for us to show that a trajectory in the ellipse does indeed move on a torus, i.e. that the billiard is integrable.

3.8.1 Integrability of the ellipse

The conserved quantity we are looking for is the product of the angular momenta L_1 , L_2 with respect to the two foci F_1 , F_2 . This is stated without proof in Ref.⁷², and we shall verify this result here. The crucial property of the ellipse billiard is that the normal at any point A on the boundary exactly bisects the angle $\angle F_1 A F_2$ subtended by the two foci.

To show this, write the equation for the elliptical boundary in cartesian coordinates,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (3.115)$$

and find the normal by taking the gradient of this,

$$\mathbf{n} \propto \begin{pmatrix} x/a^2 \\ y/b^2 \end{pmatrix}. \quad (3.116)$$

The unit vectors pointing to $A = (x, y)$ on the boundary from the foci are

$$\mathbf{d}_1 = \frac{1}{\sqrt{(x-c)^2 + y^2}} \begin{pmatrix} x-c \\ y \end{pmatrix}, \quad (3.117)$$

$$\mathbf{d}_2 = \frac{1}{\sqrt{(x+c)^2 + y^2}} \begin{pmatrix} x+c \\ y \end{pmatrix}, \quad (3.118)$$

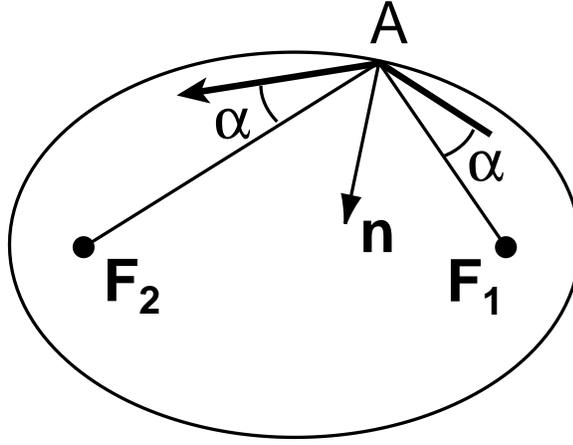


Figure 3.3: The normal bisects the angle between incoming and outgoing ray in the ellipse. Since it also bisects the angle between the focal rays, the two angles denoted by α are the same.

where $c = \sqrt{a^2 - b^2}$ is the distance of the foci from the origin, assuming without loss of generality $a > b$. Now we form the ratio

$$\frac{\mathbf{n} \cdot \mathbf{d}_1}{\mathbf{n} \cdot \mathbf{d}_2} = \frac{(x-c)x/a^2 + y/b^2}{(x+c)x/a^2 + y/b^2} \sqrt{\frac{(x+c)^2 + y^2}{(x-c)^2 + y^2}}. \quad (3.119)$$

Here one can eliminate y^2 using $y^2 = b^2 - x^2b^2/a^2$, to obtain

$$\frac{\mathbf{n} \cdot \mathbf{d}_1}{\mathbf{n} \cdot \mathbf{d}_2} = \frac{1 - cx/a^2}{1 + cx/a^2} \sqrt{\frac{x^2(1 - b^2/a^2) + a^2 + 2xc}{x^2(1 - b^2/a^2) + a^2 - 2xc}} \quad (3.120)$$

$$= \sqrt{\frac{(a^4 + x^2c^2 - 2xca^2)(x^2c^2 + a^4 + 2xca^2)}{(a^4 + x^2c^2 + 2xca^2)(x^2c^2 + a^4 - 2xca^2)}} \quad (3.121)$$

$$= 1, \quad (3.122)$$

where we used the definition of c and multiplied through by a^3 . This result means that \mathbf{d}_1 and \mathbf{d}_2 form the same angle with the normal, which is what we wanted to show. Now it is a simple matter to convince ourselves that the angle α_1 between incoming ray and \mathbf{d}_1 is the same as the angle α_2 between outgoing ray and \mathbf{d}_2 , because the angles of in-and outgoing rays with the normal have to be equal, too. This is illustrated in Fig. (3.3). We denote

$$\alpha_1 = \alpha_2 = \alpha. \quad (3.123)$$

If χ is the angle of incidence, then the angle between incident ray and \mathbf{d}_2 is $2\chi - \alpha$, and the same is true for the reflected ray and \mathbf{d}_1 . Let r_i be the distance from A to F_i , then the product of angular momenta for the incoming ray is

$$L_{12} \equiv L_1 L_2 = p^2 r_1 r_2 \sin \alpha \sin(2\chi - \alpha), \quad (3.124)$$

where p is the constant magnitude of the linear momentum. But the same is found if we consider the outgoing ray. Therefore, this product is unchanged by the collision with the boundary, and hence it is a constant of the motion because the angular momenta are individually conserved between bounces.

Inspection of Eq. (3.124) reveals that the phase space of the ellipse is quite different from that of the circle. Depending on the sign of L_{12} , the real space trajectories are either so-called whispering gallery orbits ($L_{12} > 0$), or of the “bouncing-ball” type ($L_{12} < 0$). The latter always cross the x -axis in the interval $[-c, c]$, i.e.

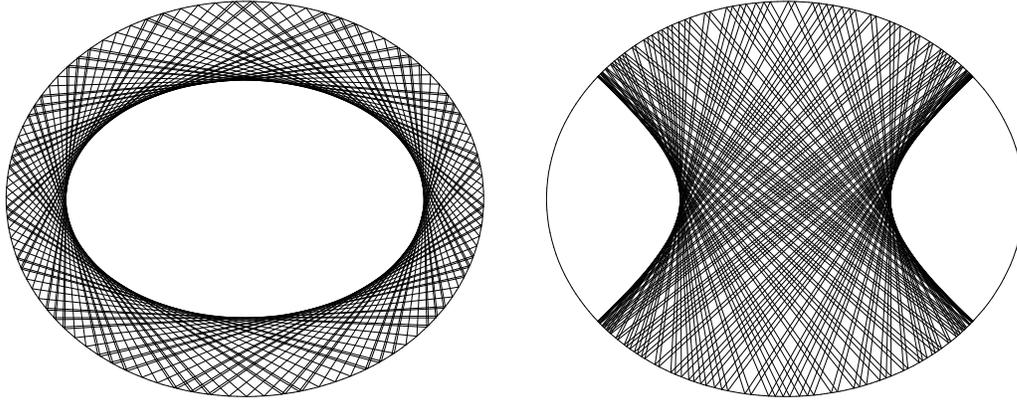


Figure 3.4: The two types of trajectories in the ellipse are whispering-gallery orbits (left) and bouncing-ball orbits (right). The caustics can be identified clearly as ellipses (left) and hyperbolae (right), to which the rays are tangency between any two reflections. For bouncing ball orbits, this tangency may require a ray segment and the hyperbola to be continued to the outside of the billiard.

between the focal points; the former cross the x axis outside that range. These two classes of trajectories are shown in Fig. 3.4. To see how the bouncing-ball orbits arise, note that the sign of L_{12} can be negative if either α or $2\chi - \alpha$ is negative. Assume without loss of generality $\alpha < 0$ (the other case can be treated by reversing time). This corresponds to trajectories that impinge on the boundary from inside the triangle F_1AF_2 , and the reflected ray will be inside the same triangle so that it has to cross the x -axis between F_1 and F_2 . At the next collision, the same scenario applies because L_{12} is conserved. The whispering gallery orbits are forbidden from crossing the x -axis between the foci, by similar arguments.

The real-space projections of the invariant torus is therefore a deformed annulus for $L_{12} > 0$, but not for $L_{12} < 0$. In the latter case the forbidden region is located around the highest-curvature points, because precisely at these points the angle between \mathbf{d}_1 and \mathbf{d}_2 shrinks to zero, making it impossible to have a trajectory impinge from “within” this degenerate triangle unless $\alpha = 0$ (which in turn implies $L_{12} = 0$).

It can be shown⁷³ that the caustics are confocal ellipses for $L_{12} > 0$ and confocal hyperbolae for $L_{12} < 0$. An important application of this knowledge is the semiclassical quantization procedure employed by Keller and Rubinow⁹⁹, where the caustics figure prominently. In view of the fact that we intend to modify that quantization procedure such that a knowledge of the caustics is not required, we will at this point only mention that the actions for a WG orbit can be calculated along the same types of loops as in the circle, with the integrations along the elliptical caustic giving rise to elliptic integrals of the second kind.

3.8.2 An expression for the angle of incidence

Let us rewrite Eq. (3.124) as

$$L_{12} = p^2 r_1 r_2 \sin\left(\chi - \frac{\Delta}{2}\right) \sin\left(\chi + \frac{\Delta}{2}\right), \quad (3.125)$$

where $\Delta = \angle F_1AF_2$ is the angle between the two focal rays through the point A where the reflection occurs. It is straightforward to bring this into the form

$$L_{12} = r_1 r_2 \left(\sin^2 \chi - \sin^2 \frac{\Delta}{2} \right) \quad (3.126)$$

$$\Rightarrow \sin^2 \chi = \sin^2 \frac{\Delta}{2} + \frac{L_{12}}{p^2 r_1 r_2}. \quad (3.127)$$

Now one can also express Δ in terms of the r_i , because the triangle $F_1 A F_2$ satisfies

$$4c^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \Delta. \quad (3.128)$$

This leads to

$$\cos \Delta = \frac{r_1^2 + r_2^2 - 4c^2}{2r_1 r_2} \quad (3.129)$$

$$= -\frac{2c^2}{r_1 r_2} + \frac{(r_1 + r_2)^2 - 2r_1 r_2}{2r_1 r_2} \quad (3.130)$$

$$= -\frac{4c^2}{2r_1 r_2} + \frac{4a^2}{2r_1 r_2} - 1 \quad (3.131)$$

$$= \frac{2b^2}{r_1 r_2} - 1, \quad (3.132)$$

where we have used $r_1 + r_2 = 2a$. This allows us to write

$$\sin^2 \frac{\Delta}{2} = \frac{1}{2} (1 - \cos \Delta) = 1 - \frac{b^2}{r_1 r_2}, \quad (3.133)$$

which we insert into Eq. (3.127) with the result

$$\sin^2 \chi = 1 + \frac{L_{12}/p^2 - b^2}{r_1 r_2}. \quad (3.134)$$

This result can be written in a different way by making use of the striking fact that the curvature of the ellipse satisfies the relation

$$\kappa = \frac{ab}{(r_1 r_2)^{3/2}}. \quad (3.135)$$

With this, we finally obtain

$$\sin \chi = \sqrt{1 + \frac{(L_{12}/p^2 - b^2)}{(ab)^{2/3}} \kappa^{2/3}}. \quad (3.136)$$

In this fashion we have expressed the dynamic variable $\sin \chi$ purely in terms of the curvature at the bounce point (and of course the constant of motion). This result is central to our approach for finding the actions, as will be discussed now. It reduces to the correct result in the circle, because then $L_{12} = L_1 L_2 = L_1^2$, $a = b = r_1 = r_2 = R$, so that $p R \sin \chi = L_1$.

3.8.3 Actions from the curvature

As advertised above, we at this point calculate the action variables for the ellipse assuming nothing about the detailed shape of the caustic. This will be done only for the whispering gallery orbits, because they are the ones of special interest later. The method will be applicable to general convex billiards as a way of finding approximate invariants, because all we assume here is that the spatial projection of the torus is two-sheeted and forms a deformed annulus, and that we possess an expression for $\sin \chi$ at the boundary as a function of ϕ ,

$$\sin \chi = \sqrt{1 - \left[\frac{\kappa(\phi)}{K} \right]^{2/3}}. \quad (3.137)$$

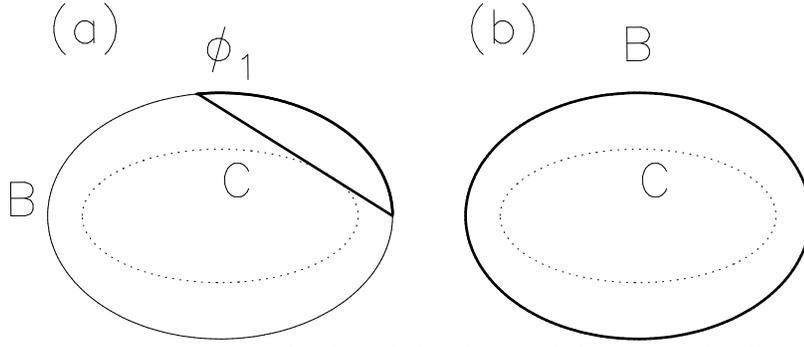


Figure 3.5: Integration contours for the radial and azimuthal action in the ellipse.

Here, K is an abbreviation for the constants appearing in Eq. (3.136). As will be shown later (chapter 5), an expression of precisely this form can in fact serve as an approximate description of the intermediate-time dynamics even in nonintegrable convex billiards.

The classical torus on which the trajectory moves is completely specified by the energy and the constant K in Eq. (3.137). The loops we choose for the action integrals are the following: Let C_1 be a closed curve infinitesimally close to the boundary, and C_2 the loop made up of a ray intersecting the boundary at $\phi_0 = 0$ and $\phi_1 \neq 0$ together with the short segment of the boundary between ϕ_0 and ϕ_1 (see Fig. 3.5). As we trace out these curves, we have to keep track on which sheet of the projection of S onto the spatial coordinate plane we are, either S_{BC} or S_{CB} . Here, we use the same terminology as in subsection 3.7.2, i.e., S_{BC} increases if one follows its rays from the boundary (B) to the caustic (C), and S_{CB} has the opposite behavior. For definiteness it is assumed from now on that we start on the boundary at $\phi_0 = 0$.

Along C_1 we can stay on one sheet, say S_{CB} , and the line element $d\mathbf{l}$ is parallel to the boundary tangent vector \mathbf{t} , so that $\nabla S_{CB} \cdot \mathbf{t} = \sqrt{2mE} \sin \chi$. At this point we make use of Eq. (3.137) which provides an explicit expression for $p(\phi)$ in terms of the unknown constant K . Thus the azimuthal action is

$$J_\phi = \sqrt{2mE} \int_0^{2\pi} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - \left[\frac{\kappa(\phi)}{K} \right]^{2/3}}. \quad (3.138)$$

Here we used the arc length element $dl = \sqrt{r^2 + \dot{r}^2} d\phi$ in polar coordinates.

For the loop C_2 , we go from the boundary to the caustic along the ray segment on S_{BC} , cross to S_{CB} and continue straight on to the boundary point at ϕ_1 . The integral along this segment is just the momentum times the length L of the straight line between ϕ_1 and ϕ_0 . To return to ϕ_0 , we cross back to S_{BC} at ϕ_1 and integrate along the arc, obtaining a contribution analogous to Eq. (3.138). However, the integration extends from ϕ_1 down to ϕ_0 ,

$$\sqrt{2mE} \int_{\phi_1}^{\phi_0} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - \left[\frac{\kappa(\phi)}{K} \right]^{2/3}}. \quad (3.139)$$

This produces a negative sign after we make ϕ_0 the lower boundary of the integral, so that the result for the radial action is

$$J_r = \sqrt{2mE} L(K) - \sqrt{2mE} \int_{\phi_0}^{\phi_1(K)} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - \left[\frac{\kappa(\phi)}{K} \right]^{2/3}}, \quad (3.140)$$

where we wrote an explicit dependence of $\phi_1(K)$ and $L(K)$ on the constant K . The reason is that K determines $\sin \chi$ at ϕ_0 through

$$p(0) = \sqrt{1 - \left[\frac{\kappa(0)}{K} \right]^{2/3}}, \quad (3.141)$$

which in turn fixes the point ϕ_1 at which the ray intersects the boundary, and consequently its length L .

The semiclassical quantization to be introduced later will rely mainly on Eqs. (3.138) and (3.140). These are the correct actions in the ellipse. For shapes other than the ellipse, similar action integrals might be surmised to exist. However, that would necessarily require integrability of the billiard, i.e. the existence of a quantity analogous to L_{12} . This leads us to the question of whether or not other integrable billiards exist.

3.8.4 How exceptional is the ellipse billiard ?

The only smooth and convex deformation of the circle for which the classical dynamics still possesses two constants of motion is the ellipse. This is a claim with far-reaching consequences, as we will see later. In fact, the proof of this statement, entitled ‘‘Smooth convex planar domains for which the billiard ball map is integrable are ellipses’’ was given only recently⁷⁴.

The Hamilton-Jacobi differential equation, Eq. (3.76), for the hard-wall ellipse billiard is separable in elliptical cylinder coordinates as we show below; one can also separate it in the so-called Jacobi variables which are derived from the elliptical ones⁷⁰. Recalling that polar coordinates allow to decouple the equations of motion for *any* rotationally invariant potential $V(r)$, it is natural to expect that the same should hold in elliptical coordinates for any V whose equipotential lines are also coordinate lines.

Surprisingly, however, the Hamilton-Jacobi equation for such a V is *not* separable in elliptical coordinates for any potential other than hard walls. The analogous statement holds for ellipsoidal billiards, too⁷⁵. To elaborate on this, we derive the Hamilton-Jacobi equation by starting from the Lagrangian. The elliptical coordinates are defined through

$$x = c \cosh \rho \cos \phi, \quad y = c \sinh \rho \sin \phi \quad (3.142)$$

where c is a positive constant corresponding to half the distance between the foci of the ellipse. The boundary is assumed for definiteness to be parametrized by $\rho = 1$. The Lagrangian in these coordinates is

$$L = \frac{1}{2} mc^2 \left[(\dot{\rho}^2 + \dot{\phi}^2) (\sinh^2 \rho + \sin^2 \phi) \right] - V(\rho, \phi), \quad (3.143)$$

and the resulting Hamiltonian is

$$H = \frac{1}{2mc^2 (\sinh^2 \rho + \sin^2 \phi)} (p_\rho^2 + p_\phi^2) + V(\rho, \phi). \quad (3.144)$$

The Hamilton-Jacobi equation for the characteristic function $W(\rho, \phi)$ is

$$\frac{1}{2mc^2 (\sinh^2 \rho + \sin^2 \phi)} \left[\left(\frac{\partial W}{\partial \rho} \right)^2 + \left(\frac{\partial W}{\partial \phi} \right)^2 \right] + V(\rho, \phi) = E. \quad (3.145)$$

This could be solved by the separation ansatz

$$W(\rho, \phi) = R(\rho) + F(\phi) \quad (3.146)$$

if one could multiply through by $2mc^2 (\sinh^2 \rho + \sin^2 \phi)$ and arrive at an equation where each term contains functions of either ρ or ϕ separately. However this is not possible if we assume the equipotential lines to coincide with the curves of $\rho = \text{const}$, because then we will be left with a cross term $\sin^2 \phi V(\rho)$.

The potential V for a hard wall billiard is $V = \infty$ for $\rho > 1$ and $V = 0$ otherwise. The fact that the only nonzero value of V is infinity can be used to write the potential in a form that makes a separation of variables possible. In close analogy to⁷⁰ we define the hard wall as a limit of the smooth potential

$$V(\rho, \phi) = \frac{V_\epsilon(\rho)}{2mc^2 (\sinh^2 \rho + \sin^2 \phi)} \quad (3.147)$$

where $V_\epsilon(\rho)$ is any differentiable function of ρ that satisfies the requirements

$$V_\epsilon(\rho) = \begin{cases} \frac{V_0}{\epsilon} & (\rho > 1 + \epsilon) \\ 0 & (\rho < 1 - \epsilon) \end{cases} \quad (3.148)$$

with some positive constant V_0 that serves only to guarantee that V blows up when the limit $\epsilon \rightarrow 0$ is taken. At the same time the transition region where the rapid increase occurs, $]1 - \epsilon, 1 + \epsilon[$, shrinks to a point. The resulting potential will thus be the hard wall, but we must perform the separation of variables at finite ϵ .

Inserting Eq. (3.148) in Eq. (3.145), we arrive at

$$\left(\frac{\partial W}{\partial \rho}\right)^2 + \left(\frac{\partial W}{\partial \phi}\right)^2 + V_\epsilon(\rho) = 2mc^2 E (\sinh^2 \rho + \sin^2 \phi), \quad (3.149)$$

which is clearly separable.

We have thus seen that the ellipse with hard walls is indeed a unique point in a large space of possible shapes and potentials, so that integrability and hence a rigorous generalization of Eq. (3.136) is not achieved in generic systems of the types discussed.

3.9 The Poincaré map

The discussion of the ellipse billiard immediately leads to the question of how the absence of integrability manifests itself and how it can be characterized. In order to address these issues, we have to introduce one of the principal tools of chaos theory, namely the Poincaré surface of section. It is used in fields as diverse as fluid dynamics⁷⁶ and astronomy³⁸.

Consider again a two-dimensional system with energy conservation, so that the phase space motion is restricted to a three-dimensional subspace. Irrespective of whether there is a second constant of motion or not, one therefore can express one phase space variable as a function of the other three, as was done in Eq. (3.54) with the function \tilde{p}_2 , which was then converted to a function only of \mathbf{q}_1 and \mathbf{q}_2 with the help of the second conservation law. Since this latter step cannot be taken in general, we could reduce the number of independent dynamical variables artificially by considering q_2 as a parameter. In particular, we could fix q_2 in such a way that trajectories repeatedly cross the plane $q_2 = q_2^{SOS}$, and then record the sequence

of tuples (q_1, p_1) generated by the trajectory upon successive crossings. We will also agree to record only those intersections with this *Poincaré surface of section* (SOS) for which q_2 crosses q_2^{SOS} in the direction of increasing q_2 . In this way the continuous-time dynamics gives rise to a discrete map of the $q_1 - p_1$ plane onto itself, the Poincaré map. Given a point (q_1^0, p_1^0) on the SOS, the map takes this point as the starting point of a trajectory and yields by definition the next intersection (q_1^1, p_1^1) of the trajectory with the plane q_2^{SOS} in the increasing direction.

3.9.1 Area-preservation

An important property of Poincaré sections for Hamiltonian systems is that their resulting mappings are *area-preserving*, i.e. a loop on the SOS is mapped in the next iteration onto a new loop which encloses the same area. This statement is not obvious, because not all points on such a loop take the same *time* Δt to cross the plane q_2^{SOS} again. In other words, if we start all trajectories on the initial loop at $t = 0$, and wait until one of them crosses the SOS again at time Δt , then some other trajectories may already have crossed the SOS while others have yet to reach it. The new loop in the SOS is therefore not taken at equal time t , and consequently we cannot immediately use the Poincaré invariant of Eq. (3.11) to show area preservation in the Poincaré map.

The invariant named after Poincaré, $\oint_C \mathbf{p}d\mathbf{q}$, exists for all Hamiltonian systems, but it is required that the loop C be taken at equal times. For *integrable* systems it was shown that such a loop can be deformed on the surface of the phase-space torus without effect on the integral. There is an even more general invariance property, the Poincaré-Cartan theorem, which concerns the *extended* phase space where time has been added as an extra coordinate, its conjugate momentum being the energy⁷⁶. In this space a loop C generates a tube along the time axis, with varying cross-sectional shape. One can then show that for conservative systems, the loop in the Poincaré invariant can be replaced by another one that is *not* taken at equal time, and the result is the same for all such loops as long as they encircle the same tube of trajectories. This is sufficient to show area-preservation in the context of the SOS. Instead of using this theorem, we give a direct proof for the case of $N = 2$ degrees of freedom. The key point is to consider the parameter q_2 as a new “time” variable. This can be done for each interval in which q_2 is a monotonic function of the real time t , so that one can invert $q_2(t)$ and substitute this into q_1, p_1 . Then one has

$$\frac{dq_1}{dq_2} = \frac{\dot{q}_1}{\dot{q}_2} = \frac{\partial H / \partial p_1}{\partial H / \partial p_2} = \frac{\partial}{\partial p_1}(-\tilde{p}_2), \quad (3.150)$$

$$\frac{dp_1}{dq_2} = \frac{\dot{p}_1}{\dot{q}_2} = \frac{-\partial H / \partial q_1}{\partial H / \partial p_2} = -\frac{\partial}{\partial q_1}(-\tilde{p}_2), \quad (3.151)$$

where we first used Hamilton’s equations for \dot{q}_i, \dot{p}_1 , and then applied Eqs. (3.56) and (3.57). This result has the interpretation that $-\tilde{p}_2(q_1, p_1; q_2)$ acts as a new Hamiltonian governing the dynamics in the *reduced* phase space q_1, p_1 , with q_2 as the new time. Note that the new Hamiltonian is explicitly time-dependent, unlike the original H . Even for time-dependent Hamiltonians, however, the first Poincaré invariant exists, as we showed. Therefore, a loop in the SOS at q_2^{SOS} encloses the same area for any q_2 , and in particular will have the same area whenever q_2 returns to the SOS. Therefore, the Poincaré map is area-preserving.

For billiard maps, area-preservation can also be proved directly, without recourse to the Hamiltonian⁶⁵. The idea is to derive a Legendre transformation of

the generating function that is needed in theorem 3.3.1.1. It is given simply by the action of the straight line trajectories between three bounces, which in turn is just the path length.

In the following sections, the classical dynamics of billiard systems will be discussed mainly from the point of view of the Poincaré map, and its area-preserving property will in fact emerge as an important ingredient in the transition from integrable to chaotic dynamics.

Chapter 4

Billiard maps and the transition to chaos

In this chapter, we begin the discussion of billiard systems, and explore how a shape deformation introduces chaotic trajectories into the dynamics. A self-contained treatment of the classical mechanics in various billiard systems, including the circle and the ellipse, has been given in Ref.⁷².

4.1 Birkhoff coordinates

To cast the billiard motion into the form of a map, we employ the so-called Birkhoff coordinates⁷⁷: For each collision of a straight-line trajectory with the hard wall, one records the arc length s along the boundary at which the reflection occurs, and the value of the tangential momentum p_t of the incident particle. This is done for a number N of consecutive reflections, and each pair of these values is labeled in ascending order by an integer n . The billiard mapping M brings us from one member of such a sequence to the next,

$$(s_{n+1}, p_{t,n+1}) = M(s_n, p_{t,n}). \quad (4.1)$$

This notation means that M is a function of s_n and $p_{t,n}$. To show that M is area-preserving, we first have to convince ourselves that these variables form a conjugate pair. If the billiard shape is given in cartesian coordinates as

$$\mathbf{r}_B(s) \equiv \begin{pmatrix} x_B(s) \\ y_B(s) \end{pmatrix}, \quad (4.2)$$

then any point \mathbf{r} can be uniquely specified by two coordinates s and a that are chosen according to the equation

$$\mathbf{r}(s, a) = \begin{pmatrix} x(s, a) \\ y(s, a) \end{pmatrix} = a \mathbf{r}_B(s). \quad (4.3)$$

In other words, a is a factor by which one scales the billiard shape such that the boundary goes through \mathbf{r} . Then the kinetic energy of the particle is

$$T = \frac{1}{2}m \left\{ \left(\frac{\partial x}{\partial s} \dot{s} + \frac{\partial x}{\partial a} \dot{a} \right)^2 + \left(\frac{\partial y}{\partial s} \dot{s} + \frac{\partial y}{\partial a} \dot{a} \right)^2 \right\}, \quad (4.4)$$

and the momentum conjugate to s is

$$p_s = \frac{\partial T}{\partial \dot{s}} = m \dot{\mathbf{r}} \cdot \frac{\partial}{\partial \dot{s}} \mathbf{r}. \quad (4.5)$$

If this is evaluated at the actual billiard wall ($a = 1$), this becomes

$$p_s = m \dot{\mathbf{r}} \cdot \frac{\partial}{\partial \dot{s}} \mathbf{r}_B = m \dot{\mathbf{r}} \cdot \mathbf{t} = p_t, \quad (4.6)$$

because \mathbf{t} is the unit tangent at the boundary. Therefore, the map on the plane s, p_t will be area-preserving, according to the result of the previous section, and we will refer to it as *the* surface of section of the billiard.

As is commonly done, we shall adopt a slight modification of this coordinate system whenever a SOS actually has to be displayed. Without losing any information, one can replace the arc length s by the conventional polar angle ϕ , as long as the dependence $s(\phi)$ is monotonic. This in turn is the case as long as the billiard is a star-shaped domain¹. Since ϕ and p_t are not canonically conjugate except in the circle, the resulting map is not in general area-preserving, and one should revert to s, p_t if this property is required.

Finally, it is convenient to scale out the energy dependence of p_t by using instead the variable

$$\sin \chi = p_t / \sqrt{2mE}. \quad (4.7)$$

This is simply the sine of the angle of incidence with respect to the normal. The pair $(s, \sin \chi)$ still generates an area-preserving map because we have only scaled one variable by a constant. For the Jacobian this just means that one row gets multiplied by $\sqrt{2mE}$ and one column divided by $\sqrt{2mE}$, leaving the determinant equal to one.

4.2 SOS of the circle and the ellipse

The simplest example of a billiard SOS is the circle, whose Poincaré map we can write down easily:

$$\sin \chi_{n+1} = \sin \chi_n \quad (4.8)$$

$$s_{n+1} = s_n + 2 \arccos(\sin \chi) \quad \text{modulo } 2\pi, \quad (4.9)$$

assuming the circle to be of unit radius (i.e. an area of π), and taking only counterclockwise trajectories into account. The distinction to clockwise orbits lies in the sign of $\sin \chi$, which is positive for counterclockwise motion and negative otherwise. In the relevant interval $\sin \chi = 0 \dots 1$, the arccos function is then monotonically decreasing with $\sin \chi$. In particular, the line in the plane $s, \sin \chi$ defined by $\sin \chi = 1$ is a stationary curve of the billiard map, which means $s_{n+1} = s_n$ for any point on it. We can convince ourselves that this type of stationary curve always exists for convex billiards of any deformation. The reason is that we must end up at $s_{n+1} = s_n$ if we try to launch a ray tangentially, because for a convex shape the tangent is always outside the domain except at the single point s_n . This is not the case in billiards with a non-convex boundary.

The SOS for the circle is shown in Fig. 4.1, and we observe both periodic orbits with rational winding numbers, cf. Eq. (3.107), and quasi-periodic orbits

¹I.e., there exists an origin from which all radial straight lines intersect the boundary once and only once.

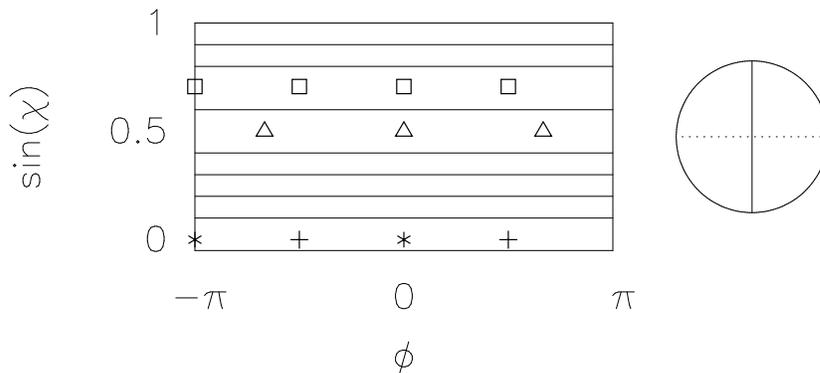


Figure 4.1: Momentum conservation in the circle leads to the invariance of $\sin \chi$, so that trajectories remain on a horizontal line in the SOS. Squares indicate a four-bounce orbit that inscribes a square in the circle, similarly triangles indicate a three-bounce orbit.

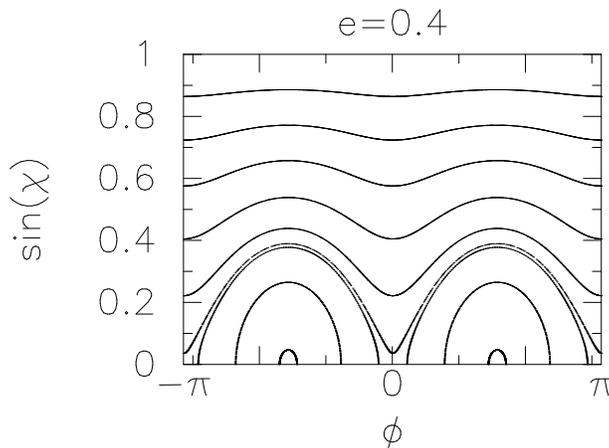


Figure 4.2: Surface of section for the ellipse at eccentricity $e = 0.4$.

which create horizontal lines because collisions can occur arbitrarily close to any point along the perimeter.

For the ellipse, the mapping equations can be derived explicitly, but they are of a more complicated form because the determination of the bounce position from the previous one requires the solution of a quadratic equation. However, we can immediately plot the SOS even without writing down the map, thanks to the integrability of the billiard and the relation for the invariant curves given in Eq. (3.136), which we rewrite here as

$$\sin \chi = \sqrt{1 + (S^2 - 1) \kappa^{2/3}}. \quad (4.10)$$

The constants have been lumped into S , which now plays the role of a “typical” $\sin \chi$ around which the invariant curve oscillates due to the oscillation in κ around $\kappa = 1$ (assuming again that the billiard area is π). Note that small values of S can lead to imaginary $\sin \chi$ near the points of highest curvature. The intervals of arc length s where this occurs are inaccessible to the orbit parametrized by the given S . Orbits for which this happens are of the bouncing-ball type described earlier in section 3.8.1. The resulting SOS shown in Fig. 4.2 shows two islands arising from the bouncing-ball trajectories. At their center lies the diametral orbit which reflects back onto itself between the two points of lowest curvature. This should be compared with the circle, where such a diametral orbit comes with an infinite family

of twins only differing by arbitrary rotations. The elliptical deformation destroys this isotropy.

In addition to the short diametral orbit at the center of the islands, the ellipse displays a long diametral orbit bouncing back and forth between the points of highest curvature. It is clear from the SOS that this orbit is of a different nature because it is not surrounded by an island. Trajectories starting in its neighborhood do not oscillate around this orbit as is the case for the bouncing-ball motion. Instead, perturbations of the long diametral orbit perform either a bouncing-ball motion (e.g., if one starts with $\sin \chi = 0$ and $\phi = \epsilon \neq 0$), or whispering-gallery motion (e.g. starting at $\phi = 0$ and $\sin \chi = \epsilon \neq 0$). As a consequence, the invariant curves in the neighborhood of the long diametral orbit have the appearance of hyperbolae. The short diametral orbit, on the other hand, is surrounded by curves that are approximately ellipses. This different behavior can be quantified by examining the map M in the neighborhood of these two orbits.

4.3 Fixed points and tangent map

The hyperbolic behavior occurs around the points $(\phi, \sin \chi) = (0, 0)$ and $(\pi, 0)$, whereas the ellipses surround the points $(\pm\pi/2, 0)$. We have

$$M(\pi/2, 0) = (-\pi/2, 0), \quad M(-\pi/2, 0) = (\pi/2, 0), \quad (4.11)$$

and

$$M(0, 0) = (\pi, 0), \quad M(\pi, 0) = (0, 0). \quad (4.12)$$

Therefore, each of these points is a *fixed point* of the squared map M^2 . More generally, any trajectory in the billiard that closes on itself after N reflections generates in the SOS a number N of fixed points of the iterated map M^N . Now one can perform a Taylor expansion of the vector-valued function M^N to linear order around one of these fixed points, say (s_0, p_0) , where we use the abbreviation p for $\sin \chi$ and reverted to the arc length as our position variable:

$$M^N(s, p) \approx M^N(s_0, p_0) + D[M^N] \cdot \begin{pmatrix} s - s_0 \\ p - p_0 \end{pmatrix} \quad (4.13)$$

$$= \begin{pmatrix} s_0 \\ p_0 \end{pmatrix} + D[M^N] \cdot \begin{pmatrix} s - s_0 \\ p - p_0 \end{pmatrix}. \quad (4.14)$$

Here, $D[M^N]$ is 2×2 the matrix of first derivatives of M^N . Since M is area-preserving, so is M^N , and consequently the matrix $D[M^N]$ has unit determinant. It is called the linearized map or *tangent map* at the fixed point (s_0, p_0) .

If we continue to iterate the map starting at the fixed point, then we will return to it every N steps, i.e. (s_0, p_0) is also a fixed point of $M^{\nu N}$ for any integer ν . Thus we can expand such a higher iterate in the same way as above to get in the neighborhood of the fixed point

$$M^{\nu N}(s, p) \approx \begin{pmatrix} s_0 \\ p_0 \end{pmatrix} + (D[M^N])^\nu \cdot \begin{pmatrix} s - s_0 \\ p - p_0 \end{pmatrix}. \quad (4.15)$$

To appreciate the effect of repeated application of the tangent map, we will need its eigenvalues $\lambda_{1,2}$ and eigenvectors $\mathbf{d}_{1,2}$, satisfying

$$D[M^N] \mathbf{d}_i = \lambda_i \mathbf{d}_i. \quad (4.16)$$

If we decompose the initial deviation from the fixed point as

$$\begin{pmatrix} s - s_0 \\ p - p_0 \end{pmatrix} = A_1 \mathbf{d}_1 + A_2 \mathbf{d}_2, \quad (4.17)$$

then the iterated deviation becomes

$$\left(D[M^N]\right)^\nu \cdot \begin{pmatrix} s - s_0 \\ p - p_0 \end{pmatrix} = \lambda_1^\nu A_1 \mathbf{d}_1 + \lambda_2^\nu A_2 \mathbf{d}_2. \quad (4.18)$$

The eigenvalues have to satisfy

$$\lambda_1 = 1/\lambda_2 \quad (4.19)$$

due to the unit determinant of the tangent map.

Now one can see how the elliptic or hyperbolic shape of the invariant curves in the vicinity of (s_0, p_0) comes about: In the expression

$$\lambda_1^\nu A_1 \mathbf{d}_1 + \lambda_1^{-\nu} A_2 \mathbf{d}_2, \quad (4.20)$$

we could formally consider $x \equiv \lambda_1^\nu$ as our variable, which would then lead to a hyperbola

$$x A_1 \mathbf{d}_1 + \frac{1}{x} A_2 \mathbf{d}_2 \quad (4.21)$$

provided that x is indeed real. This is in fact one of the two possibilities that can arise in the general case, as we show presently. Of course, x is not a continuous variable, but each ν puts a discrete point onto one and the same hyperbola, and we can cover both of its branches by admitting negative ν , corresponding to backward iteration of the tangent map.

Let t be the trace of $D[M^N]$, then a second equation for the eigenvalues in addition to Eq. (4.19) is

$$\lambda_1 + \lambda_2 = t, \quad (4.22)$$

which leads to a quadratic equation with the solution

$$\lambda_{1,2} = \frac{1}{2} t \pm \frac{1}{2} \sqrt{t^2 - 4}. \quad (4.23)$$

Now we can distinguish three cases.

$$a) \quad t < 2$$

which leads to real eigenvalues and consequently hyperbolic behavior near the fixed point as indicated above. It is also possible that

$$b) \quad t = 2$$

in which case $\lambda_1 = \lambda_2$. This non-generic special case is not of further interest at the moment. Finally, one can have

$$c) \quad t > 2$$

which leads to complex eigenvalues of the form

$$\lambda_{1,2} = \frac{1}{2} t \pm \frac{i}{2} \sqrt{4 - t^2} \quad (4.24)$$

whose absolute value is clearly unity, so they can be written with a phase factor β as

$$\lambda_{1,2} = e^{\pm i\beta}. \quad (4.25)$$

Since Eq. (4.17) is real and we apply to it the real matrix $(D[M^N])^\nu$ in Eq. (4.18), we know that the imaginary part of $\lambda_1^\nu A_1 \mathbf{d}_1$ must be canceled by the imaginary part of $\lambda_1^{-\nu} A_2 \mathbf{d}_2$ in Eq. (4.20). Here, we can again substitute

$$x = \lambda_1^\nu = e^{i\nu\beta}, \quad (4.26)$$

and take the real part, which now yields the equation of an ellipse if ν is varied.

4.4 Symmetries of the billiard map

Time reversal invariance manifests itself in a reflection symmetry of the SOS with respect to the horizontal axis, $\sin \chi = 0$. If $(\phi, \sin \chi)$ is a point in the SOS, then by definition there exists a trajectory which hits the boundary at ϕ with the given $\sin \chi$. If we reverse time for this motion, then $\sin \chi$ changes sign, and the resulting trajectory is also a solution of the equation of motion. Therefore, the point $(\phi, -\sin \chi)$ is also a point in the SOS.

Spatial symmetries lead to further symmetries in the SOS. If the billiard boundary has reflection symmetry about, say, the x -axis, then the mirror image of any trajectory under this reflection is also a valid trajectory. A point $(\phi, \sin \chi)$ is transformed into the point $(-\phi, -\sin \chi)$ by reflection at the x -axis. Combined with time reversal invariance, this also produces $(-\phi, \sin \chi)$, so that the SOS will have reflection symmetry around the line $\phi = 0$. If the billiard is, in addition, unchanged by reflections at the y -axis, then each point $(\phi, \sin \chi)$ in the SOS is accompanied by the points $(\pi - \phi, -\sin \chi)$ and (by time reversal) $(\pi - \phi, \sin \chi)$. Therefore, the SOS has reflection symmetry about the lines $\phi = \pm\pi/2$ (if we wrap all angles back into the interval $[-\pi, \pi]$).

Symmetries can be very helpful in finding periodic orbits of a map. It has been shown⁷⁸ that any periodic orbit of a reversible twist map is itself symmetric, meaning that there is a reflection axis that transforms the orbit into its time-reversed counterpart. However, there can still be periodic orbits with fewer symmetries than the billiard. For example, in the case of both x - and y - reflection symmetry, it is impossible for a single periodic orbit with odd-denominator winding number to satisfy all symmetries of the SOS. Such an orbit can still exist, but it must come with a twin of the same winding number that is obtained by reflection.

4.5 Poincaré-Birkhoff fixed point theorem

In this section it will be shown that fixed points of the hyperbolic type arise by necessity when the shape of a billiard is distorted. The billiard mapping for the circle,

$$\bar{p} = p, \quad (4.27)$$

$$\bar{s} = s + 2 \arccos p, \quad (4.28)$$

has the property that $\bar{s} - s$ is a monotonic function of p , except at the point $p = 1$. Leaving out this pathological point (see below), this is the definition a *twist map*.

We assume again that only positive p need to be considered. This twist condition has a strong effect on orbits with rational winding number when a perturbation is added to the above map, in the form

$$\bar{p} = p + \epsilon F(s, p), \quad (4.29)$$

$$\bar{s} = s + 2 \arccos p + \epsilon G(s, p), \quad (4.30)$$

where the functions F and G are chosen so as to maintain area-conservation. This is the kind of map we expect to obtain from a small deformation away from circular shape. An orbit with rational winding number $w = n/q$ corresponds, in the unperturbed case, to a set of q points in the SOS at

$$p = \sin \chi = \cos \pi w. \quad (4.31)$$

Each of these points is a fixed point of the map M^q , and there is an infinite family of such q -member sets which fills out the line $p = \cos \pi w$. The whole line is therefore *stationary* under application of M^q . If we apply M^q to points on a line at p^+ slightly above p , they will get mapped to smaller values of s due to the twist condition. Likewise, points on a line at $p^- < p$ get shifted toward higher s under M^q . Now we switch on the small perturbation of the circle. For ϵ small enough, the horizontal lines at p^+ and p^- continue to be shifted by M^q in the same direction as in the circle, even though the amount of the shift may not be the same everywhere along the respective line. Also, the line at p is in general no longer stationary for $\epsilon \neq 0$. In fact, there will generally be no stationary line left near p . The question that the Poincaré-Birkhoff theorem answers is what remains of the stationary curve when $\epsilon \neq 0$. Even if we cannot find a stationary line, we can identify a line that is stationary at least in the s -direction. By this we mean a set of points (s, p) which have only their p -component altered by M^q , with s being unchanged. We can find this line between the upper and lower limits p^\pm . For every s , one varies p in the interval $[p^-, p^+]$ until the point is found for which s is unchanged under M^q . This point exists in the said interval because the shift in s under M^q goes from positive to negative between p^- and p^+ . This yields the curve $p_1(s)$ for which the new s equals the old.

For some points along the curve $p_1(s)$, it may happen that not only s but also p remains the same under M^q . These will be the desired fixed points. To identify them, we let the map M^q act on $p_1(s)$, thus obtaining a new curve $p_2(s)$. The intersections of these two curves are exactly the points where neither s nor p is changed. However, it is not clear that $p_1(s)$ and $p_2(s)$ will intersect at all, i.e. there might be no fixed points left in the perturbed map. Here one has to invoke the area-preserving property of M^q . The line $p = 1$ is a stationary curve independently of ϵ , as long as the billiard remains convex. Consider the area of the SOS enclosed between the lines $p = 1$ and $p_1(s)$. Application of M^q to this domain changes only one of the boundaries, namely the lower one from $p_1(s)$ to $p_2(s)$. Since the area enclosed between $p = 1$ and $p_2(s)$ must be the same as before, $p_2(s)$ cannot lie entirely above or below $p_1(s)$. The two curves must necessarily intersect, and it follows that there exist fixed points of M^q . This construction also implies that fixed points come in pairs, because the number of intersections between $p_1(s)$ and $p_2(s)$ is even, due to the periodicity in s . Moreover, there will be q pairs of fixed points of M^q . This follows by taking one arbitrary fixed point (s_0, p_0) and subjecting it to M . The image point is also a fixed point of M^q because

$$M^q(M(s_0, p_0)) = M(M^q(s_0, p_0)) = M(s_0, p_0). \quad (4.32)$$

The same holds for $M^r(s_0, p_0)$ with $r = 2 \dots q - 1$, and the resulting q fixed points are all distinct provided that q is the smallest integer for which $M^q(s_0, p_0) = (s_0, p_0)$. The arrangement of fixed points resulting from the above argument is shown schematically in Fig. 4.3. Arrows indicate the direction in which points of the SOS are mapped by M^N , as explained in the caption. From this it can be seen that

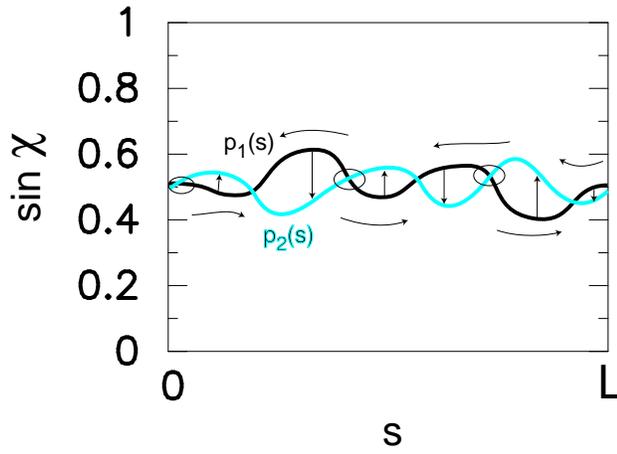


Figure 4.3: Origin of hyperbolic and elliptic points in the Poincaré-Birkhoff fixed point theorem. The solid black curve is $p_1(s)$, the blue curve is $p_2(s)$.

elliptic fixed points must always be separated from each other by a hyperbolic fixed point. The existence of hyperbolic fixed points can be viewed as the origin of chaotic dynamics in the perturbed map.

4.6 Stable and unstable manifolds and their intersections

Hyperbolic fixed points are characterized by two real eigenvalues of the tangent map that are reciprocals of each other. The corresponding eigenvectors define directions in the SOS along which points are mapped either away from or toward the fixed point under the action of $D[M^q]$. For example, any point on the line given by \mathbf{d}_1 through the fixed point (s_0, p_0) approaches this point if $\lambda_1 < 1$. Since consequently $\lambda_2 = 1/\lambda_1 > 1$, the direction \mathbf{d}_2 leads away from (s_0, p_0) . We call \mathbf{d}_1 the stable and \mathbf{d}_2 the unstable direction. In the stable direction, repeated application of $D[M^q]$ will bring us progressively closer to the fixed point, never away from it. The opposite is true for the unstable direction, at least until the validity of the tangent map breaks down.

These considerations are based on the linear approximation to the true map M^q , so the stable and unstable directions are only the linear approximations to the actual stable and unstable *manifolds* W_s, W_u . These are curves in the SOS to which $\mathbf{d}_{1,2}$ are tangent at the fixed point. Along W_s the map M^q brings us closer to the fixed point after many iterations, and along W_u we can apply the *inverse* map M^{-q} to achieve the same result of approaching the fixed point closer and closer. These manifolds are uniquely determined by the tangent directions $\mathbf{d}_{1,2}$ due to the requirement that a point (s, p) on, say, W_s not only be mapped toward the fixed point under sufficiently many applications of M^q , but that (s, p) itself be the image of some other point on W_s . In other words, $W_{s,u}$ are invariant sets and are called the *eigencurves* of M^q . One consequence of this definition is that neither the stable nor the unstable manifold can intersect itself. For W_s , such an intersection would have to be the image of two distinct points on the eigencurve under M^q , so that M^{-q} could not be single-valued. But we know that the map must be invertible because area-preservation implies

$$\det[D[M^q]] = 1 \quad (4.33)$$

on the whole SOS (here D need not be evaluated at the fixed points as it was

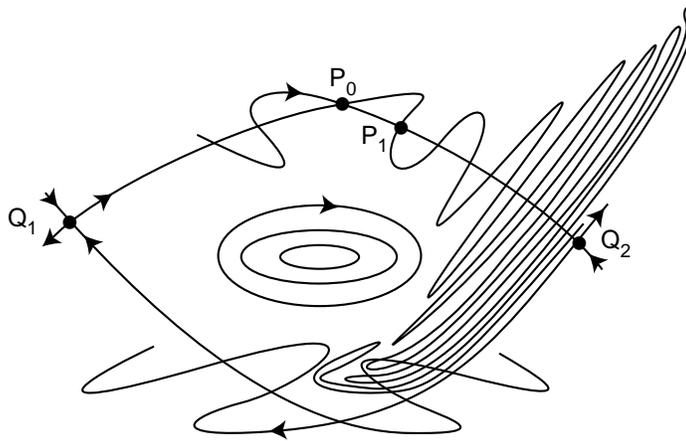


Figure 4.4: Unstable manifolds emanating from the hyperbolic fixed points Q_1 and Q_2 are indicated by an outward arrow, similarly an inward pointing arrow signifies a stable manifold. Only one of the tangles arising from the intersections of W_u and W_s is sketched. A stable island surrounds the elliptic fixed point.

so far). Therefore, the self-intersection can be ruled out. The picture becomes much more complicated if we consider intersections between W_u and W_s . The phenomenon to be described was already noted by Poincaré⁷⁹. A review of this subject was provided by Moser⁸⁰, and the qualitative discussion we rely on here is a variation of the treatment in various textbooks^{76,81,82}. Since any point of intersection must remain on both W_u and W_s after any number of iterations of M^q , there will in fact be an infinite number of intersections if there is one to begin with, which we label by P_0 . This is true whether W_u and W_s belong to the same hyperbolic fixed point or to different ones. The former type of intersections are called homoclinic, and the latter heteroclinic. In Fig. 4.4 we show the resulting complicated structure called a heteroclinic tangle. The infinite number of intersections between W_u and W_s occurs in a finite domain but with loops of ever-increasing length as the fixed point is approached. This follows again from area-preservation: let P_1 be an intersection adjacent to P_0 , forming with it a loop of area A_{01} . Then the loop formed between the image points $M^{\nu q}(P_{0,1})$ (where ν counts the iterations) must have the same area. But since the image points accumulate to one of the fixed points, this area becomes stretched and folded as we follow it to higher iterates of the map. The stretching loops described so far are, according to the figure, formed by an unstable manifold of the fixed point to the left, Q_1 , and accumulate at the other fixed point, Q_2 . Likewise, the stable manifold of Q_1 also accumulates at Q_2 , accompanied by similar stretching and folding.

An actual example of a heteroclinic tangle for a billiard system will be shown in Fig. 6.6 to illustrate its effect on phase-space transport.

4.7 Chaotic motion

The proof that this stretching and folding leads to chaotic motion near the hyperbolic fixed points was given by Smale⁸³ using the fact that the behavior of the interlocking eigencurves near these points can be described by the horseshoe map. We shall simply make the following qualitative remarks.

The complicated tapestry of interweaving stable and unstable manifolds near a hyperbolic fixed point forms an invariant set of M^q , due to the definition of W_u

and W_s . Consider two points on this invariant set near Q_2 separated from each other by some distance ϵ in the SOS. It is possible that the points are on the same or on different lobes of the tangle, either both on the same eigencurve, or one on W_u (of Q_1) and the other on W_s . This will determine the *direction* in which the two points are mapped, and very different outcomes are possible even at small separation ϵ . The *distance* by which the mapping moves the points can also vary widely, due to the wild oscillations of the eigencurves with the resulting stretching of the loops. As a result, one finds that the distance between the two neighboring points grows exponentially up to some saturation distance determined by the size of the chaotic portion of the SOS.

The exponential growth in separation δ is characterized by the *largest Lyapunov exponent* λ according to

$$\delta \approx \delta_0 e^{\lambda\nu}. \quad (4.34)$$

This law holds before the saturation time is reached, and one can often assume that this time is infinite, but it is also possible to extract λ if this is not a justified approximation⁸¹. On the other hand, one also has to wait for a certain number of iterations of the map to observe the exponential separation. Initially, the motion of the two trajectories may still be similar, if for example they are not launched in the midst of the tangle near a strongly hyperbolic point.

This behavior with a positive Lyapunov exponent in a finite neighborhood of the hyperbolic fixed point can serve as our definition of chaos. Because the Lyapunov exponent does not necessarily describe the initial evolution of neighboring points, chaos is a *long-time* property of a dynamical system. In particular, one finds that a trajectory launched in the chaotic domain comes arbitrarily close to any dynamically accessible point in that domain after a long enough time.

4.8 The KAM theorem

Chaos does not flood the whole SOS when integrability is destroyed by a perturbation. We already saw that hyperbolic fixed points alternate with elliptic ones, around which the trajectories rotate on approximate ellipses. These regions around elliptic fixed points form islands of stability, surrounded by the chaotic domains created from heteroclinic tangles of the neighboring hyperbolic fixed points. These regions connecting hyperbolic fixed points where the motion first starts to become stochastic are called separatrices. The name can be understood from the ellipse SOS where the invariant curves through the hyperbolic fixed points separate the rotating (whispering gallery) from the oscillatory (bouncing ball) motion. In nonintegrable systems, these invariant curves are the first ones to be destroyed.

From the discussion of the fixed-point theorem, it should be recalled that families of hyperbolic points were shown to emerge from stationary curves of the unperturbed map with rational winding number w . The proof fails for irrational w , so we should ask what the fate of the corresponding invariant curves is. Recall that irrational winding numbers can be defined even in the absence of action-angle variables using Eq. (3.109). The theorem that answers this question was initiated by Kolmogorov⁸⁴ for Hamiltonian flows and completed in that context by Arnol'd⁸⁵. For the case of interest here, namely area-preserving twist maps, the corresponding theorem was proved by Moser⁸⁶. The improvement over Arnol'd's work is that Moser does not need to assume the existence of an infinite number of derivatives of the map, whereas Arnol'd does assume an analytic Hamiltonian. The main result of the KAM theory is that invariant curves with irrational winding number do indeed

persist under finite perturbations of the integrable system, as expected from the non-applicability of the fixed-point argument.

The proof (reviewed in different amounts of detail in Refs.^{67,81,85,61}) is based on a perturbation theory in which one changes the initial conditions for the orbits with the small parameter ϵ (characterizing the nonlinearity of the map) in such a way that the motion maintains a given irrational winding number. It is possible to carry out the approximation in powers of ϵ to an infinite number of terms, and to show the rapid convergence of the resulting series. The difficulties arising in conventional classical perturbation theory due to small denominators associated with rational winding numbers are avoided by the requirement that the perturbed winding number remain equal to the given irrational value. The change in initial conditions necessary for this is implicit in a set of canonical transformations from the original action-angle variables with conserved momenta to a new set of variables whose momenta are again conserved and in fact equal to zero in the presence of the perturbation. The corrections to the approximate Hamilton equations are cubic in the new momenta, thus insuring the consistency of the solution with momenta equal to zero.

4.9 Resonance overlap and global chaos

The KAM theorem and the Poincaré-Birkhoff fixed point theorem seem to conflict with each other because the rationals are dense in the real numbers, so that each invariant curve with irrational winding number w has in its immediate neighborhood curves with rational w . The former are predicted to remain intact while the latter are predicted to break up. The resolution comes from the observation that the resonance zones associated with large-denominator rational $w = m/n$ are of smaller extent in the momentum direction than low-order (i.e. small n) resonances. Therefore, if we choose a sequence of rationals converging to the irrational winding number w_∞ , it could happen that the size of successive resonance zones in the momentum direction decreases faster than the distance between the resonant momenta themselves. If this occurs, the invariant curve at w_∞ does not overlap with the resonance zones of any of its rational approximates, and thus can be expected to remain intact. The reason this construction is possible is the fact that the rationals are of measure zero on the w -axis. This criterion of resonance overlap was first proposed by Chirikov in the context of escape of charged particles from a magnetic bottle⁸⁷.

4.9.1 The most irrational numbers

As our sequence of rational approximates, we should choose the $w_i = m_i/n_i$ such that they are in some sense the best possible rational approximations to w_∞ . By this we mean that m_i/n_i should come the closest to w_∞ compared to all rationals with the same or smaller denominator. This is satisfied⁸⁸ if we choose the sequence of *continued fractions* obtained by truncating the continued fraction expansion of w_∞ ,

$$w_\infty = [a_0, a_1, a_2, \dots] \equiv \left(a_0 + \left(a_1 + (a_2 + \dots)^{-1} \right)^{-1} \right)^{-1}. \quad (4.35)$$

The infinite number of terms in this expression is truncated to give

$$w_i \equiv [a_0, a_1, \dots, a_i]. \quad (4.36)$$

The irrational w_∞ hardest to approximate by rationals in this way are called “noble” winding numbers. They are characterized by the fact that their continued fraction

expansions have an infinite tail consisting only of 1's. The most irrational number in this sense is the *golden mean*,^{89,90}

$$\gamma = \frac{1 + \sqrt{5}}{2} = [1, 1, 1, \dots]. \quad (4.37)$$

As the perturbation of the dynamical system increases, these noble invariant curves are the last ones to be destroyed. The golden mean itself occurs as the most noble winding number in the standard map to be discussed presently, but it does not occur as a winding number in billiards, because there $|w| \leq 1$. However, $1/\gamma$ is also noble.

The picture of the transition to chaos that thus emerges is the following: The perturbation first causes invariant curves with rational winding number to disintegrate into chains of elliptic and hyperbolic fixed points, the latter being connected by stochastic layers along separatrices which enclose the islands around the elliptic fixed points. In between such so-called resonance zones, there remain irrational winding number invariant curves corresponding to rotational motion in the billiard. These remaining invariant curves disappear at higher values of the perturbation. When this happens, one speaks of a transition to global stochasticity.

To estimate how large the perturbation has to be in order to destroy the last KAM tori, one can again use the Chirikov overlap criterion, applied to the primary resonances, i.e. the ones with the largest separatrix regions. Using a simplified Hamiltonian that produces only one resonance zone (similar to the ellipse or the nonlinear 1D pendulum), one can calculate the value of the nonlinearity at which their separatrices begin to overlap, and that will be the desired critical perturbation. In fact, since the separatrix regions themselves contain regular islands surrounded by their own separatrix regions etc., one can improve the above estimate by taking this self-similar structure into account, using renormalization group arguments⁸¹.

4.9.2 The Chirikov standard map

The paradigm system in the study of area-preserving maps is the Chirikov standard map⁹¹, which can be derived as the (real-time) stroboscopic phase space portrait of a rigid rotor (one degree of freedom) subjected to time-periodic delta-function kicks⁸¹. The standard map describes the generic KAM scenario in the separatrix region between a family of hyperbolic fixed points. It can be written as

$$\begin{aligned} \bar{p} &= p + K \sin \phi \\ \bar{\phi} &= \phi + \bar{p}. \end{aligned} \quad (4.38)$$

Here, p and ϕ have the physical meaning of angular momentum and azimuthal angle, respectively, and the overbar denotes the same quantities after one period of the driving force. In this map, it has been found⁸¹ that the last invariant curve to be destroyed as the perturbation K increases from $K = 0$, is the “inverse-golden-mean” curve with winding number $1/\gamma$. The critical value of the *chaos parameter* at which this occurs is $K_c = 0.9716354$ according to numerical studies. From Chirikov's overlap criterion, one overestimates this value to be $K = \pi/4 \approx 2.47$. This is due to the fact that the separatrices of the primary resonances themselves have a finite width filled with stochastic motion and other island chains, as mentioned above. The estimate for K_c can be improved taking this width into account. In practice, whenever one can approximate a nonlinear map by the form of the standard map in some region of phase space⁹⁴, a reasonable estimate for the critical chaos parameter is simply

$$K \approx 1. \quad (4.39)$$

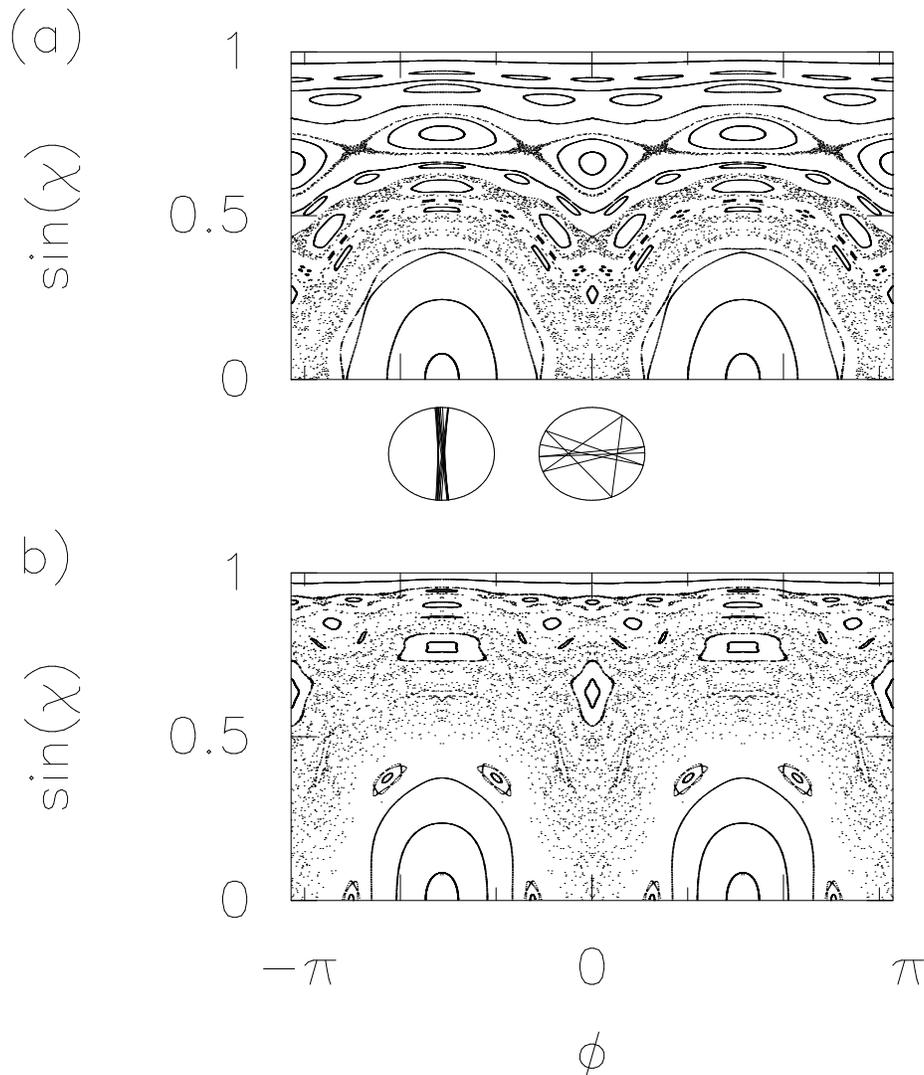


Figure 4.5: Poincaré surfaces of section for quadrupolar deformations of the circle (cf. Fig. 1.2) with fractional deformations $\epsilon = 0.065$ (a) and $\epsilon = 0.1$ (b). In the schematic below (a), trajectories close to the horizontal and vertical diametric orbits are plotted (each below its bounce position in the SOS).

It must be kept in mind that Chirikov’s overlap criterion is not a rigorous mathematical result, so that it should be viewed as a rule of thumb that works especially well for mappings that are very similar to the standard map, for which this criterium has been checked numerically. As regards rigorous results, it is fair to say that there are none that can be applied to rigorously to predict the onset of global chaos, i.e. the breaking of the last KAM curve, in general area-preserving maps.

4.10 Lazutkin’s theorem

The description given so far still does not capture the *global* transition to chaos correctly in the case of convex billiards. Locally, i.e. in a certain range of winding numbers, the KAM scenario takes place as described. This is illustrated in Fig. 4.5. However, this SOS also shows that there remain unbroken invariant curves near $\sin \chi = 1$ even at a deformation that has already destroyed all of the “most irrational” invariant curves: For the undeformed billiard map, we know from Eq. (3.107) that the momentum is $p = \cos \pi w$, and that $0 \leq w \leq 1$. This means that $w =$

γ does not occur, but other noble w like $1/\gamma$ do exist. In fact, $1/\gamma = [0, 1, 1, 1, \dots]$, so this is the most irrational winding number possible in the billiard, corresponding to $\sin \chi \approx \pm 0.36$, using time reversal, Eq. (3.108). It is a peculiarity of billiards that $1/\gamma^2 = [0, 2, 1, 1, 1, \dots]$ is just the time-reversed counterpart of the torus with $1/\gamma$. This follows from $1/\gamma^2 = 1 - 1/\gamma$ and the fact that $w = 1$ is the stationary curve. The figure shows stochastic motion in the region of the inverse golden mean curve, whereas the extreme whispering gallery orbits remain on invariant curves. But the latter correspond to winding numbers that tend to zero because the number of reflections per round trip becomes very large. In fact, $w = 0$ is a stationary curve of the convex billiard, even though it is clearly rational.

The existence of infinite families of invariant curves for any convex plane billiard with a sufficiently smooth boundary was proved by Lazutkin²⁹. The proof is actually concerned with the existence of caustics, and it is shown that the union of all such caustics forms a set of positive Lebesgue measure in the billiard plane. The existence of a caustic, however, is equivalent to the existence of an invariant curve in the SOS. This follows from the fact that a caustic per definition is a curve to which a given trajectory is tangent between any two consecutive reflections. Given a position s at which the reflection occurs and the curve C defining the caustic, there is only one counterclockwise (and one clockwise) straight line trajectory that we can draw tangent to C through the boundary point s . This also fixes $\sin \chi$, which is hence a unique function of s . If the trajectory does not close on itself, it will come arbitrarily close to any s , and thus trace out the curve $\sin \chi(s)$ in the SOS. This is the invariant curve; it is smooth because the caustic and boundary are smooth.

A chaotic trajectory does not form a caustic, because it fills a positive area in the SOS. The latter implies that a reflection at s can eventually (after some number of intermediate bounces) be followed by another reflection arbitrarily close to s , but both events may have values of $\sin \chi$ that differ by a finite amount. If the two rays were both tangent to a caustic, they would have to intersect, and this intersection would have to approach the caustic since the bounce positions approach each other. But this is possible with a finite difference in χ only if the intersection comes very close to the boundary at the point s . Since this argument applies to a finite interval of s , we conclude that the caustic is identical to the boundary in this interval. But this is a contradiction, because only reflections with $\sin \chi = 1$ are consistent with this condition, although we started with two different values of $\sin \chi$ near s . Hence the chaotic orbit does not form a caustic.

We still have to understand why the supposedly most noble winding numbers do not give rise to the most robust invariant curves in the billiard system. It has been noticed previously⁹² that the last invariant curves of a Hamiltonian system are not always associated with the golden mean, and the actual most stable winding number was found to be determined by the two primary (largest) resonance zones that bracket the KAM torus. In the billiard system, there is one additional reason for deviations from the typical KAM behavior, namely the fact that the twist condition is violated at $\sin \chi = 1$. This condition is one of the requirements for the KAM theorem, and is called “nondegeneracy condition” in that context. Our approach to an understanding of Lazutkin’s theorem does not follow the literature. Instead, we rely on the effective map to be discussed shortly, and find that the nonlinearity of the billiard map is itself a local quantity, dependent on $\sin \chi$. That is to say, regions of the SOS near $\sin \chi = 0$ are more strongly perturbed than regions near $\sin \chi \approx 1$. Hence, the whispering gallery orbits move on invariant curves because they are locally (in the $\sin \chi$ -direction) described by a map with small nonlinearity.

At the same deformation, the nonlinearity felt at $\sin \chi = 0.36$ is already so large that even the most noble invariant curve has disappeared.

The effective map approach cannot make any statements about non-convex billiards, so we briefly mention one of the few general theorems that are known for this case. The theorem proved by Mather⁹³ states that for a smooth billiard which has at least one point of vanishing curvature (but is still C^2), there are trajectories that reflect from the boundary both arbitrarily close to $\sin \chi = 1$ and also to $\sin \chi = -1$. Such a trajectory must reverse its sense of rotation between two such events, and that implies that there are no more caustics (or invariant curves) associated with rotation around the whole billiard boundary. To prove this implication we have to know that invariant curves are *barriers* in phase space, in the sense that a trajectory starting above such a curve which spans the whole SOS, can never appear below it. Given this, Mather's theorem immediately implies that there can be no invariant curves between $\sin \chi = 1$ and $\sin \chi = -1$. We will return to the notion of KAM curves as barriers in the SOS when we take up the discussion of phase space transport.

Chapter 5

Effective map for planar convex billiards

As long as the billiard is convex, there will be whispering gallery orbits near its boundary, as we saw in the discussion of Lazutkin's theorem. This suggests that approximate results for the billiard motion can be derived, using as a small parameter the proximity to the boundary. This corresponds to an expansion of the equations of motion in the limit $\sin \chi \rightarrow 1$, and it should be contrasted with a perturbation theory in which the degree of deformation is the small parameter: the whispering gallery limit can be taken for any convex billiard, independently of the deformation.

This will yield an approximate description of the billiard mapping from which we can draw insights into the way Lazutkin's caustics break up. We shall derive an "effective map" for convex billiards, which exhibits a kick strength that depends non-analytically on $\sin \chi$. The same $\sin \chi$ dependence also appears in the adiabatic invariant that was derived by Robnik and Berry⁵⁹, and we will show the close connection between both problems.

5.1 Non-analyticity of the kick strength

A non-analyticity of the mapping equations in $\sin \chi$ should be expected for $\sin \chi \rightarrow 1$ because $\sin \chi$ cannot be greater than one so that we are unable to expand around this value. If we expand instead in $\cos \chi$ around 0, then there is the possibility that the non-analyticity is removed, in the same way that one can expand for example the function $f(s) = \sqrt{1-s^2}$ in powers of $x := \sqrt{1-s^2}$ around $x = 0$, but not in s around $s = 1$. In this example, the function $f(s)$ is clearly nonnegative, whereas the substitution of variables yields $f(x) = x$ which can be negative. This is only an apparent contradiction, of course, because the substitution implies that x is never negative for any s .

In the case of the billiard map, we will similarly perform an expansion in $\cos \chi = \sqrt{1 - \sin^2 \chi}$, and obtain the final mapping function by re-substituting $\sin \chi$. Our hypothesis is that the only non-analyticity of the problem is contained in this substitution, so that expansions in $\cos \chi$ can in fact be performed.

Hence we consider the mapping equation

$$\cos \bar{\chi} = F(\bar{s}, \cos \chi). \quad (5.1)$$

Here \bar{s} and $\cos \bar{\chi}$ are the results of applying the map once to the point $(s, \cos \chi)$. It will become clear below that we should write \bar{s} as the argument of the function

F because we intend to construct a generating function. For now it shall suffice to remark that the mapping equations can in principle always be inverted to obtain this form. We assume that F can be expanded at least to second order,

$$\cos \bar{\chi} \approx f_0(\bar{s}) + f_1(\bar{s}) \cos \chi + f_2(\bar{s}) \cos^2 \chi + \dots \quad (5.2)$$

We can already draw some important conclusions without actually solving the geometry of the billiard motion. Firstly, we expect $f_0 \equiv 0$ since a trajectory starting with $\cos \chi = 0$ must end up with $\cos \bar{\chi} = 0$. The particle does not move at all in this limit of momentum tangential to the (convex) boundary.

Transforming Eq. (5.2) to the variable $\sin \chi$, we obtain

$$\sin^2 \bar{\chi} = 1 - \cos^2 \bar{\chi} \approx 1 - f_1^2(\bar{s}) + f_1^2(\bar{s}) \sin^2 \chi - 2f_1(\bar{s}) f_2(\bar{s}) (1 - \sin^2 \chi)^{3/2} \dots \quad (5.3)$$

Factoring out $\sin^2 \chi$, taking the square root and using $\sin \chi \approx 1$, we arrive at

$$\sin \bar{\chi} \approx \sin \chi \left[f_1^2(\bar{s}) + \frac{1 - f_1^2(\bar{s})}{\sin^2 \chi} - 2f_1(\bar{s}) f_2(\bar{s}) \frac{(1 - \sin^2 \chi)^{3/2}}{\sin^2 \chi} \right]^{1/2} \quad (5.4)$$

$$\approx \sin \chi - f_1(\bar{s}) f_2(\bar{s}) (1 - \sin^2 \chi)^{3/2}. \quad (5.5)$$

This exhibits the universal non-analytic form of the kick strength,

$$(1 - \sin^2 \chi)^{3/2}, \quad (5.6)$$

which tells us the way in which the nonlinearity of the billiard map decreases as $\sin \chi = 1$ is approached. Figure 5.1 shows how this general law compares with numerical simulations of real billiard systems. As an important consequence of this

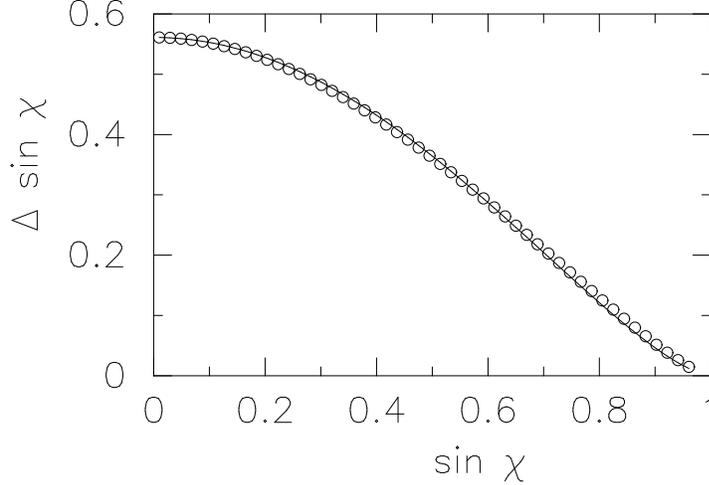


Figure 5.1: Kick strength in the momentum mapping equation, as a function of the initial $\sin \chi$ at eccentricity $e = 0.56$ in the quadrupole. The solid line is obtained from the true map by averaging the squared kick amplitude over all final ϕ and taking the square root. The circles are a fit with $(1 - \sin^2 \chi)^{3/2}$.

behavior, we will see later that the billiard map cannot be linearized around the stationary curve at $\sin \chi = 1$. Linearization of the position mapping equation is used in⁹⁴ to bring the Kepler map (describing a periodically driven Kepler problem) into the form of the Chirikov standard map. We therefore anticipate that the map

to be derived here will not reduce to the standard map in a similar way. This is reflected in the fact that the KAM transition in the standard map leaves as the last unbroken torus the one with winding number equal to the *golden mean*, $\gamma = (1 + \sqrt{5})/2$, whereas in a convex billiard γ is not a valid winding number, and there even remain unbroken tori long after the inverse golden-mean torus ($w = 1/\gamma$) has been destroyed by the perturbation. As mentioned in the discussion of Lazutkin's theorem, the surviving invariant curves in the latter system are those whose winding numbers approach zero, and this is to be expected since the curve with $w = 0$ is the stationary curve $\sin \chi = 1$.

Note that Eq. (5.5) yields the conservation law $\sin \bar{\chi} = \sin \chi$ if $f_2 \equiv 0$, independently of f_1 . On the other hand, Eq. (5.2) does not have the same property for $\cos \chi$, unless $f_1 \equiv 1$. To find the \bar{s} -dependence of f_1 and f_2 in Eq. (5.5), we have to resort to more detailed geometric considerations. These will in fact show that $f_1 \equiv 1$ while f_2 is a measure of the deformation from circularity. This makes sense because we expect the conservation of $\sin \chi$ to hold only in the circle. Recall that $\sin \chi$ is the tangential component of the linear momentum, which in turn equals the conserved angular momentum if the circle has unit radius. Before we embark on the geometric derivation of f_1 and f_2 , it is useful to sum up the principal result that will be obtained. The billiard mapping can be described approximately by the two implicit mapping equations

$$\sin \bar{\chi} = \sin \chi - \frac{2\kappa_1(\bar{s})}{3\kappa^2(\bar{s})} (1 - \sin^2 \chi)^{3/2}, \quad (5.7)$$

$$s = \bar{s} - 2 \arccos(\sin \chi) + 2 \left(1 - \frac{1}{\kappa(\bar{s})}\right) \sin \chi (1 - \sin^2 \chi)^{1/2}. \quad (5.8)$$

Here, κ is the curvature of the billiard as a function of arc length, and κ_1 is its derivative. Equation (5.7) shows explicitly that $\sin \chi = 1$ is a stationary curve of the map, because $\arccos(\sin \chi) \rightarrow 0$ there.

5.2 Geometric derivation of the momentum mapping equation

We start with some exact geometric relations⁵⁹, illustrated in Fig. 5.2. Let $\mathbf{t}(s')$ be

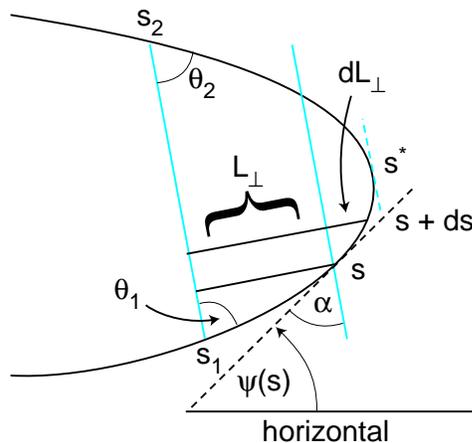


Figure 5.2: A ray intersects the convex boundary at arc lengths s_1 and s_2 .

the unit tangent vector at point s' on the boundary and $\psi(s')$ the angle between $\mathbf{t}(s')$ and the horizontal. The curve $\mathbf{t}(s')$ describes a unit circle as s' goes around

the boundary. Therefore the angle between the tangent vectors in two points s and s^* is simply given by the arc length of $\mathbf{t}(s')$ between s^* and s ,

$$\Delta\psi = \int_{s^*}^s \left| \frac{d\mathbf{t}}{ds'} \right| ds'. \quad (5.9)$$

But the integrand is the curvature of the boundary, so that we can write

$$\Delta\psi = \int_{s^*}^s \kappa(s') ds'. \quad (5.10)$$

This relation holds even for non-convex shapes because it correctly subtracts segments where \mathbf{t} rotates in the backward direction.

A second exact statement can be made if we consider now a straight line trajectory intersecting the boundary at arc lengths s_1 and s_2 . Let ψ^* be its angle with the horizontal. We can then ask what is the perpendicular distance L_\perp from a boundary point s to the line defined by the trajectory. To derive an integral expression for L_\perp , determine the increment dL_\perp when we move from s to $s + ds$. As shown in Fig. 5.2, one obtains a right angled triangle by drawing a parallel to the trajectory through s . The hypotenuse is ds , and the angle opposite the side dL_\perp is $\alpha = \psi^* - \psi(s)$. The perpendicular length is then obtained by integrating

$$dL_\perp = ds \sin \alpha, \quad (5.11)$$

between s and some end point. The latter is now chosen to be s_2 , and for the former we set $s = s_1$. Since these are the intersections between the trajectory and the boundary,

$$L_\perp(s_1) = L_\perp(s_2) = 0. \quad (5.12)$$

The integral now yields the condition

$$0 = L_\perp(s_2) = L_\perp(s_1) + \int_{s_1}^{s_2} dL_\perp = \int_{s_1}^{s_2} \sin(\psi^* - \psi(s)) ds. \quad (5.13)$$

There exists exactly one point s^* along the boundary between s_1 and s_2 at which the tangent is parallel to the trajectory. There, the tangent also is at an angle $\psi(s^*) = \psi^*$ to the horizontal. This s^* will now be our reference point in Eq. (5.10), i.e.,

$$\psi(s) = \psi^* + \int_{s^*}^s \kappa(s') ds'. \quad (5.14)$$

The relation we are trying to obtain is one between the angles of incidence and the positions s_1, s_2 . Consider the angles of incidence with respect to the tangent directions first, denoted by θ_1 and θ_2 . They can be combined with the tangent angles $\psi(s_{1,2})$ to give back the angle of the trajectory:

$$\psi^* = \psi(s_1) + \theta_1 = \psi(s_2) - \theta_2. \quad (5.15)$$

Adding these two equations yields an equation for the difference in the angle of incidence,

$$\theta_2 - \theta_1 = \psi(s_1) + \psi(s_2) - 2\psi^*. \quad (5.16)$$

Here we can use Eq. (5.14) to obtain

$$\theta_2 - \theta_1 = \int_{s^*}^{s_1} \kappa(s') ds' + \int_{s^*}^{s_2} \kappa(s') ds'. \quad (5.17)$$

So far, all expressions are exact, and the only remaining step is to eliminate the unknown s^* . To do this, one can apply Eq. (5.13), again using Eq. (5.14) to eliminate ψ^* :

$$\int_{s_1}^{s_2} \sin \left(\int_{s^*}^s \kappa(s') ds' \right) ds = 0. \quad (5.18)$$

This fixes s^* as a function of s_1 and s_2 , but it cannot be solved explicitly for s^* to insert into Eq. (5.17). Therefore, we expand $\kappa(s)$ to second order around s^* ,

$$\kappa(s) = \kappa(s^*) + \kappa'(s^*)(s - s^*) + \dots \equiv \kappa_0 + \kappa_1 (s - s^*), \quad (5.19)$$

which permits to calculate the integrals of the curvature. Then the sine function can also be expanded. As our expansion parameter we choose

$$\Delta s = s_2 - s_1, \quad (5.20)$$

and we write

$$s_1 - s^* = \alpha \Delta s + \beta \Delta s^2 + \dots, \quad (5.21)$$

$$s_2 - s^* = s_1 - s^* + \Delta s = (\alpha + 1) \Delta s + \beta \Delta s^2 + \dots, \quad (5.22)$$

Substituting the expansions into Eq. (5.18), we obtain

$$\begin{aligned} 0 &= \int_{s_1}^{s_2} \sin \left[\int_{s^*}^s \{ \kappa_0 + \kappa_1 (s' - s^*) \} ds' \right] ds \\ &\approx \int_{s_1}^{s_2} \left[\kappa_0 (s - s^*) + \frac{1}{2} \kappa_1 (s - s^*)^2 \right] ds \\ &= \frac{1}{2} \kappa_0 \left[(s_2 - s^*)^2 - (s_1 - s^*)^2 \right] + \frac{1}{6} \kappa_1 \left[(s_2 - s^*)^3 - (s_1 - s^*)^3 \right] \\ &= \frac{1}{2} \kappa_0 \left[(\alpha + 1)^2 - \alpha^2 \right] \Delta s^2 + \frac{1}{2} \kappa_0 \left[2(\alpha + 1)\beta - 2\alpha\beta \right] + \\ &\quad + \frac{1}{6} \kappa_1 \left[(\alpha + 1)^3 - \alpha^3 \right] \Delta s^3 + O(\Delta s^4). \end{aligned}$$

On the last line, we have ordered terms in ascending powers of Δs . Each coefficient has to vanish separately, leading to

$$\alpha = -\frac{1}{2} \quad (5.23)$$

for the Δs^2 term, and

$$\kappa_0 \beta + \frac{1}{6} \kappa_1 \frac{1}{4} = 0 \quad (5.24)$$

$$\Rightarrow \beta = -\frac{1}{24} \frac{\kappa_1}{\kappa_0} \quad (5.25)$$

from the cubic terms. Therefore,

$$s_1 - s^* = -\frac{1}{2} \Delta s - \frac{1}{24} \frac{\kappa_1}{\kappa_0} \Delta s^2, \quad (5.26)$$

$$s_2 - s^* = \frac{1}{2} \Delta s - \frac{1}{24} \frac{\kappa_1}{\kappa_0} \Delta s^2. \quad (5.27)$$

This can be used in Eq. (5.17), after we again used the expansion of the curvature:

$$\begin{aligned} \theta_2 - \theta_1 &= \kappa_0 (s_1 - s^*) + \kappa_0 (s_2 - s^*) + \frac{1}{2} \kappa_1 (s_1 - s^*)^2 + \frac{1}{2} \kappa_1 (s_2 - s^*)^2 \\ &= \frac{1}{6} \kappa_1 (\Delta s)^2. \end{aligned} \quad (5.28)$$

This result was already obtained in Ref.⁵⁹, but without any of the intermediate steps after Eq. (5.19). It was important to convince ourselves that the coefficient 1/6 in Eq. (5.28) is correct because it enters in our effective map, whereas this coefficient *drops out* in the derivation of the adiabatic invariant curves which was the goal in Ref.⁵⁹. This will be seen later when we derive the adiabatic invariant curves from the effective map.

To get the momentum mapping equation, a second equation is needed to eliminate Δs from Eq. (5.28). Therefore, we subtract the two Eqs. (5.15) to get

$$\theta_1 + \theta_2 = \psi(s_2) - \psi(s_1). \quad (5.29)$$

After performing the same expansions as above, one arrives at

$$\theta_1 + \theta_2 = \kappa_0 \Delta s. \quad (5.30)$$

Here, we only need the first order in Δs because the expression is substituted into Eq. (5.28) where Δs appears squared already. Eliminating Δs in this way, we get

$$\frac{6\kappa_0^2}{\kappa_1} (\theta_2 - \theta_1) = \theta_1^2 + \theta_2^2 + 2\theta_1\theta_2 \quad (5.31)$$

which is a quadratic equation for θ_2 :

$$\theta_2 = -\theta_1 + \frac{3\kappa_0^2}{\kappa_1} - \sqrt{\frac{9\kappa_0^4}{\kappa_1^2} - \frac{12\kappa_0^2}{\kappa_1} \theta_1}. \quad (5.32)$$

The minus sign for the square root is chosen to insure $\theta_2 = 0$ when $\theta_1 = 0$. Now expand to second order in θ_1 :

$$\theta_2 \approx -\theta_1 + \frac{3\kappa_0^2}{\kappa_1} - \frac{3\kappa_0^2}{\kappa_1} \left[1 - \frac{2\kappa_1}{3\kappa_0^2} \theta_1 - \frac{2\kappa_1^2}{9\kappa_0^4} \theta_1^2 \right] \quad (5.33)$$

$$= \theta_1 + \frac{2\kappa_1}{3\kappa_0^2} \theta_1^2. \quad (5.34)$$

If one now uses $\theta_1 \approx \cos \chi$ and $\theta_2 \approx \cos \bar{\chi}$, this becomes identical to the desired mapping equation, Eq. (5.2), with $f_1 \equiv 1$ and $f_2 = 2\kappa_1/(3\kappa_0^2)$. As a result, we have from Eq. (5.5)

$$\sin \bar{\chi} = \sin \chi - \frac{2\kappa_1(\bar{s})}{3\kappa_0^2(\bar{s})} (1 - \sin^2 \chi)^{3/2}. \quad (5.35)$$

We are allowed to use \bar{s} as the argument of κ , instead of s^* , because the difference is proportional to Δs , which would have led only to higher-order terms in Eqs. (5.28) and (5.30).

It should be kept in mind that there was no assumption of small deformation involved in the derivation of this equation, only the property of whispering gallery orbits that they bounce close to the boundary. We shall later determine the conditions for the validity of this assertion by determining what values of deformation and $\sin \chi$ are allowed so that trajectories do not break away from the boundary due to chaos. In this way we also provide a criterion for the applicability of Robnik's and Berry's adiabatic approximation. Such a criterion has been lacking so far.

An alternative way of writing the mapping equation for $\sin \chi$ is to use Eq. (5.34) but not the further approximation of Eq. (5.5). Then one obtains

$$\sin \bar{\chi} = \sqrt{1 - \left[\sqrt{1 - \sin^2 \chi} + \frac{2\kappa_1(\bar{s})}{3\kappa_0^2(\bar{s})} (1 - \sin^2 \chi) \right]^2}. \quad (5.36)$$

This has the advantage that it can be solved explicitly for $\sin \chi$ since it is a quadratic equation in $\sqrt{1 - \sin^2 \chi}$. By contrast, Eq. (5.35) cannot be inverted analytically. However, Eq. (5.36) will be less useful to us because its form is too complicated for the explicit implementation of area preservation which we will discuss below. It was noted e.g. in the treatment of the hydrogen atom in a microwave field⁹⁴ that the implicit nature of approximate maps is generally an unavoidable price to be paid for area preservation.

5.3 The position mapping equation

With Eqs. (5.35) or Eq. (5.36), we only have one of the two equations that describe the billiard mapping from bounce to bounce, namely that for the new momentum. The second equation must relate the old and new positions s and \bar{s} . We could try to derive it using the same approximations as above, and this was in fact done by Robnik and Berry⁵⁹. Their result is

$$\Delta s \approx \frac{2\sqrt{1 - \sin^2 \chi}}{\kappa \sin \chi}, \quad (5.37)$$

which has the unphysical property of diverging at low $\sin \chi$. This is not a problem in the limit $\sin \chi \rightarrow 1$ that underlies the expansion, and we could correct for it by including higher-order terms, to obtain instead

$$\Delta s \approx \frac{2 \sin^m \chi \sqrt{1 - \sin^2 \chi}}{\kappa}, \quad (5.38)$$

which agrees with the previous expression to second order in $\cos \chi$ for any choice of the integer m . However, in choosing some m , we have to satisfy the additional requirement that the map be area-preserving. If the map is to be useful as an approximation to the full billiard SOS, area-preservation must hold not just to order $\cos^2 \chi$ which is the accuracy of our derivation, but to all orders. Depending on the value of m , modifications must then be made in the momentum mapping equation, Eq. (5.35).

However, we already verified the $\sin \chi$ -dependence of Eq. (5.35) in Fig. 5.1, so that it should be a high priority to leave this equation unchanged. The approach taken here is to start from the momentum mapping equation already derived

above, find a generating function that reproduces this equation according to theorem 3.3.1.1, and then directly read off the position map. In this way, area-preservation is guaranteed and all corrections due to this are absorbed in the position map. We argue that this is consistent at least with some of the physical questions we want to answer with this map. Of interest is, e.g., the diffusion in the direction of decreasing $\sin \chi$ which provides the mechanism for resonance broadening in the deformed billiards. The fast variable s is not of immediate interest, and one in fact can make a “random phase approximation” by averaging over s to obtain the diffusion constant in $\sin \chi$. It will have to be seen if the position map can in fact be trusted enough to answer questions that do depend on the distribution of bounce positions.

Let us call the desired generating function $F(\bar{s}, p)$, where p stands for the conjugate momentum,

$$p \equiv \sin \chi. \quad (5.39)$$

If F is chosen such that the new momentum and old positions are obtained as partial derivatives,

$$\bar{p} = \left. \frac{\partial F}{\partial \bar{s}} \right|_p, \quad s = \left. \frac{\partial F}{\partial p} \right|_{\bar{s}}, \quad (5.40)$$

then the map is area preserving because the Jacobian is unity. Using the momentum mapping equation, Eq. (5.35), we can obtain the generating function by applying the first of Eqs. (5.40) backwards, i.e. integrating over \bar{s} . This leads to

$$F(\bar{s}, p) = p \bar{s} + \frac{2}{3\kappa(\bar{s})} (1 - p^2)^{3/2} + c(p), \quad (5.41)$$

where $c(p)$ is the integration constant which may still depend on p . Now we obtain the desired equation for the positions from the second of Eqs. (5.40),

$$s = \bar{s} + c'(p) - \frac{2p}{\kappa(\bar{s})} (1 - p^2)^{1/2}. \quad (5.42)$$

The mapping equation for s that we obtained from area preservation differs from the one obtained by Robnik and Berry not only in that p is absent in the denominator, but also in the fact that we have an additional p -dependent constant of integration. The additional $c'(p)$ in our relation gives us the freedom to achieve better agreement between effective map and the real billiard for lower p . First note that $c'(p)$ must vanish for $p \rightarrow 1$ with some power $(1 - p^2)^{3l/2}$, where l is an integer, $l \geq 1$. This is required in order to be consistent with the geometrically derived Eq. (5.37) to second order in $\cos \chi$.

The particular choice we make for c' is brought on by the additional requirement – not implied by the approximations made so far – that the effective map should reduce exactly to the correct map in the limit of the circular billiard. There, the position equation is

$$s = \bar{s} - 2 \arcsin(\cos \chi). \quad (5.43)$$

Comparing this with Eq. (5.42) in the limit $\kappa_0 \equiv 1$,

$$s = \bar{s} + c'(p) - 2 \sin \chi (1 - \sin^2 \chi)^{1/2}, \quad (5.44)$$

the two equations agree if

$$c'(p) = 2 \sin \chi (1 - \sin^2 \chi)^{1/2} - 2 \arccos(\sin \chi) \quad (5.45)$$

$$= 2 \cos \chi (1 - \cos^2 \chi)^{1/2} - 2 \arcsin(\cos \chi). \quad (5.46)$$

If this is expanded, one finds

$$c'(p) \approx 2 \cos \chi + \cos^3 \chi - 2 \cos \chi - \frac{1}{3} \cos^3 \chi + \dots \quad (5.47)$$

$$= \frac{2}{3} \cos^3 \chi = \frac{2}{3} (1 - p^2)^{3/2}. \quad (5.48)$$

This is explicitly of higher than second order in $\cos \chi$, and therefore consistent with our derivation. Inserting this into the mapping equation, we finally obtain the result summarized in Eq. (5.7),

$$\sin \bar{\chi} = \sin \chi - \frac{2\kappa_1(\bar{s})}{3\kappa^2(\bar{s})} (1 - \sin^2 \chi)^{3/2}, \quad (5.49)$$

$$s = \bar{s} - 2 \arccos(\sin \chi) + 2 \left(1 - \frac{1}{\kappa(\bar{s})}\right) \sin \chi (1 - \sin^2 \chi)^{1/2}. \quad (5.50)$$

A remark is in order about how to iterate the effective map, given its implicit nature. If we had to start with $(s, \sin \chi)$, it would be necessary to invert the complicated curvature expressions for the new position \bar{s} . However, this is entirely unnecessary if we simply reverse the meaning of old and new variables. That is, we iterate the map backwards, which is allowed due to time reversal invariance. Then one has to consider $(\bar{s}, \sin \bar{\chi})$ as given. Next, we solve Eq. (5.49) for $\sin \chi$. This must still be done numerically, but the functional form of the curvature does not enter the search algorithm. With the new $\sin \chi$ thus obtained, we directly read off the new s from Eq. (5.50) without the need for another root search.

As will be seen in the next chapter from the Poincaré sections for the real billiards as compared to the effective map, the resonance structure is reproduced quite well not just in the location along the $\sin \chi$ axis, but also with respect to island positions along the ϕ axis.

Chapter 6

Phase space structure and transport theory

6.1 Some model deformations

To establish some experience with the phenomenology of the chaotic transition in billiards, it is instructive to discuss some specific shapes. These will subsequently be used both in our classical and wave mechanical calculations. Only convex deformations of the circle are of interest to us, because that is the requirement for the existence of whispering gallery orbits. To compare different shapes among each other, a measure of the deformation is required. Since a convex billiard can be parametrized in polar coordinates, see section 4.1, the most general representation is

$$r(\phi) = \frac{1}{2}a_0 + \sum_{l=1}^{\infty} a_l \cos l\phi + \sum_{l=1}^{\infty} b_l \sin l\phi. \quad (6.1)$$

For billiards with a reflection axis, one can choose $b_l = 0$ for all l . This is true for all the cases we are dealing with. We can then refer to Eq. (6.1) as the two-dimensional *multipole* decomposition.¹The choice of origin is not unique. For example, Eq. (6.1) describes a circle if all Fourier amplitudes except a_0 vanish; but it is also possible to have the same circle *not* centered at the origin, which would require nonzero a_l ($l > 0$). The origin can be fixed by additional conditions. One natural choice would be the center of mass. One could also require that the multipole expansion have no dipole component $l = 1$. The resulting origins are not identical in general, but they do agree if there are two or more reflection axes in the billiard: in this case the intersection of the symmetry lines is the proper origin, and the lowest possible multipole component is quadrupolar. If there is exactly one reflection axis, we translate the origin along this line so as to eliminate any dipole component. This makes a comparison with the higher symmetry shapes most meaningful. All the deformations we consider leave the area constant to isolate the effect of shape alone.

¹When the b_l vanish, the Fourier expansion can also be regarded as an expansion in Chebyshev polynomials $T_l(\cos \phi)$. This means that the error made by truncating is smaller than for any other expansion in orthogonal polynomials of $\cos \phi$.

6.1.1 Quadrupole

The boundary is parametrized in polar coordinates by

$$r(\phi) = \frac{1}{\sqrt{1 + \epsilon^2/2}} (1 + \epsilon \cos 2\phi). \quad (6.2)$$

The area of this domain is π . All other deformations will have as their dominant multipole component this term, and we can therefore use the strength of the quadrupole part as a measure of the deformation that allows a comparison between different shapes.

6.1.2 Robnik billiard

The quadratic Robnik billiard is defined by the conformal mapping of the unit circle,

$$w = \frac{1}{\sqrt{1 + 2b^2}} (z + bz^2) \quad (6.3)$$

where $z = e^{i\phi}$, ($\phi = -\pi \dots \pi$), and b is real. It is possible to shift the origin so that the polar coordinate representation Eq. (6.1) of this shape only has a dipole component. This is why Eq. (6.3) will also be called the dipole billiard. To show this equivalence, write Eq. (6.3) as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{1 + 2b^2}} \begin{pmatrix} \cos \phi + b \cos 2\phi \\ \sin \phi + b \sin 2\phi \end{pmatrix} \quad (6.4)$$

$$= \frac{1}{\sqrt{1 + 2b^2}} \begin{pmatrix} \cos \phi [1 - b + 2b \cos \phi] \\ \sin \phi [1 + 2b \cos \phi] \end{pmatrix} \quad (6.5)$$

$$= \begin{pmatrix} \cos \phi r(\phi) - \frac{b}{\sqrt{1+2b^2}} \\ \sin \phi r(\phi) \end{pmatrix}, \quad (6.6)$$

where we defined

$$r(\phi) := \frac{1}{\sqrt{1 + 2b^2}} (1 + 2b \cos \phi). \quad (6.7)$$

This is the dipole form of the Robnik billiard, and the dipole parameter is $\epsilon = 2b$ if we define it in analogy to Eq. (6.2). The shape is shifted along the x -axis by an amount $\frac{b}{\sqrt{1+2b^2}}$.

However the polar origin we choose here yields a multipole expansion that has *no* dipole component to lowest order in b . In cartesian coordinates, we rename ϕ in Eq. (6.6) as θ to get

$$w = \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \frac{1}{\sqrt{1 + 2b^2}} \begin{pmatrix} \cos \theta + b \cos 2\theta \\ \sin \theta + b \sin 2\theta \end{pmatrix} \quad (6.8)$$

and define a translation through

$$\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} w_x - b \\ w_y \end{pmatrix}. \quad (6.9)$$

The corresponding polar coordinates (r, ϕ) satisfy the relations

$$\begin{aligned} \tan \phi &= \frac{y}{x} = \frac{\sin \theta + b \sin 2\theta}{\cos \theta + b \cos 2\theta - b} \\ r &= \frac{1}{\sqrt{1 + 2b^2}} \sqrt{(\cos \theta + b \cos 2\theta - b)^2 + (\sin \theta + b \sin 2\theta)^2}. \end{aligned} \quad (6.10)$$

Simplifying r , we arrive at

$$r = \frac{1}{\sqrt{1+2b^2}} \sqrt{1+4b^2-4b^2 \cos^2 \theta}. \quad (6.11)$$

To eliminate θ in favor of ϕ , use

$$x \sqrt{1+2b^2} = \cos \theta + b \cos 2\theta = \cos \theta + b(2 \cos^2 \theta - 1) \quad (6.12)$$

and solve this for $\cos \theta$ under the assumption that b is small. The result to lowest order is simply

$$\cos \theta \approx x = r \cos \phi. \quad (6.13)$$

Inserting this into Eq. (6.11), we can write to lowest order in b

$$r^2 \approx 1 + 2b^2 - 4b^2 \cos^2 \phi \quad (6.14)$$

which yields

$$r \approx 1 + 2b^2 - 4b^2 \cos^2 \phi = 1 + 2b^2 \cos 2\phi. \quad (6.15)$$

This is a pure quadrupole to leading order in b ; all higher multipoles are at least of order b^3 . Comparing this to Eq. (6.2), we can identify to leading order

$$\epsilon = 2b^2. \quad (6.16)$$

This relation is useful because it allows us to parametrize both the Robnik and the quadrupole billiard by ϵ , giving the strength of the quadrupolar deformation.

6.1.3 Ellipse

The usual polar coordinate representation of the ellipse is not useful here because it assumes the origin to coincide with one of the focal points. To retain the full C_{2v} symmetry around the origin, we start from the implicit equation

$$\frac{x^2}{a^2} + \frac{y^2}{c^2} = 1, \quad (6.17)$$

which becomes in polar coordinates

$$\frac{r^2 \cos^2 \phi}{a^2} + \frac{r^2 \sin^2 \phi}{c^2} = 1. \quad (6.18)$$

Solving for r and imposing $a = 1/c$ to keep the area constant and equal to π , we obtain

$$r(\phi) = \frac{c}{\sqrt{1 + (c^4 - 1) \cos^2 \phi}}. \quad (6.19)$$

To compare this to the quadrupole as we did for the Robnik billiard, we set $c \equiv 1 - \epsilon$ and expand in ϵ . The result is

$$r(\phi) \approx (1 - \epsilon)(1 + 2 \cos^2 \phi) = 1 + \epsilon \cos 2\phi \quad (6.20)$$

Thus, ϵ as defined here can be identified with the quadrupole parameter. We can also establish a relationship to the eccentricity,

$$e \equiv \frac{\sqrt{a^2 - c^2}}{a} = \sqrt{1 - c^4} \approx 2\sqrt{\epsilon}. \quad (6.21)$$

Table 6.1: Conversion table for the parameters of various billiards.

	Ellipse	Quadrupole	Dipole
Ellipse	e	$e = 2\sqrt{\epsilon}$	$e = \sqrt{8}b$
Quadrupole	$\epsilon = e^2/4$	ϵ	$\epsilon = 2b^2$
Dipole	$b = e/\sqrt{8}$	$b = \sqrt{\epsilon/2}$	b

6.2 Poincaré sections of the model billiards

For reference in later chapters, we append to this section a small gallery of Poincaré sections at selected values of the deformation. We have chosen to characterize the deformations by the strength of the quadrupolar component (in a frame where any dipole moment has been shifted out). To make this physically more meaningful, we give the *eccentricity* of an ellipse whose quadrupolar component has the same value as that of the given shape. To convert from e to the parameters native to the other billiards, one can use table 6.1. One observation that we will be able to explain below is the larger amount of chaos that seems to prevail in the dipole as compared to the quadrupole at the same eccentricity. For example, at $e = 0.5$ there is clearly more chaos in the dipole than in the quadrupole. This behavior can be understood in terms of the effective map.

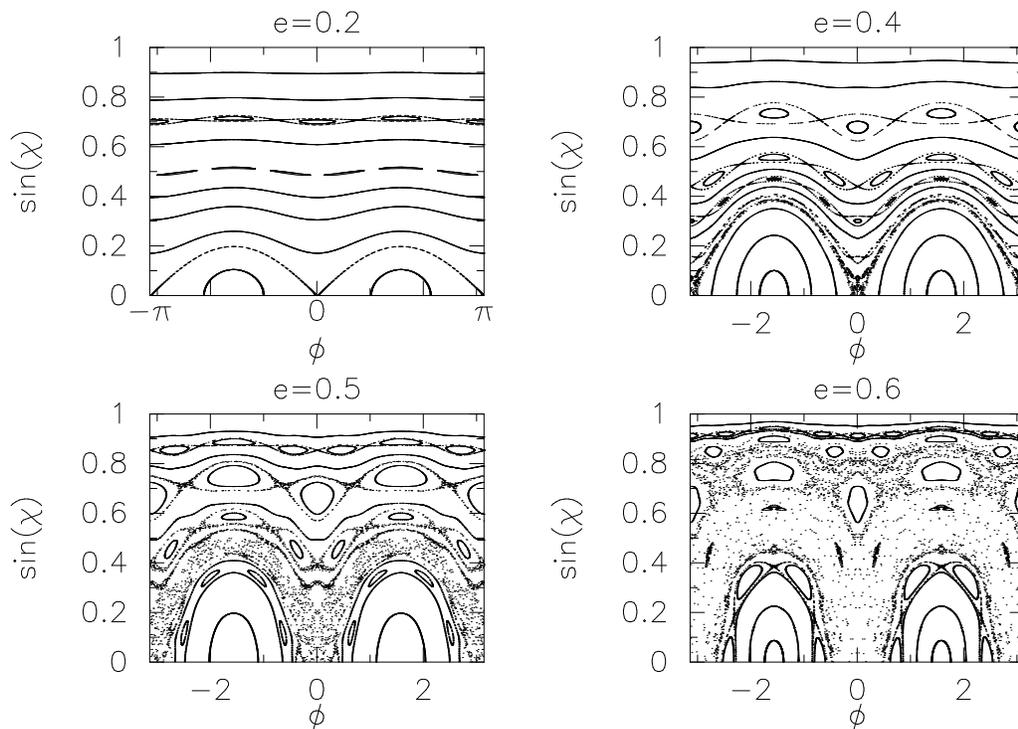


Figure 6.1: Poincaré sections for the quadrupole billiard. Chaos is seen to spread beginning from the low- $\sin \chi$ region where the stable and unstable two-bounce diametral orbits dominate the phase portrait. The next largest islands belong to the four-bounce orbit, and they survive to large deformation. The whispering gallery orbits near $\sin \chi = 1$ are largely unaffected by the deformation.

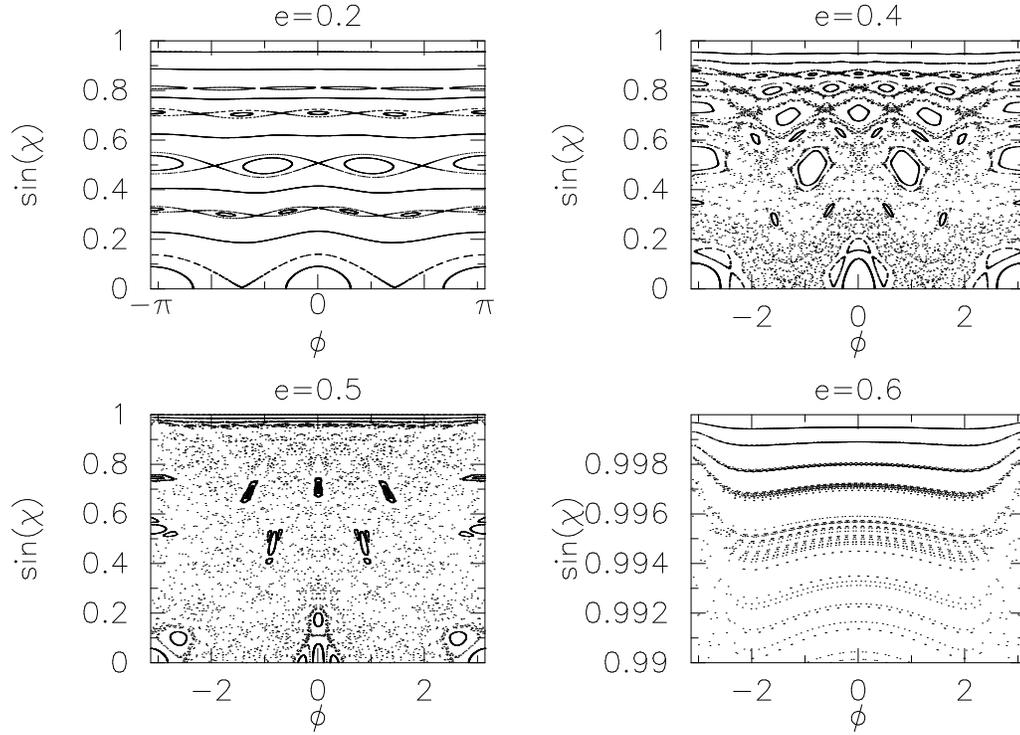


Figure 6.2: Poincaré sections for the dipole, showing strong influence of the three-bounce orbit. Surviving WG orbits and their break-up at large e are shown on an expanded scale.

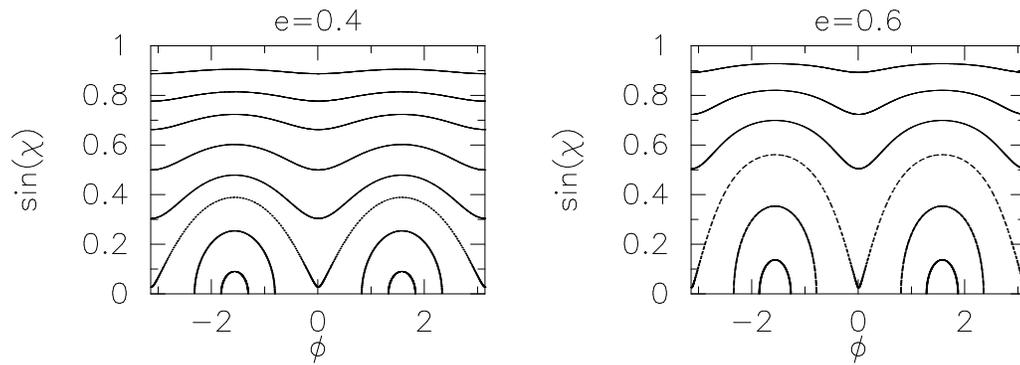


Figure 6.3: Poincaré sections for the ellipse billiard.

6.3 Local nonlinearity and its relation to Lazutkin's theorem

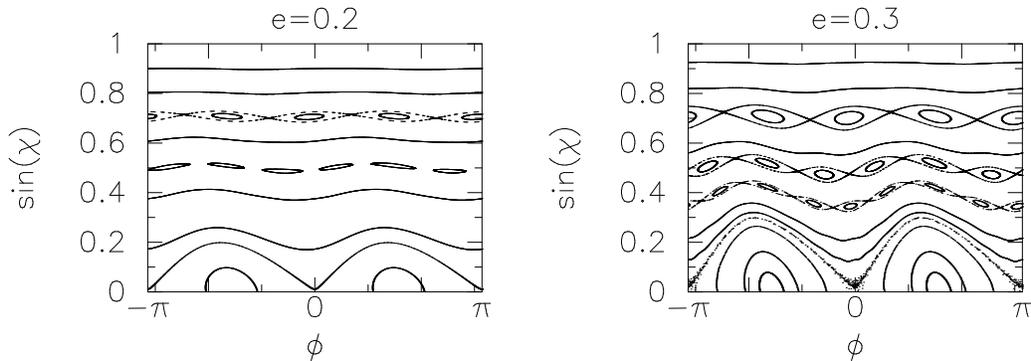


Figure 6.4: Poincaré sections for the effective map of the quadrupole. As in the real system, the four-bounce orbit produces the strongest island structure after the two-bounce orbit. The SOS for $e = 0.3$ is compared to the $e = 0.4$ SOS in Fig. 6.1, see text.

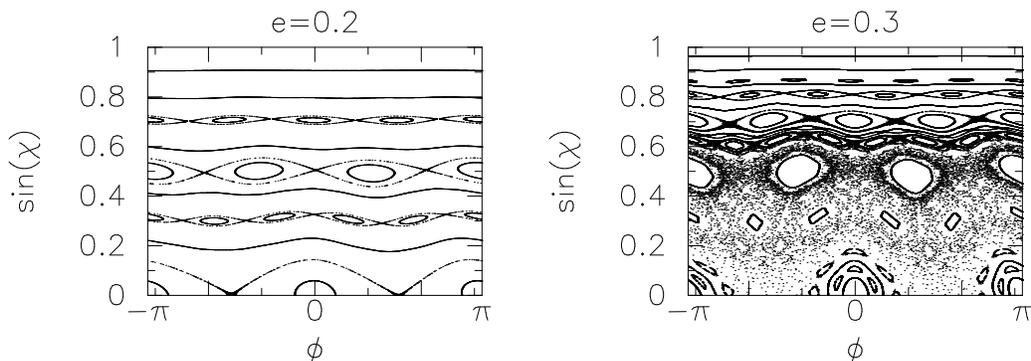


Figure 6.5: Poincaré sections for the effective map of the dipole. The three-bounce orbit creates larger islands here than the four-bounce orbit.

The various shapes introduced above can be used to generate the corresponding effective map, Eqs. (5.49) and (5.50). To that end, we use the polar coordinate parametrizations of the billiards to calculate the curvature,

$$\kappa(\phi) = \frac{r^2(\phi) + 2\dot{r}^2(\phi) - r(\phi)\ddot{r}(\phi)}{(r^2(\phi) + \dot{r}^2)^{3/2}}. \quad (6.22)$$

In Fig. 6.4, the resulting Poincaré sections are shown for the quadrupole billiard, and similarly for the dipole in Fig. 6.5. The island structure is seen to be reproduced in great detail, for both shapes, and in particular the characteristic differences between the real surfaces of sections are mirrored correctly. Comparing to the real SOS, we note however, that with increasing deformation the absolute value of e in the effective map does not coincide with the true eccentricity. The SOS obtained in the effective map for $e = 0.3$ looks more like the true billiard SOS at $e = 0.4$. Also, one notes that the effective map produces a skewed SOS, breaking the symmetry between positive and negative $\sin \chi$. This is due to the fact that the effective map does not preserve time reversal invariance to all orders in the deformation.

Now we want to use the effective map to make more quantitative statements about the chaotic transition as well as the phase space transport. Assume that the billiard shape is

$$r(\phi) = \frac{1}{\sqrt{1 + \epsilon^2/2}} (1 + \epsilon \cos m\phi), \quad (6.23)$$

which is normalized for any integer m so that the area is π . The normalization factor can be dropped if we want to perform an expansion in ϵ . In accordance with the periodicity of the boundary, it is useful to take the position variable to be $m\phi$ instead of ϕ , yielding for the derivatives

$$\dot{r}(m\phi) \approx -\epsilon \sin m\phi, \quad (6.24)$$

$$\ddot{r}(m\phi) \approx -\epsilon \cos m\phi, \quad (6.25)$$

$$\kappa(m\phi) \approx 1 + \epsilon \cos m\phi, \quad (6.26)$$

$$\dot{\kappa}(m\phi) \approx -\epsilon \sin m\phi. \quad (6.27)$$

We also make a *local approximation*⁴⁹, assuming $p = p_0 = \text{constant}$ in the kick strength. This is justified because p does not vary strongly in each iteration of the map at small deformation or for $p \rightarrow 1$, as is evident from Eq. (5.49). With this approximation, the effective map becomes

$$\bar{p} = p + \frac{2}{3} \epsilon (1 - p_0^2)^{3/2} \sin m\phi, \quad (6.28)$$

$$m\phi = m\bar{\phi} - 2m \arccos(p) \quad (6.29)$$

which is still area-preserving. Now we wish to make progress in the direction of reducing this to the form of the Chirikov standard map. The first equation is already of this form, compare Eq. (4.38). One still has to linearize the second equation with respect to p ⁹⁴, dropping the constant term in the Taylor expansion as an inessential phase offset. However, at $p_0 = 1$ the linear approximation diverges, and it will be poor for p_0 close to the whispering gallery region. This problem was already anticipated when we discussed the non-analytical kick-strength in the effective map.

Therefore we specialize to the limit of small ϵ , such that the local approximation, $p \approx p_0$, remains valid at lower p , where the linearization is more accurate. Small ϵ was also assumed in the derivation of Eq. (6.28). The linear expansion of Eq. (6.29) yields (dropping the constant terms)

$$\bar{p} = p + \frac{2}{3} \epsilon (1 - p_0^2)^{3/2} \sin m\phi, \quad (6.30)$$

$$m\phi = m\bar{\phi} + 2m \frac{1}{\sqrt{1 - p_0^2}} p, \quad (6.31)$$

which can be used in the Chirikov criterion, Eq. (4.39). We simply have to bring this into the standard form by using as our new momentum variable

$$P = 2m \frac{1}{\sqrt{1 - p_0^2}} p, \quad (6.32)$$

so that Eq. (6.30) becomes

$$\bar{P} = P + K \sin m\bar{\phi}, \quad K \equiv \frac{4}{3} \epsilon m (1 - p_0^2). \quad (6.33)$$

Here, K is the *local nonlinearity* of the effective billiard map. It is not only a function of deformation ϵ , but also of the characteristic momentum p_0 of the trajectories under consideration. We will discuss the limits of its validity shortly, but the result will be that at least the qualitative content is correct: As we go to higher p_0 at fixed deformation, chaos will decrease. This is the physical picture of Lazutkin's theorem which was already discussed.

The Chirikov criterion, Eq. (4.39), for the onset of chaos is $K \approx 1$, which means

$$\epsilon \approx \frac{3}{4} \frac{1}{m(1-p_0^2)}. \quad (6.34)$$

This result becomes obviously wrong for $p_0 \rightarrow 1$, because the billiards become non-convex at finite ϵ , and at that point we know that the last KAM curves should vanish. In the opposite extreme, $p_0 = 0$, one might hope to gain some information from this formula. The prediction then is that chaos sets in near $p_0 = 0$ for $\epsilon_d \approx 3/4$ in the dipole and $\epsilon_q \approx 3/8$ in the quadrupole billiard. Clearly, one has to expect that at these perturbations the linear expansion in ϵ is no longer accurate. However, we can compare these numbers with the actual billiard maps, and the tendency is reproduced correctly. Converting all deformation parameters to the Robnik parameter b as defined in Eq. (6.7) for the dipole and Eq. (6.16) for the quadrupole, we get

$$e_d = \sqrt{2}\epsilon_d = 1.06, \quad e_q = \sqrt{\epsilon_q/2} = 1.22. \quad (6.35)$$

This means that one needs a *smaller* quadrupolar admixture in the Robnik (dipole) billiard than in the pure quadrupole billiard to create the same amount of chaos near $p_0 = 0$. This is a nontrivial statement, and one could have expected the opposite since the radius changes more rapidly with angle in the quadrupole. It is seen to be qualitatively correct by comparing the quadrupole and the dipole in the gallery of Poincaré sections above, even though the numerical values for e are very different from the ones predicted. In fact, both e_d and e_q are much too high. The quality of the linearization of $\arccos p_0$ decreases because the p -interval in which it is valid shrinks as $p_0 \rightarrow 1$. But the interval over which p varies shrinks in this limit, too. Therefore, the non-analyticity at $p_0 = 1$ may not cause any problems if the allowed variation Δp is sufficiently small. From the adiabatic curve, Eq. (6.53), in the whispering-gallery limit $K \rightarrow \infty$, one finds that the amplitude Δp goes like $\eta/K^{2/3}$, where $\eta = O(\epsilon)$ is much smaller than one, given by the amplitude of the curvature. But K is related to the typical p around which the adiabatic curve oscillates by

$$K = (1 - p_0^2)^{-3/2}. \quad (6.36)$$

Therefore, we get

$$\Delta p \sim \eta(1 - p_0^2) = \eta(1 - p_0)(1 + p_0). \quad (6.37)$$

Due to the smallness of η , this is smaller than the distance from p_0 to the non-analyticity $(1 - p_0)$. Therefore, a standard map description is not inconsistent with the effective map, even in the whispering-gallery limit. The fact remains, however, that Eq. (6.34) is incorrect in this regime. The failure lies in the fact that we have to go to such high ϵ to see chaos near $p = 1$, that the expansion of the curvature to first order in ϵ is insufficient. But higher-order terms can have either a $\sin m\phi$ or $\cos m\phi$ dependence, and their superposition will produce deviations from the sinusoidal standard map, which in turn means that the Chirikov criterium cannot be applied without modifications.

We conclude that Eq. (6.34) describes the physics qualitatively, but we did not succeed in reducing the billiard map to the standard map without losing information about the billiard shape that would be essential for quantitative predictions. Nonetheless, the picture of Lazutkin's invariant curves as arising from the vanishing of the local nonlinearity near $p \rightarrow 1$ has been confirmed. Although further investigation is needed to try and reduce the billiard maps to other known maps, the above discussion seems to indicate that the billiard map is to be considered as an independent class of mappings in its own right.

In recent work by Borgonovi *et al.*⁴⁹, an approximate map for the stadium billiard was derived which also does not reduce to the standard map, but is similar to the saw-tooth map⁹⁵. They use the true polar coordinates instead of Birkhoff coordinates, and their momentum mapping equation displays a non-analyticity of the form $\sqrt{1-L^2}$ instead of the 3/2 power we obtained. Their position map is identical to Eq. (6.31), with $m = 1$, after making the local approximation in momentum. The important difference between the stadium billiard and our systems is that the former has boundary segments where $\kappa \equiv 0$, so that Lazutkin's invariant curves do not exist (even though there are orbits that rotate in one direction along the boundary for very long times).

6.4 Diffusion in the SOS

In the context of Mather's theorem, cf. section 4.10, it was noted that unbroken KAM curves corresponding to whispering gallery modes with a caustic constitute barriers in phase space. To understand this, we have to return to the continuous-time Hamiltonian flow from which the map originated. If we fix the energy, the motion takes place in a three-dimensional submanifold of the four-dimensional phase space. A KAM curve in the SOS, assumed for the sake of argument to be given by some function $p(s)$, is the projection of the two-dimensional surface of a torus that is embedded in the energy shell. The caustic results from the projection of the same torus onto the coordinate plane. The trajectory corresponding to a point (s, p_1) with $p_1 > p(s)$ cannot have touched the caustic since its last reflection because its angle of incidence is larger than for the trajectory belonging to the caustic. In other words, the trajectory giving rise to (s, p_1) was inside the torus when it went through its radial turning point. Then it must remain inside the torus at all times, because to leave, it would have to cross the surface of the torus. The latter is impossible because we know that KAM curves are invariant curves, and this property translates to the KAM torus. Hence, a phase space trajectory that is on the torus at some point in time will be there at all times in the future and the past. It is interesting to note that this argument is restricted to two degrees of freedom, because KAM tori do not partition a four- or higher-dimensional energy shell into an inside and an outside. Therefore, KAM curves in the SOS of systems with three degrees of freedom or more can in fact be crossed. This process may however be very slow or even be absent for certain initial conditions, and is called Arnol'd diffusion⁸². We have studied a 3D system in which this process may be occurring, namely the cylindrical billiard with a tilted cap, but more work is needed for a satisfactory understanding, so this work will not be expounded in this thesis.

As long as no phase space barriers of the type just discussed are encountered, a trajectory in the chaotic part of the SOS executes a stochastic motion during which it will in time come arbitrarily close to every point of the SOS belonging to the chaotic domain, and it will do so infinitely many times⁸². This *ergodicity* on the

energy shell is a long-time property, in the sense that a large number of iterations are needed before a trajectory has indeed explored all the available regions of the SOS. We can see this from the form of the effective map,

$$\bar{p} = p - \frac{2\dot{\kappa}(\bar{\phi})}{3\kappa^2(\bar{\phi})} (1 - p^2)^{3/2}, \quad (6.38)$$

$$\phi = \bar{\phi} - 2 \arccos(p) + 2 \left(1 - \frac{1}{\kappa(\bar{\phi})}\right) p (1 - p^2)^{1/2}. \quad (6.39)$$

Even if a trajectory does not move on an invariant curve, it is still not possible for \bar{p} to differ widely from p , especially when p is large, simply due to the fact that the kick strength depends on p , and the kicks are not only small (for large p) but also of either positive or negative sign due to the oscillation in curvature. Therefore, a large number of iterations is needed to see a change in p that is large compared to the kick strength.

The angle ϕ also advances slowly for whispering-gallery orbits, but it is taken modulo 2π and thus can cover the whole interval $[-\pi, \pi]$ evenly before p has departed much from its original value. This qualitative behavior is the motivation for a *random phase approximation* aimed at providing information only about the evolution of the slow variable p . The phases ϕ are assumed to become uncorrelated to their initial values after times long enough to produce a significant change in p . We calculate the phase average of $(\Delta p)^2$ by choosing an ensemble of initial conditions (ϕ_0, p_0) with fixed p_0 and evenly distributed ϕ_0 , and defining after n iterations

$$\langle (\Delta p(n))^2 \rangle \equiv \frac{1}{2\pi} \int_0^{2\pi} (\Delta p(n))^2 d\phi_0. \quad (6.40)$$

The sequence of points leading from (ϕ_0, p_0) to (ϕ_n, p_n) is determined by the equations of motion, and we have

$$\Delta p(n) = - \sum_{i=0}^n \frac{2\dot{\kappa}(\bar{\phi}_i)}{3\kappa^2(\bar{\phi}_i)} (1 - p_i^2)^{3/2}. \quad (6.41)$$

Inserting this into the previous equation, yields

$$\langle (\Delta p(n))^2 \rangle = \frac{4}{9} \sum_{i=0}^n \sum_{j=0}^n \left\langle \frac{\dot{\kappa}(\bar{\phi}_i)}{\kappa^2(\bar{\phi}_i)} \frac{\dot{\kappa}(\bar{\phi}_j)}{\kappa^2(\bar{\phi}_j)} (1 - p_i^2)^{3/2} (1 - p_j^2)^{3/2} \right\rangle. \quad (6.42)$$

Now ϕ_j is uncorrelated to ϕ_i except in an interval $|i - j| \ll n$, so that we can approximate the expectation value above by

$$\begin{aligned} & \left\langle \frac{\dot{\kappa}(\bar{\phi}_i)}{\kappa^2(\bar{\phi}_i)} \frac{\dot{\kappa}(\bar{\phi}_j)}{\kappa^2(\bar{\phi}_j)} (1 - p_i^2)^{3/2} (1 - p_j^2)^{3/2} \right\rangle \\ & \approx \delta_{i,j} \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{\dot{\kappa}(\bar{\phi}_i)}{\kappa^2(\bar{\phi}_i)} (1 - p_i^2)^{3/2} \right]^2 d\phi_0 \\ & = \delta_{i,j} \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{\dot{\kappa}(\phi)}{\kappa^2(\phi)} (1 - p_0^2)^{3/2} \right]^2 d\phi. \end{aligned} \quad (6.43)$$

With this, we obtain a *diffusive* growth in the variance of p ,

$$\langle (\Delta p(n))^2 \rangle = \left\{ \frac{4}{9} \frac{1}{2\pi} (1 - p_0^2)^3 \int_0^{2\pi} \left[\frac{\dot{\kappa}(\phi)}{\kappa^2(\phi)} \right]^2 d\phi \right\} n \quad (6.44)$$

$$\equiv D(p_0) n. \quad (6.45)$$

The diffusion constant $D(p_0)$ in the billiard map depends on p_0 , which is not the case in other maps like the standard map⁸¹ or the Fermi accelerator⁸². Specifically, we observe that D increases with decreasing p_0 , so that the diffusion will be *biased* toward smaller momenta. This behavior is central to our investigation of resonance lifetimes later on.

6.5 Adiabatic invariants from the effective map

Important corrections to this diffusive behavior arise due to the presence of phase space structure, but before these effects are introduced, we can use the slow diffusion in the momentum direction to derive the adiabatic invariant curves of Robnik and Berry⁵⁹.

Assume we are interested in the evolution of a billiard trajectory near p_0 on a time scale large compared to the phase randomization time but small compared to the diffusion time, i.e. we are considering n such that $\langle (\Delta p(n))^2 \rangle \rightarrow 0$. On this intermediate time scale, a trajectory comes close to any value of ϕ in the interval $[-\pi, \pi]$, but the possible values of p at two neighboring angles ϕ and $\phi + d\phi$ will still be infinitesimally close. In other words, the orbit traces out a smooth curve $p(\phi)$ in the SOS, and we can to a first approximation act as if $dp/d\phi$ exists. The existence of a one-dimensional curve to which orbits are confined means that the motion is approximately on a torus in phase space. This is synonymous with the existence of an almost-conserved quantity (in addition to the energy) which provides the constraint needed to force the trajectory onto a two-dimensional submanifold of the three-dimensional energy shell. The invariant turns out not to be p itself, but the typical value \tilde{p} around which $p(\phi)$ oscillates. Since we have to require the diffusion time in p to be much longer than the time scale on which we follow an orbit, we shall call \tilde{p} an *adiabatic invariant*.

We shall derive the adiabatic constraint equation connecting \tilde{p} to p and ϕ by making use of the effective map, Eqs. (6.38) and (6.39). One can write the two mapping equations as

$$\Delta\phi \equiv \bar{\phi} - \phi = \frac{2p(1 - p^2)^{1/2}}{\kappa} - c'(p), \quad (6.46)$$

$$\Delta p \equiv \bar{p} - p = -\frac{2\kappa_1(1 - p^2)^{3/2}}{3\kappa^2} \quad (6.47)$$

Our goal is to derive an equation of the form

$$p = p(\phi, \tilde{p}), \quad (6.48)$$

with \tilde{p} as a parameter. The idea is to find this function for the limiting case $\tilde{p} \rightarrow 1$ (i.e., the whispering-gallery limit), and then generalize by substituting arbitrary values of \tilde{p} into the function thus obtained. Therefore, we first take $\Delta s \rightarrow 0$, which implies from the $\Delta\phi$ -equation that $\cos\chi \rightarrow 0$. In this limit, $c'(p)$ can be neglected compared to the first term because $c' \propto \cos^3\chi$. As $\cos\chi \rightarrow 0$, Eq. (6.47) implies

$\Delta p \rightarrow 0$. But this means that we can expect a finite value for $\Delta p/\Delta s$, i.e. the derivative dp/ds exists. The above difference equations can then be converted to a differential equation by dividing the two lines above,

$$\frac{dp}{d\phi} = -\frac{\kappa_1(1-p^2)}{3\kappa p}. \quad (6.49)$$

This can be solved by separation of variables because we can write

$$\frac{y}{1-p^2} dp = -\frac{\kappa_1}{3\kappa} d\phi. \quad (6.50)$$

Since $\kappa_1 = \kappa'$ is just the derivative at ϕ in this limit, we can integrate to obtain

$$\tilde{K} \frac{1}{2} \ln(1-p^2) = \frac{1}{3} \ln \kappa \quad (6.51)$$

$$\Rightarrow K = \frac{\kappa}{(1-p^2)^{3/2}} \quad (6.52)$$

$$\Rightarrow p = \sqrt{1 - \left[\frac{\kappa(\phi)}{K} \right]^{2/3}}. \quad (6.53)$$

The integration constant K is therefore a conserved quantity for the map in the limit $\Delta s \rightarrow 0$, and it could serve as our adiabatic invariant. To endow the invariant with the desired physical meaning, we define

$$\tilde{p} \equiv \sqrt{1 - \frac{1}{K^{2/3}}}, \quad (6.54)$$

so that the adiabatic invariant curves become

$$p = \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}}. \quad (6.55)$$

This is the constraint function, and p is seen to oscillate around \tilde{p} because κ oscillates around the value for a circle of equal area, $\kappa_0 = 1$. We can now insert the whole allowed range of invariants, $0 \leq \tilde{p} \leq 1$ from Eq. (6.54). Of course, this course of action is no longer a controlled approximation because terms of higher order in $\cos \chi$ may become important in the constraint equation. However, there exists one additional limit that we can take to check the validity of the adiabatic curve for small \tilde{p} (away from the whispering gallery region): it consists in making the shape purely elliptical, i.e. integrable.

As has been noted in Ref.⁵⁹, Eq. (6.55) becomes exact for the whole phase space in the ellipse billiard where there is no chaos. This is by no means obvious from our derivation (nor from theirs), and we can only comment that the derivation of Eq. (6.51) ignores the existence of resonance zones of the effective map, and such regions are in fact absent in the ellipse. This remarkable identity for the ellipse of arbitrary deformation means that the functional form of the adiabatic curves is given by our result even at small p . This is a strong indication that other billiard shapes should be described reasonably well, too.

However, the question arises why the effective map can exhibit chaos at all, given the fact that it was derived as an approximation to second order in $\cos \chi$, while the differential equation Eq. (6.49) is exact to the same order and invariably provides a constant of integration. It is the separation of diffusion and phase randomization time scales that breaks down with increasing local nonlinearity. Thus, if ϕ and

$\phi + d\phi$ are close, the corresponding values of p can retain a finite difference due to the diffusion that already took place while the phase angle wrapped around a few times to come back to $\phi + d\phi$. Then, the trajectory no longer describes a smooth curve in the SOS, and the differential equation leading to the adiabatic invariant curves is inapplicable.

Without attempting any quantitative predictions, we can show that the adiabatic invariant curves become exact in the whispering gallery limit because the separation of scales becomes infinitely large. This is a physicist's way of proving Lazutkin's theorem. Define the (discrete) diffusion time in p , τ_p to be the number of iterations it takes to diffuse across the whole allowed p -interval, and the phase randomization time τ_ϕ as the number of reflections necessary before ϕ has wrapped once around the boundary. Then from Eq. (6.44),

$$\tau_p \propto \frac{1}{(1-p^2)^3}, \quad (6.56)$$

and from Eq. (6.46),

$$\tau_\phi \propto \frac{1}{p\sqrt{1-p^2}}. \quad (6.57)$$

The proportionality constants are functions of the deformation alone. As the whispering-gallery limit is approached, τ_p diverges much faster than τ_ϕ , so that the existence of invariant curves can be inferred. In the limit $p \rightarrow 0$, τ_ϕ does not diverge but saturates because of the term $c'(p)$ which was neglected in the position mapping equation. It can then be of the same order of magnitude as τ_p (depending on the deformation).

To show that the effective map, unlike the ellipse, is indeed described only approximately by the adiabatic invariant curves, it suffices to start a single trajectory on the invariant curve and demonstrate that one (backward) iteration of the map will cause it to depart from this curve. Note that the adiabatic curve has minima at the *points of highest curvature*. In the quadrupole, these are at $\phi = 0$ and $\phi = \pi$. A trajectory that ends up at $\phi = 0$ with $\Delta p = 0$, must have originated from $\phi = \pi$, i.e. $\Delta\phi = \pi$, if it stayed on the same invariant curve. This is true irrespective of the value of p at the minimum. But the position map, Eq. (6.46), then reads

$$\pi = \frac{2p(1-p^2)^{1/2}}{\kappa} - c'(p) \quad (6.58)$$

for all p , where κ is the value of the curvature maximum. In our choice of $c'(p)$, there was no curvature dependence, so that the equation cannot hold in general. Hence we arrived at a contradiction by assuming that starting and end points of the trajectory are on an invariant curve.

The above discussion leads to the following description of the diffusion process in phase space: On time scales where diffusion can be neglected, p is not actually constant, but oscillates around \tilde{p} according to the adiabatic invariant curve. This means that the diffusion process in p is in fact roughly a sequence of steps during which the trajectory tends to follow an adiabatic curve around \tilde{p} for an intermediate time interval, where \tilde{p} decreases on average as time progresses. An important conclusion is that the probability of finding the trajectory at momentum p_0 depends on ϕ . The reason is that the given p_0 will be reached earliest by a trajectory moving along an adiabatic curve $p(\phi)$ whose minimum value is just p_0 . But the minimum is reached at only a certain number of points ϕ_0 , namely those at which the curvature

is a maximum, cf. Eq. (6.55). At the time when this adiabatic curve is reached, p_0 can occur only near the ϕ_0 .

To reconcile this ϕ -dependence in the distribution of p with the fact that we made a random phase approximation for ϕ , we have to introduce a more refined understanding of the diffusion in which it is the adiabatic invariant \tilde{p} , not p , which evolves according to the diffusion law, Eq. (6.44). The anisotropy of the p distribution in ϕ will be of great importance in understanding the directional emission of asymmetric resonant cavities.

6.6 Transport in the presence of phase space structure

In the study of optical cavities, it has also turned out to be crucial to understand how deviations from the above diffusion model occur due to the presence of phase space structure. We give here a brief overview of the modifications to transport properties that can arise in a generic KAM system.

6.6.1 Cantori

In a convex billiard, chaotic regions always coexist with KAM curves, and the transition between these domains exhibits island chains of high order (large denominator in the winding number), corresponding to the rational approximates to noble KAM curves. A special structure arises just after a KAM curve (with irrational winding number) has disintegrated due to the overlap with the resonances of the rational approximates. The resulting invariant set no longer forms a continuous curve $p(\phi)$, but a Cantor set, which is appropriately called a cantorus. This means that the KAM torus acquires holes at arbitrarily fine scales, such that any point in the invariant set is arbitrarily close to other points in the cantorus, but at the same time is itself a boundary point of the set^{76,81,96}.

Cantori of noble winding number form partial barriers to phase space diffusion in the momentum direction, which is intuitively clear considering that a trajectory has to circumnavigate the dense island structure of the rational approximates. Of course transport only becomes measurable if the deformation is sufficient to go beyond the rigorous Cantor set stage to a set with slightly larger voids. The aggregation of minuscule gaps to larger gaps with increasing deformation is a consequence of the fact that the smallest gaps are produced by overlap with the separatrices of the smallest islands, which in turn are the first to disappear when the nonlinearity increases. The smallest islands are generally the first to vanish, because they correspond to fixed points with less stability. What remains are chains of slightly larger islands whose separatrix regions spread toward the cantorus and thus create larger gaps in the latter. Transport across this partial barrier was reduced by MacKay *et al.* to the action of turnstiles formed in the gaps, and a scaling approach was used by them and independently by Bensimon and Kadanoff⁹⁷ to derive a power-law dependence of the crossing time on the chaos parameter.^{96,97} Renormalization is needed because of the self-similar structure consisting of island chains around islands in other island chains⁸¹. The effect of the self-similar structure in the SOS is that the diffusion in the vicinity of island chains is not of the simple type derived above, but instead behaves *anomalously* in close analogy to a Lévy flight: since islands are themselves surrounded by remnants of cantori, a trajectory can get captured inside such a secondary cantorus for long times, during which it may wander into the vicinity of an even smaller island by which it may get captured in the same way,

leading to the possibility of extremely long dwell times. The result of these theories is that the time needed to cross the cantorus region depends on the chaos parameter through a power law,

$$T \propto (K - K_c)^\alpha, \quad (6.59)$$

where $\alpha \approx 2.55 \dots 3.026$, where the lower value was found in numerical studies by Chirikov,⁹⁸ and the highest value is due to Bensimon and Kadanoff (also a numerical result). MacKay *et al.* find $\alpha \approx 3.012$.

6.6.2 Stable and unstable manifolds as barriers

Transport in phase space is also obstructed by intersecting stable and unstable manifolds, and this is in fact a generalization of the turnstile concept applied in the study of cantori⁵¹. The idea is that, just like a cantorus or a KAM curve, unstable and stable manifolds are also invariant curves, however not necessarily of the Poincaré mapping M itself, but rather of higher iterates M^q . If q is small, then these curves may act very much like true barriers. The consequence of $q > 1$ is simply that there is a chain of q hyperbolic fixed points, each of which radiates stable and unstable manifolds that intersect. From one iteration to the next, one then jumps between these different manifolds, returning to the original manifold with M^q . Due to the intersections, there will be heteroclinic tangles, which are the reason why the barrier action is not complete: one can say that the barrier itself is so foliated and its loops reach into large regions of phase space, that a decomposition into smoothly connected domains separated by the barrier is meaningless. However, the heteroclinic tangles are concentrated mostly near the fixed points, and thus there can be portions of the stable and unstable manifolds in between the hyperbolic points that are largely free of intersections. This is seen in the sketch of Fig. 4.4, and one can define a boundary line separating an “inside” from an “outside” by connecting the two fixed points through the shortest path along either the stable or the unstable manifold. The loops are then either inside or outside this line, and one has to be inside a loop to get mapped across the boundary. This is the reason why fixed points with low q create better barriers: The fixed points are in general farther apart, and the space inbetween is less occupied by the loops of the tangle (which forms near the fixed points).

The boundary defined above becomes the separatrix in the ellipse billiard, because in that case the stable manifold of one fixed point smoothly becomes the unstable manifold of the next. This means that there are no tangles and hence no loops that could propel a trajectory across the boundary. The loops are analogous to what MacKay *et al.* call turnstiles, and it is the area of the loops that determines the time needed to cross these turnstiles^{51,81}.

The effect of intersecting manifolds on the phase space diffusion is thus similar to that of cantori, in that transport can slow down significantly. However, there is no threshold behavior or power law associated with these barriers because they exist in the separatrix regions between hyperbolic fixed points, which are locally the first regions to become chaotic and thus will be sandwiched by KAM curves upon their inception. They become dominant as transport barriers only after the last cantori in the neighborhood have disappeared, even though they already exist prior to that point.

To demonstrate in what way heteroclinic tangles appear in a billiard system, we show a concrete example in Fig. 6.6 for the quadrupole shape.

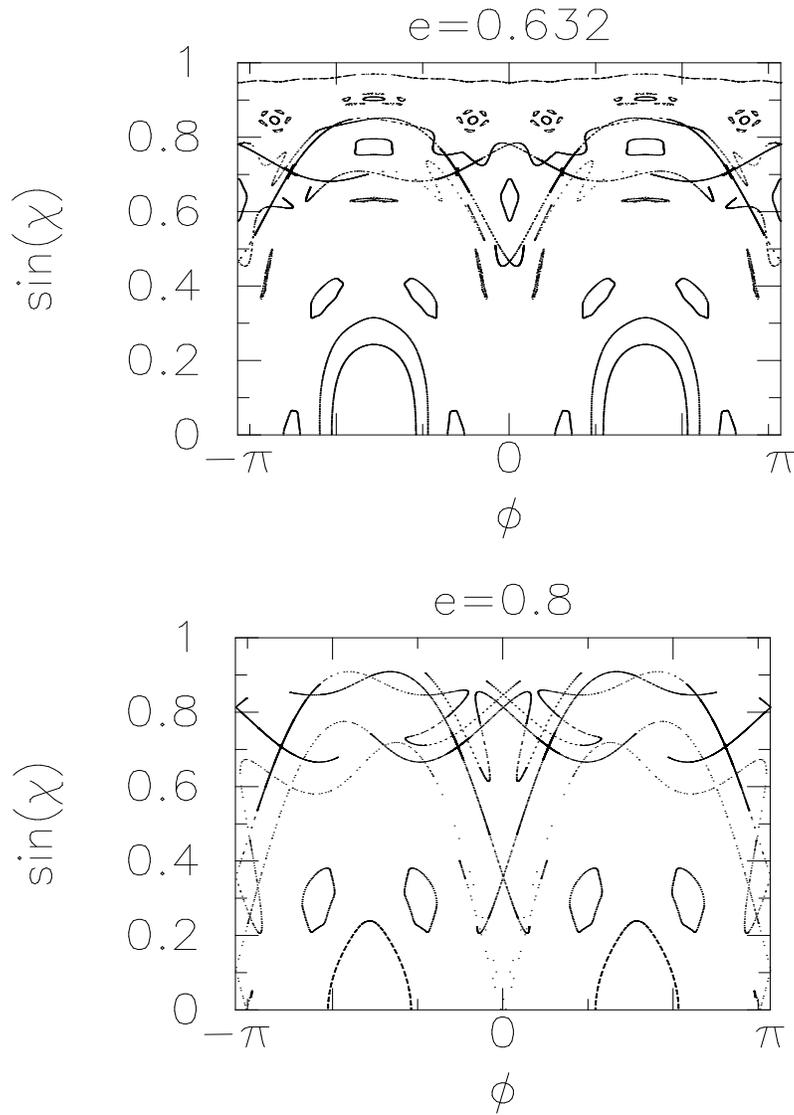


Figure 6.6: Poincaré sections of the quadrupole, showing 10 iterates of a cloud of trajectories launched close to each unstable fixed point of winding number 4. Both directions of time are plotted to obtain the stable and unstable manifolds. Some stable trajectories are also plotted to illustrate the phase space topology. Compare to the schematic in Fig. 4.4.

The location of the unstable four-bounce orbit in the SOS can be calculated easily at any deformation because symmetry implies that it always corresponds to a rectangle inscribed into the quadrupole. The latter means that $\sin \chi = \sin(45^\circ) = 1/\sqrt{2} \approx 0.7$ at all four reflections, independently of the deformation. This orbit is thus located in a region of phase space that has to be traversed by trajectories that are diffusing chaotically to an escape window corresponding to realistic indices of refraction ($n = 1.4 \dots 2$). At $e = 0.632$, we see that the turnstiles made up of the interweaving lobes are small in area and concentrated near the fixed points, thus forcing trajectories to spend long times in the separatrix region before they enter the vicinity of the islands from above and then eventually leave it toward the bottom. The transport is more efficient at $e = 0.8$ where the four-bounce islands have in fact disintegrated already. What is most notable at large deformations is the pronounced funneling of downward-moving orbits along the V-shaped intersection of manifolds around $\phi = 0$ and $\phi = \pi$. Clearly this implies that the motion is not random even at this large eccentricity, at least for 10 iterations, and that downward diffusion occurs in certain well-defined preferred directions along the ϕ axis.

6.6.3 Stable islands

The last remark also holds for islands of stability, which cause perhaps the most obvious breakdown of the diffusion from one adiabatic invariant curve to the next. The reason is that trajectories cannot stray from the chaotic domain into an island or *vice versa*, since each island is the intersection with the SOS of a phase space torus that has a noncommunicating inside and outside, as we argued previously for the KAM curves. This is why a chaotically diffusing orbit must avoid the islands. However, the adiabatic curves do not reproduce any island structure except that corresponding to the bouncing-ball orbit (see the ellipse), so that these approximate curves $p(\phi)$ can intersect the islands of the actual SOS and thus fail to describe the true motion in that case.

Stable islands are surrounded by intersecting manifolds from the accompanying hyperbolic points, cf. the above subsection. However, these may not be the dominant factor governing the phase space flow, since the islands are also surrounded by a transition region which contains cantori and smaller island chains. Trajectories may be captured by these satellite structures and rotate around the main island for long times before finding a way out. If the main islands form a chain of order q , then the rotation around one member is seen by plotting only every q -th intersection with the SOS, because the intermediate points lie near the other islands.

We will see later how islands acting as transport barriers can have a direct influence on the emission directionality of dielectric resonators.

Chapter 7

Semiclassical quantization

7.1 Einstein-Brillouin-Keller quantization on fuzzy caustics

At this point it is useful to recall our goal of describing resonant states with ray optics. The connection between a given resonance and the ray picture is made using a semiclassical approximation. The approach is described in this chapter for closed systems, and the necessary modifications in open cavities are discussed later in section 10.1. The aim is to obtain an expression for the eigenenergies or wavenumbers in terms of classical paths, applicable to the strongly deformed billiard where WG orbits can diffuse classically instead of moving on exact invariant curves.

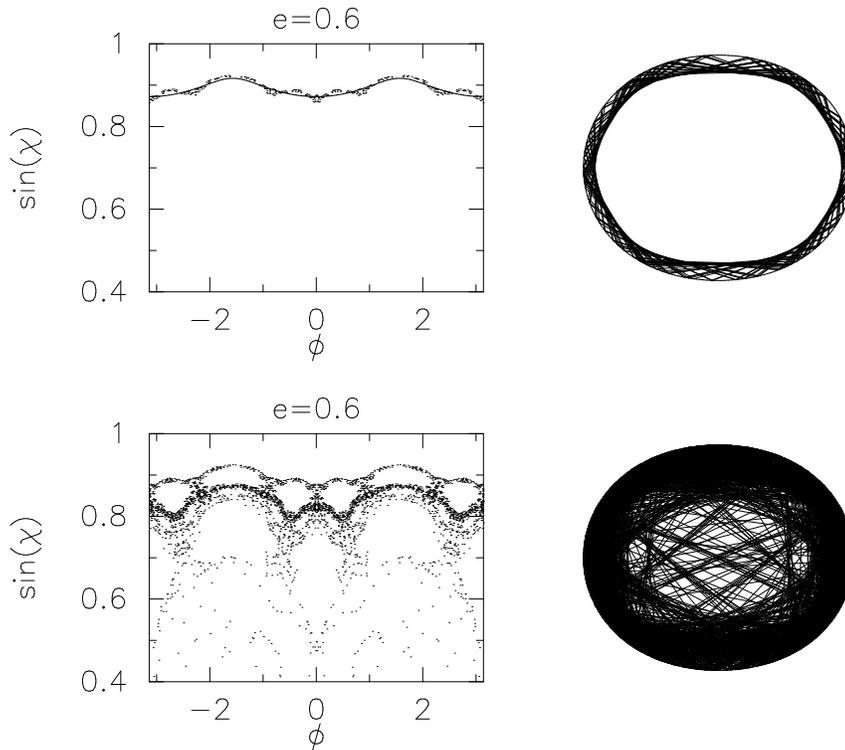


Figure 7.1: Top: Trajectory following the adiabatic curve (solid line) for 200 reflections (left), and corresponding real-space trajectory (right), exhibiting a “fuzzy” caustic. The deformation is quadrupolar. Bottom: same trajectory followed for 2000 bounces.

The billiard systems of interest to us are generically in the KAM regime,

which means that there is an appreciable portion of phase space for which the classical motion is on invariant tori. The general problem of semiclassically quantizing a mixed phase space is as yet unsolved, and a status report on this still evolving subject can be found in Ref.⁵¹. What is comparatively well-understood are the limiting cases of fully integrable and fully chaotic dynamics. The semiclassical theory of the former is based on the WKB approximation as reviewed by Berry and Mount¹⁰⁰. For chaotic systems, the theory of periodic orbits, based on Gutzwiller's trace formula for the density of states, has been the key approach to a semiclassical quantization³⁸. The benefits of trace formulas, and their shortcomings in particular for the treatment of WG modes, are discussed later in appendix B.

We propose an alternative approach that makes use of the slow classical diffusion in $\sin \chi$. The description of chaotic trajectories by adiabatic curves, discussed above, is used to apply the Einstein-Brillouin-Keller quantization procedure. This use of the adiabatic curves is novel, and its value lies in the fact that it replaces an infinite sum over periodic orbits by only two action integrals appearing in a system of two simultaneous equations. The unknowns determined from this system are k (the wavenumber of the bound state) and \tilde{p} (the parameter characterizing the adiabatic curve).

Most underlying proofs will be omitted in this chapter, and we instead try to illuminate the subject from an intuitive angle. A more detailed account of the method and some relevant derivations relying on theorems from classical mechanics can be found in appendix A.

The Einstein-Brillouin-Keller quantization scheme is based on the existence of caustics for all (non-periodic) trajectories, i.e. the absence of chaos (see appendix A). We can use it provided the adiabatic curves are a good description of the ray dynamics for intermediate times, because then the motion is on "almost" invariant curves as discussed above. In real space, the latter implies that the trajectory forms "fuzzy" caustics. This is illustrated in Fig. 7.1, where we also show the disintegration of the fuzzy caustics after long times. The existence of caustics is used in the quantization procedure to guarantee that one can write the semiclassical wavefunction in the WKB approximation with two and only two terms⁹⁹:

$$\psi(\mathbf{r}) = A_1(\mathbf{r}) e^{ik S_1(\mathbf{r})} + A_2(\mathbf{r}) e^{ik S_2(\mathbf{r})}, \quad (7.1)$$

where A_i are amplitudes and S_i are the phases, also called eikonals in optics. The classical rays propagate perpendicular to the phase fronts. In classical mechanics, a particle moves with constant momentum $p = \hbar k$ along the direction ∇S_i . Therefore, $\hbar k S_i$ is just the action function of Eq. (3.70), satisfying $\hbar k \nabla S_i = \mathbf{p}$. By definition, ∇S_i is then a unit vector along a ray direction.

We need the two terms above because there are exactly two possible ray directions through every point \mathbf{r} between the caustic and the boundary, assuming one only admits counterclockwise propagation around the billiard, see Fig. 7.2. The situation is illustrated for the simple case of the circle, where the caustic is also a circle.

The two possible trajectories can be labeled by their direction: one goes through \mathbf{r} on the way from the caustic to the boundary (abbreviated as CB), the other (BC) *vice versa*. Only the corresponding parts of the straight line trajectory between B and C have been drawn. If we now select for all \mathbf{r} *either* the BC or the CB ray, this effectively defines a field on the annulus, as shown in Fig. 7.3. As is clear from inspection of the figures, these fields are irrotational between C and B , even though they certainly display a vorticity which however circulates around the

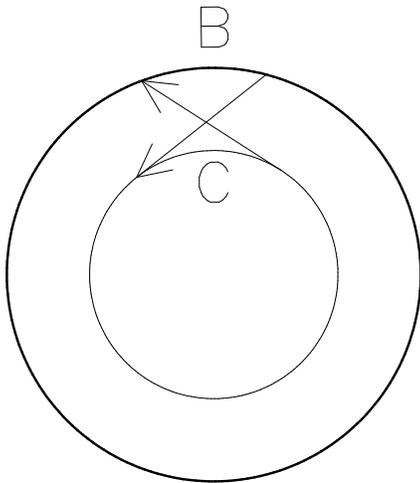


Figure 7.2: Two possible counterclockwise trajectories through an arbitrary point in the annulus between caustic (C) and boundary (B). They are uniquely determined by the requirement of tangency to the caustic.

caustic itself, not around a point inside the annulus defined by B and C . Therefore, the ray fields can in fact be written as the gradient of a scalar function. The two corresponding functions are just the eikonals S_1 and S_2 . This discussion makes it clear that one needs both wavefronts S_1 and S_2 to describe a complete ray segment between two successive reflections, since each S_i is associated only with the BC or CB portions, respectively. The switching between S_1 and S_2 occurs at the boundary or at the caustic, and one can derive relations between the two eikonals on these lines. For example, S_1 and S_2 have to differ exactly by π at B if one imposes Dirichlet boundary conditions, and there is a jump by $\pi/2$ at C . Since ∇S_i is a unit vector along the BC or CB ray direction, the phase advance along any ray path is

$$k \Delta S = k \oint_{\Gamma} \nabla S_i(\mathbf{r}) \cdot d\mathbf{r} - \alpha - \mu \frac{\pi}{2}. \quad (7.2)$$

where i in S_i is 1 or 2 depending on which type of ray segment is being followed during different parts of the contour Γ . The constants α and μ take into account the phase offsets due to each switching between S_1 and S_2 . The *Maslov index* μ counts the number times C crosses a caustic, and α is the boundary phase shift. The particular values these constants take on are discussed in appendix A. These two phase shifts subtract from the phase advance accounted for by the action integral, provided that the corresponding integration loop runs in the direction of increasing phase (S) before and after the crossings with caustic and boundary. Otherwise, the phase shifts enter with the opposite sign.

The value of the integral above is path-independent for each segment using a single S_i , simply because the CB and BC rays form a gradient field on the annulus between B and C . One can in particular consider *closed* contours Γ , which however include switching between the S_i , or encircle the caustic fully. In either of these cases, the integrals are *not* zero as they would be otherwise. To be closed, these integrations must start and end with the same eikonal, say S_2 . The nonvanishing phase advance upon return to the starting point then implies that S_2 (and likewise S_1) is multi-valued, in the same way that the value of the complex square root $z^{1/2}$ depends on how many times one winds around the origin. Recalling the interpretation of $\hbar k S_i$ as the action function of classical mechanics, we observe that any closed loop Γ in

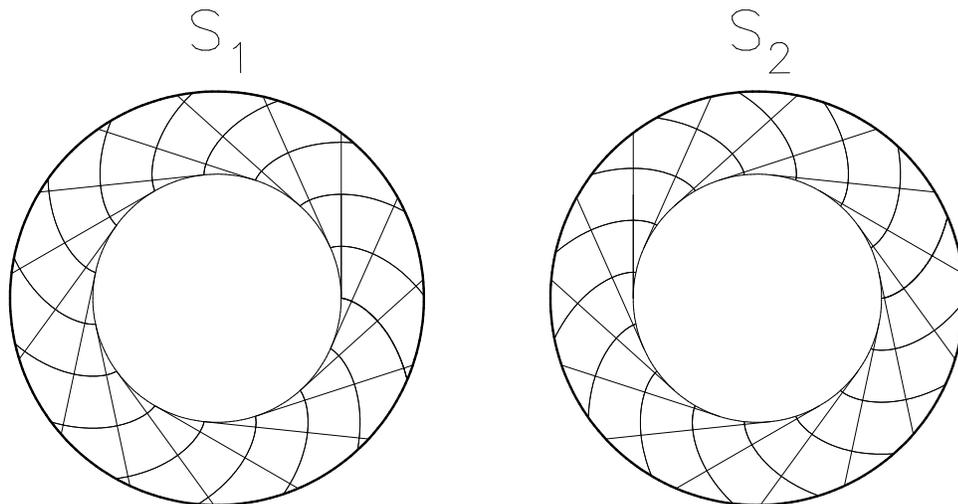


Figure 7.3: The two fields defined by the CB and BC rays, respectively. Superimposed are the wavefronts S_1 and S_2 to which the families of rays are perpendicular.

Eq. (7.2) will yield precisely one of the action variables, Eq. (3.50), up to the factor \hbar .

7.2 Quantization conditions

The quantization conditions now follow from the requirement that the wavefunction ψ be single-valued, i.e. the phase advance along any closed loop must be an integer multiple of 2π so that it drops out in Eq. (7.1). In terms of the classical actions, this leads to

$$J_\nu/\hbar = 2\pi n_\nu + \alpha_\nu + \mu_\nu \frac{\pi}{2}. \quad (7.3)$$

Here, n_ν are the (integer) *quantum numbers*. This is the central semiclassical formula to be used in our work, and it is also called EBK quantization after Einstein, Brillouin and Keller. It was Einstein¹⁰³ who proposed in 1917 how one could generalize the well-known Bohr-Sommerfeld quantization to include integrable systems where a complete separation of variables is not achieved. Brillouin¹⁰⁴ derived quantization conditions that also apply to nonseparable systems but are incorrect in that they do not contain the Maslov indices. Keller¹⁰⁵ gave the final form of this “torus quantization”.

In generic convex deformations of the circle, this procedure is hard to implement even if one has realized the existence of “fuzzy” caustics. The reason is that one still does not in general know the eikonals S_i , nor their gradients. Our solution to this problem is to employ a particular choice of integration contours Γ for which the integrand is known. Unlike previous work⁹⁹, the integrals then require no knowledge whatsoever of the real-space shape of the caustic.

This is possible because we possess the explicit form of the adiabatic curves in the Poincaré section, see Eq. (2.1), reproduced here for convenience:

$$p = \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}}, \quad (7.4)$$

which is of use provided that one can neglect the effects of island structure. This is the basic idea behind our approach to a semiclassical quantization of the whispering gallery orbits. *We rely on the existence of a slowly diffusing classical action*

parametrizing an approximate torus in phase space and an approximate caustic in real space, in analogy to the ellipse billiard, where Eq. (7.4) is exact, cf. Eq. (3.136).

Noting that

$$\nabla S_i \cdot d\mathbf{r} = \pm \sin \chi ds, \quad (7.5)$$

where ds is an arc length element along the curve Γ and χ the angle between ∇S and the normal to Γ , we can replace $\sin \chi$ in the integrand by the adiabatic formula for all parts of Γ that coincide with (or are infinitesimally close to) the boundary. Then $\sin \chi$ is identical to the $\sin \chi$ of the Poincaré section. The first contour thus chosen is the boundary itself.

The second path is depicted for the circle in Fig. 7.4. The loop is drawn

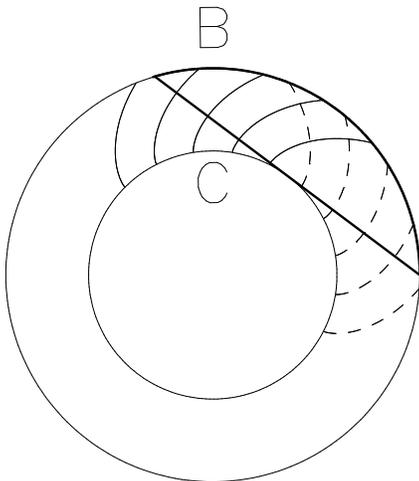


Figure 7.4: The second integration path, depicted for simplicity in the case of the circle. The same topology can be employed in deformed cavities.

as the thicker line. Along the ray, we first have to use S_2 to get from B to C , cf. Fig. 7.3. The corresponding wave fronts are drawn dashed. Following the ray from C back to B , we use S_1 whose wavefronts are drawn solid. Along the boundary, we return to the starting point by moving in a clockwise sense with respect to the caustic. At some point one also has to switch back from S_1 to S_1 to complete the loop. Where on B this happens is immaterial, as long as one takes into account that clockwise motion is by definition the direction of decreasing phase, yielding a minus sign for this part of the integration. There is a phase jump due to the encounter with C where S_2 connects to S_1 , and a second one due to the switching from S_1 to S_2 at B . The ray segment is fixed by the value of the adiabatic curve $\sin \chi(\phi)$ at the point of intersection with the boundary, which has been chosen in the figure to be at $\phi = 0$. Therefore the only unknowns in the problem are the wavenumber k in front of the phase integrals, and the parameter \tilde{p} of the adiabatic curve that is to be used in the integrand and for the determination of the ray segment.

We do not have to do any work at this point to write out the resulting quantization conditions, because the two integrals have already been considered in section 3.8.3 where we treated the ellipse. Using \tilde{p} as a parameter (instead of K), Eqs. (3.138) and (3.140) read

$$J_\phi = \sqrt{2mE} \int_0^{2\pi} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}}, \quad (7.6)$$

$$J_r = \sqrt{2mE} \left[L(\tilde{p}) - \int_{\phi_0}^{\phi_1(\tilde{p})} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}} \right] \quad (7.7)$$

We only have to add the Maslov and boundary phase shifts in accordance with the above description of the contours. The caustic and boundary are crossed once each for J_r , and not at all for J_ϕ . Therefore, we arrive at

$$J_\phi = 2\pi \hbar m, \quad (7.8)$$

$$J_r = 2\pi \hbar \left(n + \frac{3}{4} \right). \quad (7.9)$$

From this system of equations, one unknown can be eliminated immediately if one divides Eq. (7.9) by Eq. (7.8), with the result

$$\begin{aligned} & m \left[L(\tilde{p}) - \int_{\phi_0}^{\phi_1(K)} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}} \right] \\ &= \left(n + \frac{3}{4} \right) \int_0^{2\pi} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - (1 - \tilde{p}^2) \kappa(\phi)^{2/3}}. \end{aligned} \quad (7.10)$$

For any given integers n and m , this equation can be solved for \tilde{p} , using a numerical root finding routine. This provides us with the quantization of the adiabatic invariant, $\tilde{p}_{m,n}$. The quantized energy levels are then obtained by substituting $\tilde{p}_{m,n}$ into Eq. (7.8), which yields

$$k_{m,n} = \sqrt{2mE}/\hbar = \frac{2\pi m}{\int_0^{2\pi} d\phi \sqrt{r^2 + \dot{r}^2} \sqrt{1 - (1 - \tilde{p}_{m,n}^2) \kappa(\phi)^{2/3}}}. \quad (7.11)$$

In the special case of the circle, the adiabatic curves are simple $\sin \chi = \tilde{p}$, and $\kappa \equiv r \equiv 1$ (if $R = 1$), so that we obtain from Eq. (7.11) the *semiclassically exact* expression for the angle of incidence,

$$\sin \chi = \frac{m}{k_{m,n} R}, \quad (7.12)$$

if $k_{m,n}$ is the semiclassically quantized wavenumber. The integers m, n reduce to the angular momentum and radial quantum numbers in the circle. To get the quantized wavenumber, one needs to solve Eq. (7.10) first. We defer this calculation to appendix A.2.

We can also easily verify that for the deformed case, the EBK approach produces the correct result in the whispering-gallery limit, $\tilde{p}_{m,n} \rightarrow 1$. Since the curvature is finite, the result is

$$k_{m,n} = \frac{2\pi m}{\int_0^{2\pi} d\phi \sqrt{r^2 + \dot{r}^2}} = \frac{2\pi m}{L_0}, \quad (7.13)$$

where L_0 is the circumference of the boundary, in agreement with the result of Keller and Rubinow quoted in Eq. (A.18), neglecting the correction term. It should be mentioned that this relation implies a *red shift* with increasing circumference at

constant m . This situation is encountered in all the deformations we consider, since they are designed to maintain a constant area and hence increase L_0 .

The important point of this quantization based on the adiabatic curves is that the solution of Eq. (7.10) automatically yields the parameter \tilde{p} of the adiabatic curve that corresponds to the quantized state (m, n) . This curve then can serve as the desired initial condition for an ensemble of classical rays whose long-time diffusion toward lower $\sin \chi$ eventually causes classical escape and hence deterioration of the resonance lifetime. In this way we are able to make quantitative predictions for the resonance lifetime at strong deformation where classical escape is the dominant decay mechanism.

In order to take corrections due to tunneling and above-barrier reflection into account, we again have to find a way of connecting these wave phenomena with classical quantities that can be included in a ray simulation. This will be done in chapters 8 and 10.

7.3 Reliability considerations

The quality of this approximate semiclassical quantization depends on how well the adiabatic invariant curves describe the motion. We will return to this question when we present the results of this approach, but at this point it is worth mentioning that a similar situation arises in the context of the quantization of hyperbolically unstable periodic orbits¹¹⁴ in the discussion of scarred wavefunctions in strongly chaotic systems. In spite of a positive Lyapunov exponent in their neighborhood, these orbits are found to produce an imprint (“scar”)⁴² on the wavefunctions, both at low *and* high energies in the spectrum. This can be explained if one accepts the notion of “adiabatic instability”¹¹⁵, which is a weaker criterion than Lyapunov instability. It means that trajectories stay close to a hyperbolic periodic orbit for at least one real-time period, and this has been shown¹¹⁶ to be sufficient for an approximate semiclassical quantization.

In our case, the whispering-gallery orbits are not periodic in general, but the above remarks are encouraging as to the potential value of the adiabatic quantization. One can draw additional support for this hope from the work of Bohigas *et al.*¹¹⁷ which aims at a statistical classification of the spectra of mixed systems. There, it is found that partial barriers to classical phase space transport, as discussed previously, can cause the quantum Hamiltonian to be almost block-diagonal with the separate blocks being semiclassically quantized on subregions of the chaotic domain that are only weakly connected classically. The mixing of the blocks increases both with the classical transport and with decreasing \hbar . The latter means that far from the classical $\hbar \rightarrow 0$ limit, the effect of classical diffusion will be less significant for the quantization of the system.

Chapter 8

The wave equation for symmetric resonators

Up to now we have treated the classical and semiclassical analysis of nonintegrable closed systems, in particular billiards with hard wall boundaries. This has been necessary to lay the groundwork for the investigations that gave rise to the title of this thesis. Before we can turn to the problem of nonintegrable open resonators, we shall derive some necessary results concerning the *integrable* open cavity of circular shape.

We have in mind an application of concepts from nonlinear dynamics to the study of dielectric optical resonators, and it turns out that it is helpful to introduce terminology known from quantum mechanics (such as tunneling) in discussing the optical systems. At first sight it may seem trivial to do this, in view of the fact that electromagnetic waves have been used extensively to simulate closed quantum systems at microwave frequencies^{118–121}. However, the analogy between quantum mechanics and the scalar wave equation of optics is not complete if one has to deal with finite potentials or refractive indices, respectively, that are not constant in space. A homogenous dielectric in air is of this type, and it corresponds to an attractive potential well of finite depth, but only with some caveats to be discussed later. It must also be kept in mind that one has to take the polarization of electromagnetic waves into account. Fortunately, polarizations can be chosen such that they are preserved (not mixed) by the geometry to be discussed now.

8.1 Electromagnetic waves in the circular cylinder

We consider an insulating circular cylinder of refractive index $n > 1$ surrounded by vacuum, i.e., in polar coordinates

$$n(r) = \begin{cases} n & (r < R) \\ 1 & (r > R) \end{cases}. \quad (8.1)$$

This is a problem that is treated in textbooks, e.g. Refs.^{122,123}, but it will be necessary for us to expand significantly on the treatment of resonances found there. The interpretation we shall give is closely related to the work by Johnson concerning resonances in spherical dielectrics¹²⁴, which have been of greater applied interest than cylinders and were therefore studied more extensively. Since in this thesis solutions of the wave equation for deformed spheres were not attempted, we do not reproduce all the known results for spherical dielectrics but instead fill the existing gaps for the cylindrical case, in preparation for the full wave solutions of deformed

cylinders which will be presented later. We will also need the results for the circular cylinder as an ingredient in the ray dynamics simulations that are introduced later to understand resonance widths in deformed cavities.

8.1.1 Metastable well in the effective potential

Let us first see how one arrives at an effective potential with a centrifugal barrier from Maxwell's equations. Assuming all fields to have the time dependence $e^{-i\omega t}$, Maxwell's equations become

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} = ik\mathbf{H} \quad (8.2)$$

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} = -ikn^2\mathbf{E}, \quad (8.3)$$

where the wave number is given by $k = \omega/c$. Combining these equations, we get the wave equations

$$\nabla \times \nabla \times \mathbf{E} = (nk)^2\mathbf{E}, \quad (8.4)$$

$$\nabla \times \nabla \times \mathbf{H} = (nk)^2\mathbf{H}. \quad (8.5)$$

Since charge density can only appear at the surface of the dielectric, we have $\nabla \cdot \mathbf{E} = 0$ in each domain of constant n .¹ In these regions, Eq. (8.4) therefore becomes

$$-\nabla^2\mathbf{E} = (nk)^2\mathbf{E}. \quad (8.6)$$

Following Johnson¹²⁴, we can rewrite this in a form similar to the Schrödinger equation,

$$-\nabla^2\mathbf{E} + k^2(1 - n^2)\mathbf{E} = k^2\mathbf{E} \quad (8.7)$$

This suggests that regions with $n > 1$ correspond to an attractive potential, except that the potential is itself multiplied by the eigenvalue k^2 .

Let us now discuss how such a purely attractive potential can give rise to resonances that are so long-lived that they can be the building blocks of lasers and other important devices. When classical particles scatter off a short-ranged, purely attractive potential in two dimensions, their motion is accelerated as they approach the target, and they slow back down to their initial velocity as they escape to infinity. For particles launched inside the potential well, truly bound states exist even if the energy is everywhere larger than the scattering potential, simply due to conservation laws that exist in two (or three) dimensions, such as angular momentum conservation. In wave mechanics, these states acquire a finite lifetime, but they still create extremely sharp resonances because their decay is classically forbidden and proceeds via tunneling. In one dimension, one finds oscillations in the transmission coefficient of an attractive well due to above-barrier resonances, but the corresponding lifetime is short.

We illustrate the effect of rotational symmetry for the wave equation, Eq. (8.7), by performing a separation of variables in polar coordinates. The radial equation then reads

$$-\left[\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} \right] \mathbf{E}(r) + V_{eff}(r) \mathbf{E}(r) = k^2 \mathbf{E}(r), \quad (8.8)$$

¹If \mathbf{E} is parallel to the z axis, then the divergence vanishes even at the interface because E is not a function of z . This is the case we focus on.

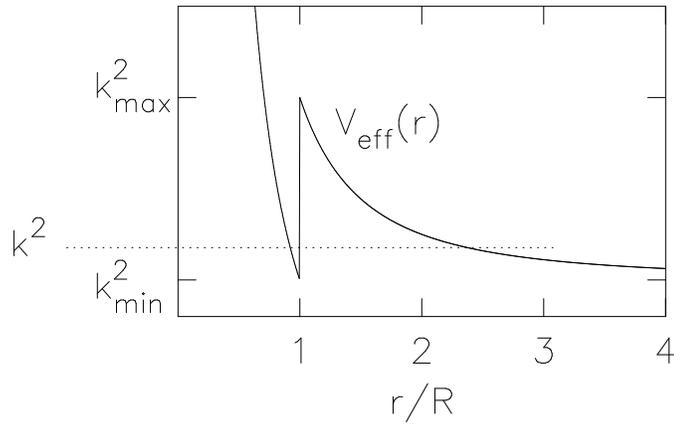


Figure 8.1: Effective potential picture for whispering gallery resonances; $k_{min} = m/(nR)$ and $k_{max} = m/R$.

where the effective potential is

$$V_{eff}(r) = k^2 \left(1 - n^2(r)\right) + \frac{m^2}{r^2}. \quad (8.9)$$

Here, an additional centrifugal potential term has appeared as a consequence of angular momentum conservation. The resulting sum of the attractive well due to the dielectric and the repulsive angular momentum barrier is shown in Fig. 8.1. The effective potential is seen to form a metastable well. This interpretation requires some additional thought, however, because of the fact that the kinetic energy term in the radial equation, Eq. (8.8) involves a first derivative so that it is not in the form of a one-dimensional Schrödinger equation. This can be corrected by using the substitution $\xi := \ln(kr)$ to obtain the normal form ²

$$\frac{d^2}{d\xi^2} \mathbf{E}(\xi) + q^2(\xi) \mathbf{E}(\xi) = 0 \quad (8.10)$$

with an “effective wavenumber”

$$q(\xi) := \sqrt{n^2 e^{2\xi} - m^2}. \quad (8.11)$$

Now one can see that the field will be oscillatory whenever $q^2(\xi) > 0$, and show decaying behavior otherwise. In the original variable r , one has

$$q^2(r) = (nkr)^2 - m^2 = r^2 \left(k^2 - V_{eff}(r)\right), \quad (8.12)$$

so that oscillatory, “classically allowed”, solutions exist whenever k^2 exceeds the effective potential. This justifies the interpretation of V_{eff} as defining a well separated from the propagating solutions outside by a tunnel barrier. The depth of the well is

$$k_{min}^2 \equiv V_{eff}(r = R^-) = k_{min}^2 \left(1 - n^2\right) + \frac{m^2}{R^2} \quad (8.13)$$

$$\Rightarrow k_{min}^2 = \frac{m^2}{(nR)^2}, \quad (8.14)$$

²Alternatively, one could decompose $\mathbf{E} = \psi/\sqrt{r}$ and obtain a Schrödinger-like equation for ψ (similar to the textbook approach for the Hydrogen problem), but this introduces an additional term in the effective potential, tantamount to replacing m^2 by $m^2 - 1/4$, which however makes no difference in the large- m limit. The method chosen here is more convenient for our illustrative purposes, and the resulting variables are closely related to the elliptical coordinates at large radius, which in turn are useful in the deformed case.

and the height of the barrier top is

$$k_{max} = \frac{m}{R}. \quad (8.15)$$

Consequently, we expect narrow resonances for wavevectors satisfying

$$m/(nR) < k < m/R, \quad (8.16)$$

because that places the metastable level below the top of the barrier, so that decay can only occur by tunneling. States at the lowest possible k for a given m will be concentrated most strongly near the interface $r = R$. Note that the tunnel barrier gets higher with increasing angular momentum m , which is a first indication that the whispering gallery modes propagating close to the dielectric interface have especially long lifetimes. If the refractive index is changed, the barrier top remains the same, but the well grows deeper, so that the lowest allowed k will give rise to a narrower resonance. It will be shown later that the metastable states satisfying Eq. (8.16) are discrete in the sense that their spacing in k is larger than their decay width through the barrier. This brings with it useful similarities to truly bound states that one would obtain by making the barrier impenetrable.

8.1.2 Boundary conditions

Our considerations so far have not addressed the boundary conditions at the interface between dielectric and vacuum (or air), which in optics depend on the polarization of \mathbf{E} . The above arguments for the existence of narrow resonances are, however, independent of the matching conditions because we rely only on the fact that a tunnel barrier is formed where the field has to decay. For this reason, we choose the polarization such that the matching conditions are simplest. This is achieved by requiring \mathbf{E} to be parallel to the cylinder axis $\hat{\mathbf{z}}$, with the propagation direction perpendicular to the cylinder. One denotes this as TM polarization because the magnetic field is transverse to the cylinder axis. For a dielectric insulator without any current flows, \mathbf{H} must be continuous everywhere³. From Eq.(8.2), one has in polar coordinates for the azimuthal component of the magnetic field

$$H_\phi = -\frac{i}{k} [\nabla \times E\hat{\mathbf{z}}]_\phi = \frac{i}{k} \frac{\partial E}{\partial r}. \quad (8.17)$$

The tangential components of \mathbf{E} are also continuous at the interface, which for TM polarization applies to E itself. Therefore, both E and $\partial E/\partial r$ must be continuous at the interface. This shows that TM polarization leads to the same matching conditions for E as in quantum mechanics. TE polarization (\mathbf{H} along $\hat{\mathbf{z}}$) does not result in this familiar requirement, and we are going to use only TM polarization in what follows.

8.1.3 Comparison to quantum mechanics

Having arrived at the common quantum mechanical matching conditions, we still have not achieved full equivalence between optics and quantum mechanics. The reason is of course that V_{eff} depends on k . For the time-independent Schrödinger equation with a constant attractive potential

$$V(r) = \begin{cases} -V_0 & (r < R) \\ 0 & (r > R) \end{cases} \quad (8.18)$$

³We neglect any variation in the magnetic permeability

we obtain, similarly to the above treatment, the quantum mechanical effective potential in the radial equation,

$$V_{qm}(r) = V(r) + \frac{\hbar^2 m^2}{2Mr^2}, \quad (8.19)$$

where M is the mass. The metastable well formed here is similar to that in Fig. 8.1, except that the depth of the well remains fixed as we change the energy of the particle. In the optics case, the well actually gets deeper if we raise k . We will return to this effect later as an explanation for the Fresnel reflection formula.

The upper limit on $k = \sqrt{2mE}/\hbar$ for the existence of long-lived resonant states is again $k = m/R$, as in the optical counterpart.

8.1.4 Total internal reflection

We can attach a well-known physical interpretation to the condition $k < m/R$ for narrow resonances, by making use of the exact semiclassical relation Eq. (7.12), which is also valid in the open system, except that k is now the discrete resonance wavevector of the light outside the cavity, corresponding to nk inside. Therefore, the relation for the angle of incidence is

$$\sin \chi = \frac{m}{nkR}. \quad (8.20)$$

Inserting this into Eq. (8.16), m is eliminated in favour of $\sin \chi$ to yield

$$1 > \sin \chi > 1/n. \quad (8.21)$$

The first inequality is trivial, but the second (corresponding to the barrier top condition) is precisely the relation a ray has to satisfy if it is to undergo total internal reflection. When $\sin \chi = 1/n$, there appears a refracted ray in the optically thin medium on the outside, tangent to the interface. Total internal reflection would imply that the below-barrier resonant states are in fact stable, not metastable. The reason they do escape is that the surface has a finite curvature, while total internal reflection requires a plane interface. The realization that total internal reflection is the classical reason behind the long lifetimes of modes in the symmetric dielectric cavity is going to be one of the key elements in our treatment of asymmetric cavities.

8.2 Quasibound states

We can answer all questions about the elastic scattering of light by a dielectric cylinder if we know the scattering states,

$$E(r) = \begin{cases} A_m J_m(nkr) & (r < R) \\ H_m^{(2)}(kr) + S_m H_m^{(1)}(kr) & (r > R) \end{cases}, \quad (8.22)$$

for all m . These are solutions to the radial equation, Eq. (8.8), with the refractive index given by Eq. (8.1), provided that we determine the coefficients such that they satisfy the matching conditions for E and its derivative, E' ,

$$\begin{aligned} A_m J_m(nkR) &= H_m^{(2)}(kR) + S_m H_m^{(1)}(kR) \\ A_m nJ'_m(nkR) &= H_m^{(2)'}(kR) + S_m H_m^{(1)'}(kR). \end{aligned} \quad (8.23)$$

This can be solved for the scattering amplitude S_m as a function of the size parameter $x \equiv kR$. In terms of quantum mechanical scattering theory, S_m is an element

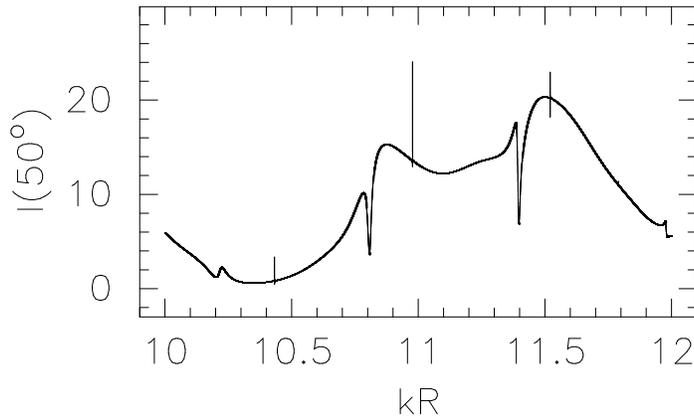


Figure 8.2: Intensity scattered off a circular cylinder at 50° with respect to the direction of the incoming wave, which is assumed to be plane and TM polarized.

of the S-matrix which for the rotationally invariant scatterer is diagonal in the basis of angular momentum states m . Using these solutions, one can for example calculate the scattering properties for an incident plane wave, because the latter can be decomposed into a combination of Bessel functions, thus prescribing the coefficients with which to superimpose the scattering states (8.22). In this way, one can obtain a plot like the one shown in Fig. 8.2. The resonances are of the Breit-Wigner or Fano shape as long as they are sufficiently isolated, and it is then a simple matter to extract their widths. What can immediately be observed is that there appear to be two distinct classes of resonances, namely broad ones and very narrow ones. The latter will be seen to be precisely the whispering-gallery modes confined by total internal reflection.

Resonances are directly connected with quasibound states or poles of the S-matrix. To determine the position and width of a resonance without recourse to fits of the scattering line shapes, one can determine real and imaginary part of the pole position in the complex k plane. This is done by recalling that a pole of the S-matrix corresponds to the possibility of having an outgoing wave in the absence of an incoming wave. Therefore we have to find the complex k at which there exists a nontrivial solution to the homogenous part of the linear system (8.23), which is tantamount to searching for the zeroes of the determinant

$$D = \begin{vmatrix} J_m(nkR) & -H_m^{(1)}(kR) \\ nJ'_m(nkR) & -H_m^{(1)'}(kR) \end{vmatrix}. \quad (8.24)$$

This leads to

$$-J_m(nkR) H_m^{(1)'}(kR) + nJ'_m(nkR) H_m^{(1)}(kR) = 0. \quad (8.25)$$

Using the recursion relations for the Bessel functions to eliminate the derivatives, we arrive at the resonance condition

$$n \left[\frac{m}{nkR} J_m(nkR) - J_{m+1}(nkR) \right] H_m^{(1)}(kR) \quad (8.26)$$

$$= J_m(nkR) \left[\frac{m}{kR} H_m^{(1)}(kR) - H_{m+1}^{(1)}(kR) \right] \\ \Rightarrow nJ_{m+1}(nkR) H_m^{(1)}(kR) = J_m(nkR) H_{m+1}^{(1)}(kR). \quad (8.27)$$

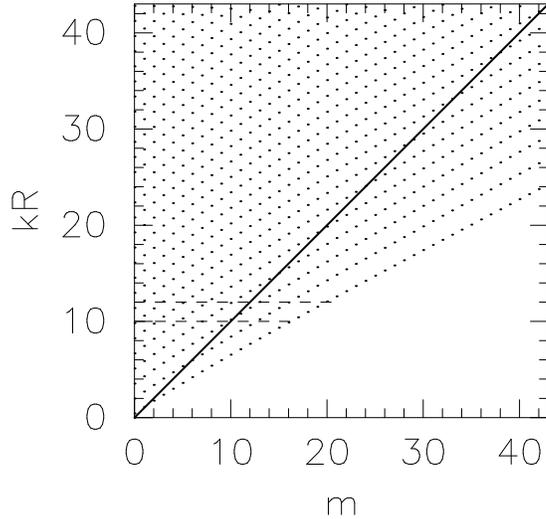


Figure 8.3: Resonance positions in the circle of refractive index $n = 2$ with TM polarization for the lowest 44 angular momentum numbers m . Each dot corresponds to one resonance, and the solid line $kR = m$ represents the dividing line between broad and narrow resonances, with narrow widths expected below it according to both the barrier picture and the total internal reflection interpretation. The horizontal dashed lines enclose a kR interval in which we count a total of 9 resonances below the critical line. Compare this observation with the scattered intensity in Fig. 8.2. The three narrowest resonances there correspond to the three lowest- kR points at $m = 17, 18, 19$. The closer the points are to the critical line, the broader the resonances get.

This can then be solved numerically to find the real and imaginary parts of kR at which a metastable state occurs. The results for the real parts are summarized in Fig. 8.3. Clearly, the discreteness of the points in this grid suggests that the resonant states can be characterized by two quantum numbers as in the closed circle. One is m , the other is the analog of the radial quantum number, and in the same way should count the number of radial nodes in the effective potential well.

Approximate analytical results can be obtained when the Bessel functions can be replaced by their asymptotic expansions. Here, one can make different approximations depending on whether k is below the barrier top of the effective potential or not. We now discuss the cases that arise, in order to make connections to important physical interpretations such as Fresnel reflection and semiclassical formulas for positions and widths of resonances.

8.3 Above-barrier resonances

According to Eq. (8.16), only broad resonances are possible in angular momentum channel m if $kR > m$. If kR is far above this threshold, we can use the large-argument expansion in Eq. (8.27) to obtain

$$n \cos \left(nkR - \frac{\pi}{2}m - \frac{3\pi}{4} \right) \exp i \left(kR - \frac{\pi}{2}m - \frac{\pi}{4} \right) \quad (8.28)$$

$$= \cos \left(nkR - \frac{\pi}{2}m - \frac{\pi}{4} \right) \exp i \left(kR - \frac{\pi}{2}m - \frac{3\pi}{4} \right). \quad (8.29)$$

Noting that $\cos(\alpha - \pi/2) = \sin \alpha$, the transcendental equation simplifies to

$$\tan \left(nkR - \frac{\pi}{2}m - \frac{\pi}{4} \right) = \frac{1}{n} e^{-i\pi/2} = -\frac{i}{n}. \quad (8.30)$$

This can be solved immediately for nkR , yielding

$$nkR = \frac{\pi}{2}m + \frac{\pi}{4} + \frac{i}{2} \left[\ln \frac{1 - 1/n}{1 + 1/n} - i2\pi l \right]. \quad (8.31)$$

Thus the real and imaginary parts, $kR = x + iy$, are

$$x = \frac{\pi}{n} \left[\frac{m}{2} + l + \frac{1}{4} \right] \quad (8.32)$$

$$y = \frac{1}{2n} \ln \frac{1 - 1/n}{1 + 1/n}. \quad (8.33)$$

It is interesting to note that the expansion for the real part of these broad resonances *coincides* with the semiclassical spectrum of a closed billiard with von-Neumann boundary conditions, cf. Eq. (A.20). The radial quantum number l runs from $0 \dots \infty$, because negative l would label a lower- m state. This shows explicitly what was already noted in the context of Fig. 8.3, namely that the resonant states can be characterized by discrete quantum numbers analogous to the closed system.

The resonance width y is independent of the size parameter x . For large enough refractive index n , we can expand the logarithm on the last line to find

$$y = -\frac{1}{n^2}. \quad (8.34)$$

This result is quite puzzling if one compares it with *quantum-mechanical* above-barrier reflection: Should the decay rate not increase indefinitely as kR is increased? That is, after all, what happens in quantum mechanics, because reflection becomes negligible when the energy is much larger than the barrier height. However in the optical system, inspection of Eq. (8.9) reveals that the effective potential jump at the barrier actually grows when kR is increased: the barrier top remains fixed, but the bottom of the well decreases, and this leads to a k -independent above-barrier reflection.

The fact that y is determined solely by the index of refraction jump at the surface suggests a simple interpretation in terms of Fresnel's formula for the reflectivity of a dielectric interface: For normal incidence at a plane interface, the reflection probability is (independently of polarization)

$$p_0 = \left| \frac{E_{reflected}}{E_{incident}} \right|^2 = \frac{(1 - n)^2}{(1 + n)^2}. \quad (8.35)$$

The limit $kR \gg m$ we are interested in means that the length over which the amplitude varies along the interface satisfies

$$\frac{2\pi R}{m} \gg \frac{2\pi}{k} = \lambda. \quad (8.36)$$

On the scale of the wavelength λ , both $2\pi R/m$ and R are very large so that the wave can be considered to be a plane wave normally incident on a flat interface. This agrees with the definition of $\sin \chi = m/nkR$ which yields $\chi \rightarrow 0$ in this limit. A ray bouncing along a diameter of the circular cross section travels for a time $2Rn/c$ between reflections. After ν bounces and a corresponding time of $t = 2Rn\nu/c$, the probability for still remaining in the cavity is

$$P(t) = p_0^\nu = \exp(\nu \ln p_0) \quad (8.37)$$

$$= \exp(tc \ln p_0 / 2nR) = e^{-2t/\tau} \quad (8.38)$$

with the decay rate

$$\frac{2}{\tau} = -\frac{c \ln p_0}{2nR}. \quad (8.39)$$

The factor of two stems from the fact that $P(t)$ describes the field intensity, which is proportional to E^2 , while we want τ to be the decay time of E itself. Inserting the expression for p_0 , we find

$$\frac{1}{\tau} = -\frac{c}{2nR} \ln \frac{1 - 1/n}{1 + 1/n}. \quad (8.40)$$

But the decay rate is related to the imaginary part of kR by

$$y = -\frac{R}{c\tau} = \frac{1}{2n} \ln \frac{1 - 1/n}{1 + 1/n}, \quad (8.41)$$

which precisely reproduces Eq. (8.33).

8.4 Below-barrier resonances

When $kR < m$, the quasibound states decay only by tunneling through the effective potential barrier. It is not possible to apply Fresnel's formulas in this case because the interface no longer looks plane to the incident wave. Since $\sin \chi$ exceeds the critical angle, we would expect total internal reflection for a plane interface, with no possibility for tunneling escape. Thus, the nonvanishing curvature is crucial for determining the finite resonance lifetime. In contrast to the broad above-barrier resonances, we now expect a strong k dependence of y .

To obtain an approximate expression for y , we return to Eq. (8.27) and expand it to first order in y . This will be justified if the resulting resonances are narrow. One obtains

$$iy \left[nJ_{m-1}(nx) H_m^{(1)'}(x) + n^2 J_{m-1}'(nx) H_m^{(1)}(x) \right. \\ \left. - J_m(nx) H_{m-1}^{(1)'}(x) - nJ_m'(nx) H_m^{(1)}(x) \right] \quad (8.42)$$

$$= J_m(nx) H_{m-1}^{(1)}(x) - nJ_{m-1}(nx) H_m^{(1)}(x). \quad (8.43)$$

Here we use the recursion relations for the derivatives in such a way as to obtain only orders m and $m - 1$. The factor multiplying iy then becomes (leaving out the arguments and superscripts)

$$nJ_{m-1} \left(H_{m-1} - \frac{m}{x} H_m \right) + n^2 \left(\frac{m-1}{x} J_{m-1} - J_m \right) H_m^{(1)} \\ - J_m \left(\frac{m-1}{x} H_{m-1} - H_m \right) - n \left(J_{m-1} - \frac{m}{x} J_m \right) H_m^{(1)}(x) \\ = \frac{1}{x} (J_m H_{m-1} - nJ_{m-1} H_m) + (1 - n^2) H_m J_m. \quad (8.44)$$

This leads to

$$iy = \left[\frac{1}{x} + (1 - n^2) \frac{H_m J_m}{J_m H_{m-1} - nJ_{m-1} H_m} \right]^{-1} \\ = \left[\frac{1}{x} + (1 - n^2) \frac{1}{H_{m-1}/H_m - nJ_{m-1}/J_m} \right]^{-1}. \quad (8.45)$$

Now x is not arbitrary, but instead fixed by the resonance condition. For Eq. (8.45), this condition means that x must be chosen so as to make the righthand side purely imaginary to linear order in y (since y on the lefthand side is real). However we do not use this condition to determine x because there is a more accurate way that does not require an expansion in y . Instead, we proceed to simplify the approximate expression for y as a function of x and m provided by Eq. (8.45). The ratio of Hankel functions is

$$\frac{H_{m-1}}{H_m} = \frac{J_{m-1} + iY_{m-1}}{J_m + iJ_m} = \frac{J_{m-1}J_m + Y_{m-1}Y_m}{J_m^2 + Y_m^2} + i\frac{J_mY_{m-1} - J_{m-1}Y_m}{J_m^2 + Y_m^2}. \quad (8.46)$$

Using this, we can rewrite Eq. (8.45) as

$$iy = \frac{1}{a + \frac{1}{c+id}} \quad (8.47)$$

where we defined

$$a \equiv \frac{1}{x} \quad (8.48)$$

$$c \equiv \frac{1}{1-n^2} \left[\frac{J_{m-1}J_m + Y_{m-1}Y_m}{J_m^2 + Y_m^2} - n \frac{J_{m-1}}{J_m} \right] \quad (8.49)$$

$$d \equiv \frac{1}{1-n^2} \frac{J_mY_{m-1} - J_{m-1}Y_m}{J_m^2 + Y_m^2}. \quad (8.50)$$

We can show immediately that $d \rightarrow 0$ when $y \rightarrow 0$ by decomposing Eq. (8.47) into real and imaginary parts:

$$iy = \frac{ac^2 + c + ad^2}{(ac+1)^2 + a^2d^2} + i\frac{d}{(ac+1)^2 + a^2d^2}. \quad (8.51)$$

The imaginary part has to vanish when $y \rightarrow 0$. To linear order in y Eq. (8.47) is in fact equivalent to

$$iy = id = i\frac{1}{1-n^2} \frac{J_m(x)Y_{m-1}(x) - J_{m-1}(x)Y_m(x)}{J_m^2(x) + Y_m^2(x)}. \quad (8.52)$$

To prove this, compare the imaginary parts of the last two equations to get the identity

$$d = \frac{d}{(ac+1)^2 + a^2d^2} \Rightarrow ac^2 + 2c + ad^2 = 0. \quad (8.53)$$

Using this, the real part of Eq. (8.51) simply becomes equal to $-c$, which through Eq. (8.53) can be expressed in terms of d as

$$-c = -\frac{1}{a} \left[\sqrt{1 - a^2d^2} - 1 \right] \quad (8.54)$$

$$\approx -x \left[1 - \frac{d^2}{2x^2} - 1 \right] \quad (8.55)$$

$$= \frac{d^2}{2x}. \quad (8.56)$$

This means that the real part is of second order in y if we set $y = d$, and hence Eq. (8.52) is correct to linear order in y .

Equation (8.52) is valid for small y ($y \ll 1$) and arbitrary angular momentum m . We now let m become very large at constant $m/x > 1$ (the inequality follows from the requirement of below-barrier resonance). If moreover $(m-1)/x > 1$ one can employ the asymptotic Debye expansion (also called approximation by tangents¹²⁵) for the Bessel functions:

$$J_m\left(m\frac{x}{m}\right) \approx \frac{1}{\sqrt{2\pi m \tanh\alpha}} \exp[m(\tanh\alpha - \alpha)], \quad (8.57)$$

$$Y_m\left(m\frac{x}{m}\right) \approx \frac{-2}{\sqrt{2\pi m \tanh\alpha}} \exp[-m(\tanh\alpha - \alpha)],$$

$$\tanh\alpha \equiv \sqrt{1 - \frac{x^2}{m^2}}. \quad (8.58)$$

These functions enter with indices m and $m-1$. The value of

$$\tanh\alpha' \equiv \sqrt{1 - \frac{x^2}{(m-1)^2}} \quad (8.59)$$

is very close to $\tanh\alpha$, and the difference can be neglected in the prefactors which then cancel in Eq. (8.52), yielding

$$y = \frac{2}{n^2 - 1} \frac{1}{\exp[2m(\tanh\alpha - \alpha)] + 4 \exp[-2m(\tanh\alpha - \alpha)]} \\ \times \left\{ \exp[(m-1)(\tanh\alpha' - \alpha')] \exp[-m(\tanh\alpha - \alpha)] \right. \\ \left. - \exp[m(\tanh\alpha - \alpha)] \exp[-(m-1)(\tanh\alpha' - \alpha')] \right\} \quad (8.60)$$

In the numerator, it would be incorrect to replace α' by α if the resulting error is of order $1/m$, because of the additional factor m in the exponentials. We perform a Taylor expansion of $\tanh\alpha'$ in the small parameter $\xi \equiv \Delta m/m$ where in our case Δm will be set to -1 in the end. The result is

$$\tanh\alpha' \approx \tanh\alpha + \Delta m \frac{x^2}{m^3 \tanh\alpha}. \quad (8.61)$$

Similarly, one obtains

$$\alpha' \approx \alpha + \frac{\Delta m}{m \tanh\alpha}, \quad (8.62)$$

which leads to

$$(m + \Delta m)(\tanh\alpha' - \alpha') \\ \approx (m + \Delta m) \left[\tanh\alpha - \alpha + \frac{\Delta m}{\tanh\alpha} \left(\frac{x^2}{m^3} - \frac{1}{m} \right) \right] \\ \approx (m + \Delta m)(\tanh\alpha - \alpha) + \frac{\Delta m}{\tanh\alpha} \left(\frac{x^2}{m^2} - 1 \right) \\ = m \tanh\alpha - (m + \Delta m)\alpha \\ = m(\tanh\alpha - \alpha) + \alpha. \quad (8.63)$$

Here we have used the definition of $\tanh\alpha$ to cancel the term $\Delta m \tanh\alpha$. With these approximations Eq. (8.60) becomes

$$y = -\frac{4}{n^2 - 1} \frac{\sinh\alpha}{\exp[2m(\tanh\alpha - \alpha)] + 4 \exp[-2m(\tanh\alpha - \alpha)]} \quad (8.64)$$

The range of $\tanh\alpha$ is $0 \dots 1$. But the resonance width predicted by Eq. (8.64) vanishes for $x = m$, whereas we expect the width to increase monotonously as x approaches m from below. The reason for this failure is that the expansion Eq. (8.57) is invalid when $\tanh\alpha \rightarrow 0$ because the prefactor diverges. Therefore we should also take the limit $\tanh\alpha \rightarrow 1$. Since $m/x < n$, this requires taking the refractive index to be large. In the numerator, the second term then dominates the first one since

$$\alpha = \frac{1}{2} \ln \frac{1 + \sqrt{1 - x^2/m^2}}{1 - \sqrt{1 - x^2/m^2}} \quad (8.65)$$

diverges for $x \ll m$. In the denominator of Eq. (8.64), the first term (coming from J_m^2) becomes negligible for large m . Keeping only the dominant terms, the result is

$$y = -\frac{1}{n^2 - 1} \frac{1}{2} \exp[-(2m - 1)\alpha] \exp[2m \tanh\alpha]. \quad (8.66)$$

In the case of broad resonances, we were able to give a derivation of y in terms of Fresnel's formula for the reflection probability p_0 at each bounce of the ray. We can now ask what the form of p_0 has to be in order to reproduce the correct y for sub-barrier resonances. I.e. we are looking for the reflection probability from a curved interface that allows evanescent escape. The distance traversed by a ray between reflections is now $2nR \cos \chi$, and we obtain in analogy to Eq. (8.39)

$$\frac{2}{\tau} = -\frac{c \ln p_0}{2nR \cos \chi}. \quad (8.67)$$

Substituting τ by the width y , the reflection probability becomes

$$p_0 = \exp(-4n|y| \cos \chi) \approx 1 - 4n|y| \cos \chi, \quad (8.68)$$

i.e. the tunneling probability is $4n|y| \cos \chi$. This rule will find an application in our discussion of deformed cavities.

8.5 Resonance widths from WKB approximation

The previous expansion of the exact resonance condition for small y in the case of sub-barrier resonances yields the same exponential factor as would be obtained from the WKB approximation. We note this here because the condition for the validity of WKB is not strictly satisfied. The latter follows from the close proximity of all three classical turning points of the effective potential¹²⁶, and we will thus be unable to determine the prefactor of y except by methods like the one used above. The proximity of the turning points is precisely the reason why we could not use the large-argument approximation for the Bessel functions in Eq. (8.52). We start from Eq. (8.10) and look for the exponentially decaying solution in the forbidden region outside the disk, where

$$q(\xi) = \sqrt{e^{2\xi} - m^2} \quad (8.69)$$

is imaginary. The interval where this happens is $\ln(kR) < \xi < \ln m$. Thus, the decay factor for the intensity will be e^{-2J} where

$$J := \int_{\ln kR}^{\ln m} d\xi \sqrt{m^2 - e^{2\xi}} \quad (8.70)$$

$$= m \ln \frac{m + \sqrt{m^2 - k^2 R^2}}{kR} - \sqrt{m^2 - k^2 R^2}. \quad (8.71)$$

Now in the first term we use

$$\ln \left[\frac{m}{kR} \left(1 + \sqrt{1 - \frac{k^2 R^2}{m^2}} \right) \right] \quad (8.72)$$

$$= \ln \frac{1 + \sqrt{1 - \frac{k^2 R^2}{m^2}}}{\sqrt{1 - \left(1 - \frac{k^2 R^2}{m^2}\right)}} \quad (8.73)$$

$$= \frac{1}{2} \ln \frac{\left(1 + \sqrt{1 - \frac{k^2 R^2}{m^2}}\right)^2}{\left(1 - \sqrt{1 - \frac{k^2 R^2}{m^2}}\right) \left(1 + \sqrt{1 - \frac{k^2 R^2}{m^2}}\right)} \quad (8.74)$$

$$= \frac{1}{2} \ln \frac{1 + \sqrt{1 - \frac{k^2 R^2}{m^2}}}{1 - \sqrt{1 - \frac{k^2 R^2}{m^2}}} \quad (8.75)$$

$$= \operatorname{arctanh} \sqrt{1 - \frac{k^2 R^2}{m^2}} \quad (8.76)$$

$$= \alpha \quad (8.77)$$

with the definition of Eq. (8.58). This leads to

$$J = m(\alpha - \tanh\alpha), \quad (8.78)$$

in agreement with Eq. (8.66), and in disagreement with the J quoted in¹⁷ which is independent of m and kR/m . This dependence is crucial in explaining the influence of tunneling in the deformed billiard.

8.6 Global accuracy of the approximations

Having considered the limits $kR \gg m$ (above-barrier) and $y \ll 1$ (below-barrier), it remains to be tested how the two results, Eqs. (8.33) and (8.52), are connected in the transition region where $kR \approx m$. To this end we return to Eq. (8.52), which read

$$y = \frac{1}{1 - n^2} \frac{J_m(x)Y_{m-1}(x) - J_{m-1}(x)Y_m(x)}{J_m^2(x) + Y_m^2(x)}. \quad (8.79)$$

The condition for its validity is simply $y \ll 1$ which is in fact satisfied even for *above-barrier* resonances provided n is large enough.

One might expect that there exists a semiclassical approximation to this equation with fixed ratio $x/m > 1$, using the asymptotic forms¹²⁵

$$J_m \left(m \frac{x}{m} \right) \approx \sqrt{\frac{2}{\pi m \tan\beta}} \cos \left[m (\tan\beta - \beta) - \frac{\pi}{4} \right], \quad (8.80)$$

$$Y_m \left(m \frac{x}{m} \right) \approx \sqrt{\frac{2}{\pi m \tan\beta}} \sin \left[m (\tan\beta - \beta) - \frac{\pi}{4} \right],$$

$$\tan\beta \equiv \sqrt{\frac{x^2}{m^2} - 1}. \quad (8.81)$$

Already one sees that the denominator of Eq. (8.79) is unity in this approximation. Again we can neglect the difference between m and $m - 1$ in the prefactors but not

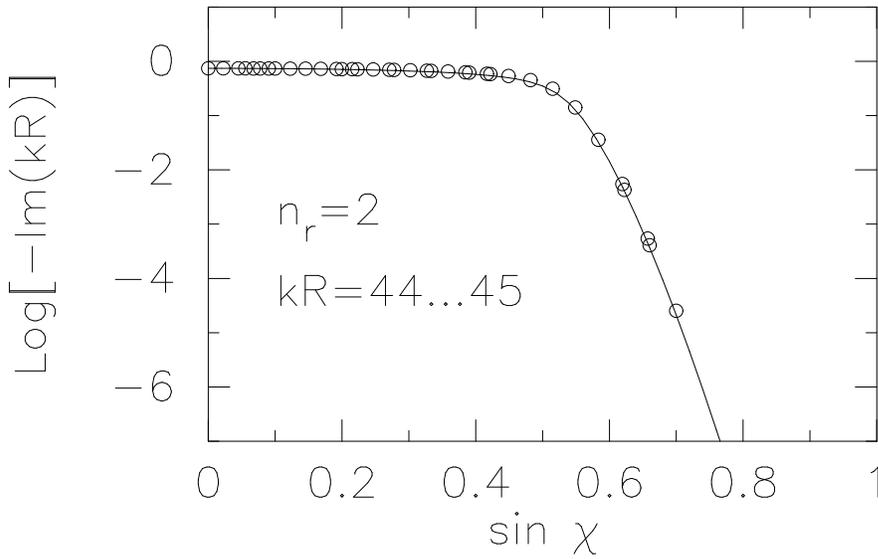


Figure 8.4: The open circles indicate exact resonance widths calculated from Eq. (8.27), for selected resonances whose real parts fall into the interval $44 < kR < 45$ at refractive index $n = 2$. To illustrate the significance of the total internal reflection condition $\sin \chi > 1/n$, the widths are plotted as a function of $\sin \chi = m/(nkR)$. The exponential falloff due to tunneling is seen to start for $\sin \chi$ above the critical value $\sin \chi_c = 1/2$. The solid line is Eq. (8.88).

in the argument of sin and cos, where we need the Taylor expansions in $\xi = \Delta m/m$. One obtains

$$\tan \beta' \approx \tan \beta - \Delta m \frac{x^2}{m^3 \tan \beta} \quad , \quad (8.82)$$

$$\beta' \approx \beta - \frac{\Delta m}{m \tan \beta} \quad , \quad (8.83)$$

$$\Rightarrow \quad (m-1)(\tan \beta' - \beta') \approx m(\tan \beta - \beta) + \beta. \quad (8.84)$$

With the abbreviation

$$A \equiv m(\tan \beta - \beta) - \frac{\pi}{4} \quad (8.85)$$

Eq. (8.52) now becomes

$$\begin{aligned} y &= \frac{1}{1-n^2} [\cos A \sin(A + \beta) - \sin A \cos(A + \beta)] \\ &= \frac{1}{1-n^2} \sin \beta. \end{aligned} \quad (8.86)$$

As in the case of below-barrier resonances, we encounter a difficulty when $x = m$, which implies $\beta = 0$ and thus vanishing decay rate. This is again due to the breakdown of the semiclassical approximation as seen from the diverging prefactors in Eq. (8.80). When $x \gg m$, Eq. (8.86) predicts $y = 1/(1-n^2)$ which differs slightly from Eq. (8.33). Both formulas show the physically expected divergence of the width at $n = 1$, and they become identical in the limit $n \rightarrow \infty$, cf. Eq. (8.34). Whereas Eq. (8.33) only involved one approximation, namely the semiclassical expansion of the Bessel functions, we obtained Eq. (8.86) by taking the semiclassical limit of Eq. (8.79) which itself resulted from a Taylor expansion in y .

We take this as a justification for adjusting the result of the Taylor expansion, Eq. (8.79) in such a way that the above semiclassical approximation leads to full agreement with Eq. (8.33) in the limit $x \gg m$. Since the condition for the Taylor

expansion was $y \ll 1$ which in turn requires $n \gg 1$, we are allowed to leading order in $1/n$ to replace the prefactor $1/(1 - n^2)$ in Eq. (8.79) by

$$\frac{1}{2n} \ln \frac{1 - 1/n}{1 + 1/n}. \quad (8.87)$$

This solves the problem, because Eq. (8.79) then automatically connects to Eq. (8.33). The formula we are therefore going to use across the whole range from $x \gg m$ to $x \ll m$ is

$$y = \frac{1}{2n} \ln \left[\frac{1 - 1/n}{1 + 1/n} \right] \frac{J_m(x)Y_{m-1}(x) - J_{m-1}(x)Y_m(x)}{J_m^2(x) + Y_m^2(x)}. \quad (8.88)$$

The comparison between this expression and the exact resonance widths is shown in Fig. 8.4. This expression will be most useful to us after it is substituted into Eq. (8.68) for the reflection probability per collision, and after the quantum number m is eliminated in terms of $\sin \chi$ and kR as done in Fig. 8.4. The resulting quantity

$$p_0(\sin \chi, kR) \quad (8.89)$$

is plotted for different values of kR as a function of $\sin \chi$ in Fig. 8.5. It is instructive to compare the results with the Fresnel formula for reflection from a plane interface

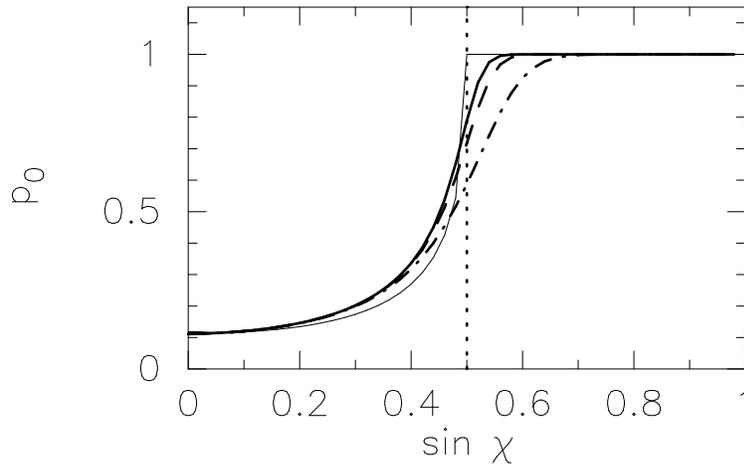


Figure 8.5: Reflection probability p_0 at each collision with a circular interface versus the sine of the angle of incidence, assuming $n = 2$. The heavy lines correspond to $kR = 45$ (solid), $kR = 27$ (dashed) and $kR = 12$ (dashed-dotted). For comparison, the Fresnel formula for a plane interface is plotted as the thin solid line. The vertical dotted line denotes $\sin \chi_c = 1/n$, above which one classically expects total internal reflection.

with electric field perpendicular to the plane of incidence. The major effect of the finite curvature in the circle is seen to consist in a nonvanishing transmission probability above the critical angle for total internal reflection. It is also important to note that even at the relatively low wavenumbers plotted here, p_0 settles into an almost k -independent shape for $\sin \chi < 1/n$. In the tunneling regime, $\sin \chi > 1/n$, the k -dependence is not seen in the figure but nonetheless is very strong on an exponentially small scale. This is of course the reason why the tunneling widths which differ by many orders of magnitude are correctly reproduced.

Chapter 9

Wave equation for asymmetric resonant cavities (ARCs)

We have been able to derive analytical formulas for the resonance widths of circular dielectric cylinders because angular momentum conservation allowed us to separate the wave equation into azimuthal and radial parts. In other words, the effective potential picture is valid only because m is a good quantum number in the circle. In the scattering state and the quasibound state for a *deformed* cavity, we then have to take into account admixtures of different m' , and their number grows with deformation.

In this and the following chapters we present two different ways in which to attack this problem: exact numerical calculations, and simulations based on a ray optics model. The two approaches are complementary in that the former provides an accuracy check for the latter, while the classical model is currently the only existing route to a theory of ARCs with explanatory and predictive power. The results of our numerical calculations will acquire physical meaning in the light of the classical model.

Historically, this thesis made its first progress toward an understanding of ARCs entirely based on classical considerations. However, the connection between ray optics and wave solutions requires a semiclassical approximation, the validity of which can be tested independently of other predictions of the model, by referring to the exact wave solutions. It therefore seems appropriate at this point to describe the numerical calculations first, followed by the discussion of the ray optics model, which then can profit from immediate comparison to the exact results.

9.1 Wave solutions for deformed cylinders

To calculate the effect of shape perturbations on the resonance width, one could start with a perturbation approach when the deformation is small⁶. However, we are interested in distortions that are not necessarily in this regime, so an exact solution of the scattering problem is required. The approach chosen here is a wavefunction matching technique. Here we discuss its justification and implementation, using the notation Ψ instead of E for the field in order to avoid confusion with the energy that appears in our classical models.

It may be worth remarking that the wave function matching technique has not been used widely because of formal divergences that can occur near the scatterer, as will be explained below. Had the author been aware of these difficulties before

embarking on the actual calculations, it is likely that a different method would have been tried first. The alternative methods that are in widespread use are mostly based on boundary integral representations of the fields. They are, however, inferior to our wave function matching technique in practice because they require finding a larger number of unknowns that grows with deformation. The resulting matrix equations are larger than for our method at the same high deformation. This may be one reason why no extensive calculations had been performed previously for the large convex deformations we are interested in. The other reason is, of course, that there was no theory that predicted what systematic behavior to expect from ARCs.

Our goal is to solve Eq. (8.6) for a *convex* domain D bounded in polar coordinates by $R(\phi)$, with an index of refraction $n > 1$ inside. In analogy to quantum scattering theory, we define the S-matrix in the angular momentum basis by looking for solutions of the wave equation with the asymptotic form

$$\Psi(r, \phi) := e^{im\phi} H^{(2)}(kr) + \sum_{l=-\infty}^{\infty} S_{lm} e^{il\phi} H_l^{(1)}(kr). \quad (9.1)$$

The first term is an incoming partial wave of azimuthal order m , and the second term represents the outgoing wave. This is the generalization of Eq. (8.22) to the case of broken rotational symmetry.

The numerical method consists in solving matrix equations for the expansion coefficients S_{lm} . In boundary-integral calculations and methods derived from it, the accuracy of the solution is increased by increasing the number of equations and of unknown coefficients. In our method, we can improve accuracy by adding more equations but *not* more unknowns (of course this implies the matrices will become rectangular). The boundary integral methods¹²⁷ (a variant of which is the T-matrix method¹²³) rely on the equivalence principle¹²⁸, replacing the scatterer by a set of induced surface polarization currents that result in zero field inside the scatterer.¹²⁹ The vanishing internal field is a sum of the incident wave and the field generated at the surface, and this allows one to determine the surface currents. From these, the scattered field can then be derived. The integrations over the surface that are needed here do not appear in the wavefunction matching technique, because the latter does not make use of the equivalence principle. The starting point of our method is the assumption that Eq. (9.1) describes the field *everywhere* outside the scatterer, not just in the asymptotic region.

9.1.1 Rayleigh Hypothesis

The question of whether or not the scattered wave expansion Eq. (9.1) in terms of Hankel functions is convergent in the *near field*, too, cannot always be answered *a priori*¹³⁰. The Rayleigh hypothesis is precisely this assumption, that Eq. (9.1) describes the wave in the entire exterior of the scatterer. We adopt this as a working hypothesis in the numerical calculation and check its validity *a posteriori* by inspection of the solution.

An immediate consequence of the Rayleigh hypothesis is that there are no evanescent channels of the type found in waveguides, since all Hankel functions become oscillatory at infinity.

The potential stumbling block that makes a general proof of the hypothesis impossible can be illustrated in the ellipse where exact solutions in terms of series of Mathieu functions are possible. The scattered field must have singularities somewhere in space because it would otherwise have to be identically zero, being

a solution of an elliptical differential equation. These singularities are located at the foci of the ellipse. If we want to expand the scattered wave in Hankel function centered in the middle of the ellipse, convergence can be guaranteed only outside the smallest circle that includes the foci. Its radius is just the distance of the foci from the origin, $\rho = \sqrt{a^2 - c^2}$. If ρ exceeds the short semiaxis c , then part of the exterior lies inside the circle where convergence of the series cannot be guaranteed. In this region we cannot be sure that the true scattered wave may be represented as a convergent series of Hankel functions (i.e. a series in which successive terms decrease sufficiently fast).

In the case of the ellipse, we see that the Rayleigh hypothesis is surely valid if $\sqrt{a^2 - c^2} < c$, corresponding to an eccentricity $e = \sqrt{a^2 - c^2}/a < 1/\sqrt{2}$ or a quadrupole parameter $\epsilon < 1/8$. Typical deformations of interest to us will be smaller than this limit.

Moreover, it has been pointed out^{131,132} that even if the Hankel series representation diverges formally (i.e. the Raighleigh hypothesis is violated), it is possible to numerically achieve any desired accuracy in the wave function matching approach. The reason for this is that although the infinite series in Eq. (9.1) may diverge at some points along the boundary, in practice one always truncates the sum after some number N of terms. The divergence of the infinite series is due to the fact that the coefficients of the large- l Hankel functions do not decrease fast enough to outpace the growth in $H_l^{(1)}(kr)$ for r on the boundary. For the truncated matrix equation, however, the coefficients adjust themselves such that the ones close to $l = N$ are smaller than the true coefficients. Still the squared deviation between true scattered field and truncated expansion, integrated over the boundary, approaches zero as N is increased. So for any N , the large-order coefficients are smaller than the true values, but the field approaches the true field. This is because partial waves with extremely large l are unphysical, i.e. cannot be significant in the far field. Their coefficients hence are practically zero, but the near-field displays an extreme sensitivity to their precise value, which in the truncated solution is kept at bay with no harm to the accuracy of the wave function. The algorithm used here in fact penalizes the use of too large a value of N by producing a more inaccurate wavefunction due to the numerically diverging Hankel functions of high order. Therefore, there is an optimum choice of N , large enough to include all physically relevant l and small enough not to cause numerical precision problems.

9.1.2 Wavefunction Matching

For the inside of D , we again expand the exact scattering state in angular momentum eigenfunctions $e^{il\phi}$, but this time only Bessel functions of the first kind can occur because all others diverge at the origin (the singularity of the Hankel functions at $r = 0$ does not affect the outside solution). We thus write

$$\Psi_D = \sum_{l=-\infty}^{\infty} A_{lm} e^{il\phi} J_l(nkr). \quad (9.2)$$

The continuity condition for E now has to be applied for all $\phi \in [-\pi, \pi]$:

$$\begin{aligned} e^{im\phi} H_m^{(2)}(kR(\phi)) + \sum_{l=-\infty}^{\infty} S_{lm} e^{il\phi} H_l^{(1)}(kR(\phi)) \\ = \sum_{l=-\infty}^{\infty} A_{lm} e^{il\phi} J_l(nkR(\phi)). \end{aligned} \quad (9.3)$$

The second condition is that $\partial\Psi/\partial r$ be continuous, cf. Eq. (8.17). Equating the r -derivatives inside and outside, we obtain for all ϕ

$$\begin{aligned} & e^{im\phi} \left[m H_m^{(2)}(kR(\phi)) - kR(\phi) H_{m+1}^{(2)}(kR(\phi)) \right] \\ &= \sum_{l=-\infty}^{\infty} e^{il\phi} \left\{ -S_{lm} \left[l H_l^{(1)}(kR(\phi)) - kR(\phi) H_{l+1}^{(1)}(kR(\phi)) \right] \right. \\ & \quad \left. + A_{lm} \left[l J_l(nkR(\phi)) - nkR(\phi) J_{l+1}(nkR(\phi)) \right] \right\}. \end{aligned} \quad (9.4)$$

Here we used the same relations for the derivatives of the Bessel functions as in Eq. (8.27), and multiplied through by $R(\phi)$.

The terms in the above equations containing the unknowns can be written in matrix form as $M^{(\nu)}\mathbf{x}$ if we define a vector \mathbf{x} through

$$x_{2l} := A_{lm}, \quad x_{2l+1} := S_{lm}, \quad (l \in \mathbb{Z}), \quad (9.5)$$

and a matrix M with ϕ as a continuous row index: For Eq. (9.3) we write

$$\begin{aligned} M_{2l}^{(1)}(\phi) &:= e^{il\phi} J_l(nkR(\phi)), \\ M_{2l+1}^{(1)}(\phi) &:= -e^{il\phi} H_l^{(1)}(kR(\phi)), \end{aligned} \quad (9.6)$$

and for Eq. (9.4)

$$\begin{aligned} M_{2l}^{(2)}(\phi) &:= e^{il\phi} \left[l J_l(nkR(\phi)) - nkR(\phi) J_{l+1}(nkR(\phi)) \right], \\ M_{2l+1}^{(2)}(\phi) &:= -e^{il\phi} \left[l H_l^{(1)}(kR(\phi)) + kR(\phi) H_{l+1}^{(1)}(kR(\phi)) \right]. \end{aligned} \quad (9.7)$$

The terms in the matching equations resulting from the incoming wave of angular momentum m can be collected into the 2-vector defined by

$$b^{(1)}(\phi) := e^{im\phi} H_m^{(2)}(kR(\phi)) \quad (9.8)$$

$$b^{(2)}(\phi) := e^{im\phi} \left[m H_m^{(2)}(kR(\phi)) - kR(\phi) H_{m+1}^{(2)}(kR(\phi)) \right]. \quad (9.9)$$

We thus can write symbolically

$$M(\phi)\mathbf{x} = \mathbf{b}(\phi), \quad (9.10)$$

which has to be satisfied for any ϕ . To solve the simultaneous equations, we need to transform to a discrete row index and then truncate the matrices.

9.1.3 Solving the matching equations

Equations (9.3) and (9.4) have been summed up in matrix form in Eq. (9.10). To find the scattering state for a given incoming m and wavenumber k , we find it is sufficient to take into account only slightly more than

$$m_{max} \approx nkR \quad (9.11)$$

angular momenta. The reason can be understood if we use the definition of the classical angle of incidence inside the billiard,

$$\sin \chi = \frac{m}{nkR}, \quad (9.12)$$

in the condition $m \leq nkR$. The result is simply $\sin \chi \leq 1$, which means that higher angular momenta give rise to rays that are no longer allowed inside the cavity. These angular momenta belong to rays that classically miss the scatterer. These partial waves will consequently be of less importance in the formation of resonant states, which is what we are looking for.

Having thus fixed the number of unknown coefficients A and S that are needed to specify the solution with sufficient accuracy, we need at least equally many independent equations to determine them. One approach would be to perform an expansion of all functions of the continuous index ϕ in some complete set of functions that are periodic in the interval $[-\phi, \phi]$. For each term in such a Fourier series, the matching conditions then have to be satisfied separately. This was tried, but found to be less efficient and no more accurate than the more direct way of discretizing ϕ itself. That is, we simply impose the matching equations at a discrete set of ϕ_q , where q enumerates at least m_{max} points along the boundary. If there are no symmetries in the problem, we can choose

$$\phi_q = (q - 1) \frac{2\pi}{N_\phi}, \quad q = 1 \dots N_\phi, \quad (9.13)$$

Here, it was *not* assumed that $N_\phi = m_{max}$, because we find it useful to impose more equations than there are unknowns, thus overdetermining the system of equations. In this way it can be insured that if the spatial discretization should by coincidence leave us with (numerically) linearly dependent equations at some ϕ_q and ϕ_r , there will be enough equations left to determine all the coefficients. If the assumption about m_{max} is correct, then a solution should exist with the same number of physically relevant unknowns, all higher-order partial waves being set to zero, no matter how many times overdetermined the system is. The solution that is thus found is then sure to be independent of the choice of ϕ_q .

The algorithm employed to solve the overdetermined system is due to Penrose, and goes by the name “pseudoinverse”, or singular-value decomposition^{133,134}. It provides the solution vector \mathbf{x} which minimizes the error $|M\mathbf{x} - \mathbf{b}|$.

9.1.4 Quasibound states at complex k

The advantage of using the S matrix formalism, i.e. calculating scattering states like Eq. (9.1) is that S is a unitary matrix, which provides us with a sum rule independent of the matching equations. The a and b coefficients used conventionally in the literature¹²² do not obey this additional relation. The accuracy of the wave function technique has been tested in this way by sweeping out an interval of (real) incident wavenumbers and verifying that unitarity is preserved while the scattering phases go through resonant behavior. However, for the subsequent analysis, it is necessary to calculate directly the quasibound states at complex k , as was done in the circle. The equations that have to be solved in this case are obtained from Eqns. (9.3) and (9.4) by leaving out the incident waves. This results in a homogenous matrix equation, $M\mathbf{x} = 0$, with the same definitions as previously. The outside field coefficients are now no longer the S-matrix elements, and they carry only one index instead of two, because m in S_{lm} labeled the incident partial wave which is absent here. In the matching equations, we should therefore replace S_{lm} by S_l and A_{lm} by A_l everywhere, to avoid confusion with the scattering state coefficients.

The fundamental difference to the real k scattering states is that a solution to the homogenous problem does not exist for all values of k , just as in the circle.

Therefore, we have to perform a search for the resonant k in the complex plane. The resonance condition is that the matrix M become singular. This means we could look for the zeros of $\det(M)$, which is however not the best way to approach the solution. The reason is first that our M should preferably be rectangular, as explained above. Secondly, we would like to determine the quasibound state wavefunctions simultaneously with k . Furthermore, $\det(M)$ has zeros resulting from all possible angular momenta, whereas we often know in advance which m will be important in the resonant state we are looking for. Therefore, it would be efficient to use this information and deal with a quantity that is more momentum-selective than the determinant.

To achieve this, we assume that angular momentum m is the dominant contribution in the quasibound state to be determined. Then we set its outside field coefficient equal to unity,

$$S_m = 1, \quad (9.14)$$

in Eqns. (9.3) and (9.4) and bring the corresponding terms to the righthand side of the matrix equation, leaving one fewer unknown on the lefthand side. The result is an inhomogenous equation

$$M' \mathbf{x}' = \mathbf{c}, \quad (9.15)$$

where M' is obtained from M by deleting the column $2m+1$, and \mathbf{c} is the negative of just this column. The vector \mathbf{x}' is likewise obtained from \mathbf{x} by dropping the element $2m+1$.

The resulting inhomogenous, overdetermined system is again subjected to singular-value decomposition, and in contrast to the scattering states there will in general be no solution. We find the complex resonant k by minimizing the output $|M' \mathbf{x}' - \mathbf{c}|$ of the singular-value decomposition. When it is equal to zero, the inhomogenous system of equations has a solution, and a quasibound state has been found in which m is a strong partial wave.

9.1.5 Symmetry considerations

The size of the problem can be reduced by taking into account possible symmetries of the domain D . In the presence of an incoming wave, this is not possible in general unless the incident direction coincides with a symmetry line. However the quasibound states can be chosen to exhibit all the spatial symmetries of the billiard. The quadratic Robnik billiard possesses reflection symmetry about the x axis. The symmetrized wave functions should therefore have even or odd parity under the operation $\phi \rightarrow -\phi$. This is achieved simply if we replace $e^{il\phi}$ by $\cos(l\phi)$ (even) or $\sin(l\phi)$ (odd) in the equations defining the matrix M . Then the sums over the angular momenta contain only $l \geq 0$.

A further simplification arises in the quadrupole and ellipse which have as an additional symmetry the reflection at the y -axis, $\phi \rightarrow \pi - \phi$. Thus the solutions can be chosen to be even or odd under this operation, too¹. This is realized by restricting the sums over l to run only over even or odd numbers, since $\cos l(\pi - \phi) = (-1)^l \cos l\phi$ and $\sin l(\pi - \phi) = -(-1)^l \sin l\phi$. The definition of $M^{(\nu)}$ can thus be simplified by writing $M_l^{(\nu)}$, $M_{l+1}^{(\nu)}$ instead of $M_{2l}^{(\nu)}$, $M_{2l+1}^{(\nu)}$ on the lefthand sides, where now l is either even or odd.

¹The full symmetry group is the direct product of the two reflection groups, yielding C_{2v}

The spatial discretization points in either case have to be chosen only from the part of the boundary that is irreducible under the respective symmetry operations which were taken into account in the definition of the matrix equation as above. For example, we choose

$$\phi_q = (q - 1) \frac{\pi}{N_\phi}, \quad q = 1 \dots N_\phi, \quad (9.16)$$

for the dipole billiard and

$$\phi_q = (q - 1) \frac{\pi}{2N_\phi}, \quad q = 1 \dots N_\phi, \quad (9.17)$$

for the quadrupole or ellipse.

Time reversal invariance implies that if Ψ is a solution of the scattering problem, then so is its complex conjugate. In the present case this provides no further reduction in the matrix sizes since both the unknown coefficients \mathbf{x} and the matrix M are complex.

9.1.6 Incrementing the deformation from the circle

Despite the above simplifications, a minimization in the complex plane can still be a tedious problem. Therefore, we make use of one more piece of information, namely the known resonance positions in the circle. Using this k as a starting value for the search, we can find the new k at a sufficiently small deformation such that the resonance has not wandered far. This process can then be iterated by incrementing the deformation in small steps and using the previously determined k as the starting point of each new search. In this way, we have the advantage of following a resonance from zero deformation to some large asymmetry, thus knowing at all times what quantum numbers this state was characterized by in the circle. Initially, the dominant angular momentum m needed in the above procedure is then equal to the quantum number in the circle. However, as we follow the state to larger deformation, we keep checking the newly found solutions to see when a different m' becomes dominant in the wavefunction. This m' is then used instead of m as the inhomogeneous term in the next iteration.

This procedure of following a resonance with deformation may seem to be severely threatened by the occurrence of avoided crossings, where states come close in k , and due to their interaction change their path abruptly. However, this was not found to be the case for the whispering gallery resonances of interest here, because the quasibound states that do cross in real part of k are usually so far separated in imaginary part that the short-lived one has to be considered part of a broad background and hence has no effect on the narrower resonance.

One disadvantage of the method, however, is that we do not retain a global picture of how other resonances in the neighborhood evolve. If, e.g., we follow all the resonances of the circle in the interval of Fig. 8.2 to some high deformation, we could not draw a plot of the resonances in that same k interval because some states may have wandered out of it and others (which we did not keep track of) could have entered it. For the purposes of our study, it was precisely the goal to follow the evolution of a given resonance with deformation, so that the method described here is suitable.

To give an impression of the reliability of the numerical algorithm and to point to the difficulties that had to be overcome in obtaining our data, Fig. 9.1 shows

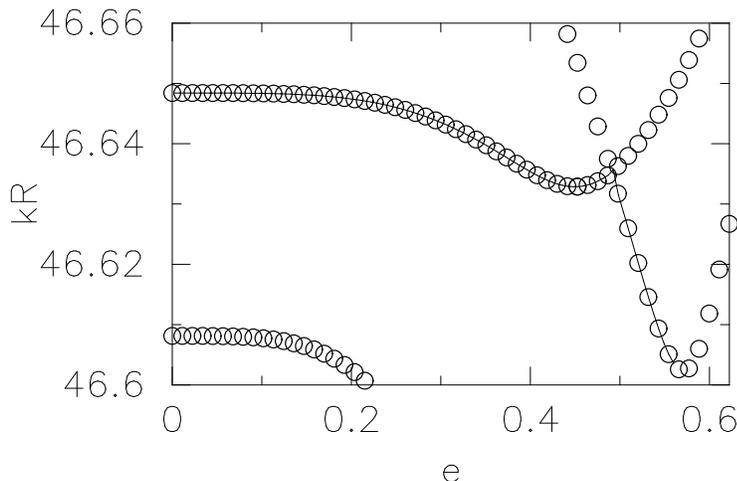


Figure 9.1: Wavenumber of an eigenstate in the closed quadratic Robnik billiard as a function of deformation. The symbols are obtained from an exact diagonalization, the solid line results from the wavefunction matching approach.

an application of an analogous technique to the problem of the *closed* Robnik billiard, for which solutions are available. The wavefunction matching in this case consists of the requirement that the field vanish identically on the boundary, so that there are no scattered field coefficients to be determined. The eigenstates are furthermore located at real k , so that only a minimization on the real axis is necessary. Our results are seen to be identical with the data obtained by Bruus^{135,136} using a diagonalization technique. Note, however, that an avoided crossing occurs on a scale not resolved in the diagonalization data, after which point the wavefunction matching algorithm follows the state that clearly does not retain the character of the WG mode. This is technically exactly the expected behavior, because the algorithm has managed to track the k position of the state through the avoided crossing. However, physically we would like to keep following the state with WG character, and thus one would have to jump across the gap between the anticrossing levels. In the open billiards we are going to consider next, close encounters of quasibound state positions in the complex k plane are comparatively rare, but due to the fact that the minimum search is in two dimensions and not in one (as for the closed system), it becomes harder to insure that the correct resonance is followed through such events.

9.2 Emission directionality of quasibound states

Whereas up to now the quasibound state was introduced only as a convenient tool for extracting resonance widths and positions that could otherwise be determined from Breit-Wigner fits for the scattered intensity, we want to examine now what physical significance it has in its own right. This question was also studied by Young and co-workers⁶, who give further references. The quasibound state can be thought of as the limiting case of a wave packet launched in the cavity and decaying to infinity. This is the appropriate description for an emission process such as lasing, where the light waves are generated in the cavity, rather than being sent in from infinity and then elastically scattered.

If the resonant state is at the complex frequency $\omega - i\gamma \equiv c(k - i\kappa)$, then the corresponding solution of the time dependent wave equation decays at a rate γ

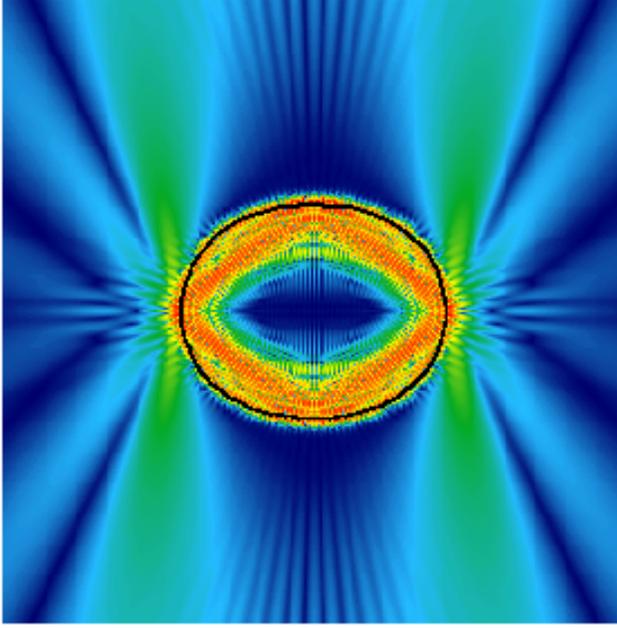


Figure 9.2: False color representation of the squared electric field in the TM mode for the $m = 68$, $kR = 45.15$ resonance of the quadrupole-deformed cylinder at eccentricity $e = 0.66$. The intensity is higher for redder colors, and vanishes in the dark blue regions. High intensity regions in the near-field occur just outside the surface at the highest curvature points $\phi = 0, \pi$, and high emission intensity lines (green) emanate from these points in the tangent directions. The high intensity inside the cavity reflects the quasibound nature of the state, which is seen to be still WG-like with no intensity in the center.

since it has the form $\psi(\mathbf{r}, t) = \psi(\mathbf{r}) e^{-i\omega t} e^{-\gamma t}$ where $\gamma > 0$. But as a function of r , the outgoing waves in fact exhibit exponential growth because

$$H_m^{(1)}(x) \approx \sqrt{\frac{2}{\pi x}} e^{i(x - m\pi/2 - \pi/4)} \quad (9.18)$$

for large values of $x = (k - i\kappa)r$. The physical reason for this growth with $e^{\kappa r}$ is a retardation effect: the field at $r \gg R$ has propagated away from the cavity where it originated a time $\Delta t \approx r/c$ in the past – but at that earlier time the field at the cavity was larger by a factor $e^{-\gamma \Delta t}$. This is equivalent to $e^{\kappa r}$.

As can be seen from Eq. (9.18), all the Hankel functions in the outgoing wave depend on r through the *same* factor $\sqrt{\frac{2}{\pi x}} e^{i(k - i\kappa)r}$ in the far-field ($r \gg R$). Pulling out this common dependence, the field of the quasibound state factorizes into radial and angular functions,

$$\psi(\mathbf{r}) = \sqrt{\frac{2}{\pi x}} e^{i(k - i\kappa)r} E(\phi). \quad (9.19)$$

This means that the directionality at large distances becomes independent of r , being contained solely in $E(\phi)$. We will then choose r in this far-field region and plot the square of the electric field (which is proportional to the intensity) as a function of ϕ to obtain the wave directionality.

Figure 9.2 shows an example of a full quasibound state wavefunction. It will be the purpose of chapter 11.3 to explain the strongly anisotropic intensity distribution observed here. Also note that the above-mentioned exponential growth

does not show up in this figure. This is clear from Eq. (9.19), which tells us that due to the prefactor the wavefunction will in fact fall off up to a distance $r = 1/(2\kappa)$, and only beyond this r begin to grow. Since $\kappa R < 1/10$ as will be seen later, the figure captures only the spatial decay. This emission process differs from elastic scattering which requires an incoming wave to excite the resonance. The directionality pattern in a scattering experiment will depend on the form of the incident wave both because of interference with the outgoing wave, and because the incident wave may couple preferentially to different senses of circulation of the rays. These effects are absent in emission, so a unique directionality profile will be observed that depends only on the quasibound state itself and should be approximately described by our ray optics model if kR is sufficiently large.

Chapter 10

Ray-optic model for resonances

This is the central chapter of the thesis, because it lays the groundwork for our understanding of the wave results. Our goal is a theory of the intrinsic properties of individual quasibound whispering-gallery modes in the deformed dielectric cylinder. These are, as listed in the introduction, the shift in the real part of k , the change in the decay width, and the anisotropic emission that develops in the deformed cavity.

The latter two are properties that obviously cannot even be studied in closed systems, whereas the frequency shift is a quantity that we already addressed when discussing the semiclassical quantization of convex closed billiards using the adiabatic invariant curves, cf. Eq. (7.13).

10.1 Semiclassical quantization for the resonance position

10.1.1 Applicability of EBK quantization

Let us therefore begin by asking whether we can use the same type of EBK quantization to predict the resonance positions in the open billiard. A necessary condition is surely that the adiabatic curve should be followed for at least one period in ϕ , which means the decay time of the resonance should be long enough for a ray to make this round trip. If $y = \text{Im}(kR)$, then the path length traversed before escaping is $L = 1/|ny|$, which has to be larger than the circumference of the billiard. An approximate condition that has to be satisfied if the resonance is to be semiclassically akin to a bound state is thus

$$2\pi R < L \quad \Rightarrow |y| < \frac{1}{2\pi n}. \quad (10.1)$$

In the circle, we know from Eq. (8.34) that above-barrier resonances have a width

$$|y| \approx \frac{1}{n^2}, \quad (10.2)$$

which for common indices of refraction, $n \leq 6$, is not sufficient to justify a classification as a quasi-bound state in the above sense. Therefore, it is necessary to restrict our attention to below-barrier resonances, i.e. whispering-gallery modes, to make headway with a semiclassical analysis similar to that for bound states. Since these orbits are confined by total internal reflection, their classical motion is identical to that of the closed system. In the semiclassical quantization, we can thus use the adiabatic curve and the resulting expression in Eq. (7.11). The only difference is that the boundary conditions are neither of Dirichlet nor von-Neumann type. Therefore, one has to reinstate the more general boundary phase shift $\alpha \in [0, \pi]$,

(not to be confused with α from Eq. (8.57)) which leads to a replacement of the $3/4$ term in the radial quantization condition by a number between $1/4$ and $3/4$. We know that $\alpha = 0$ for above-barrier resonances, see Eq. (8.32). For a general WG resonance, α can be fixed by requiring the position x at zero deformation to be equal to the semiclassical result. We then leave α constant as the deformation is introduced. This essentially means that α is a parameter in the semiclassics for the deformed case, determined uniquely by the results for the circle.

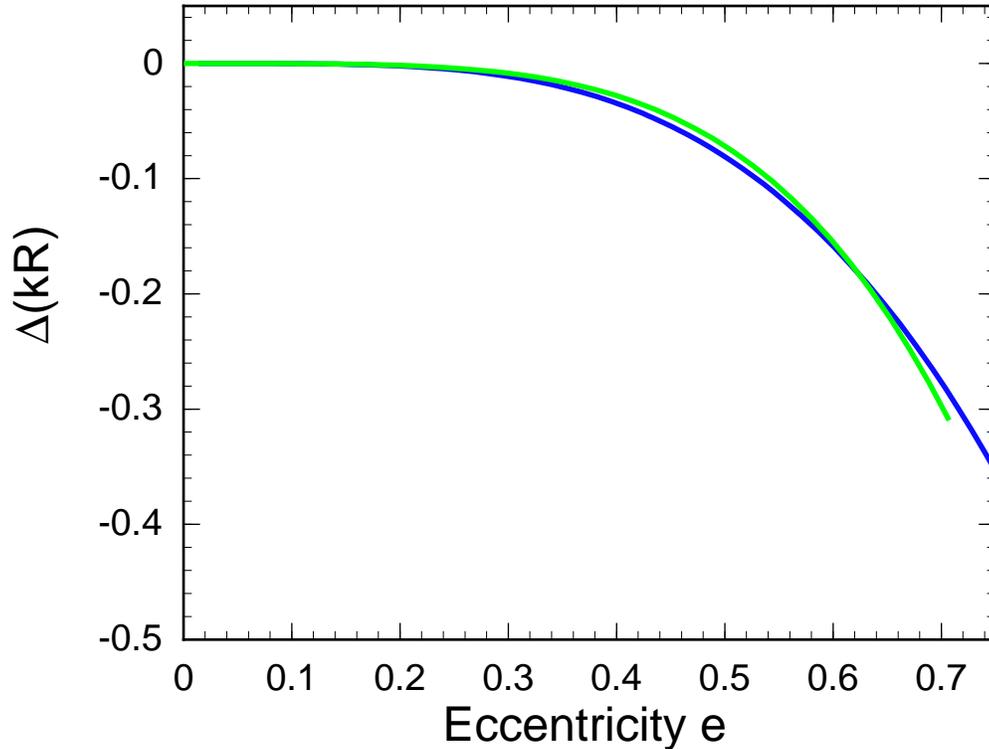


Figure 10.1: Shift in real part of kR , denoted here by ΔkR , versus deformation for a resonance with $kR = 28.6$, $m = 46$ (in the circle), under an elliptical deformation. The refractive index is $n = 2$. The green curve shows the semiclassical result.

The assumption that α is unchanged with deformation, and moreover that a single α describes the phase shifts for reflection at all points on the interface even in the deformed case, has to be tested. Figure 10.1 shows that excellent agreement between semiclassics and exact resonance position are obtained in the ellipse, after α is chosen to make the curves coincide at $\epsilon = 0$. The plot is obtained by fixing the radial and “angular momentum” quantum numbers to be those of the resonance in the circle, and then solving Eqs. (7.10) and (7.11) for the deformed shape with curvature $\kappa(\phi)$.

This example represents an open system, but the internal dynamics of the ellipse is integrable. Recall that the ellipse with a finite wall is in fact nonintegrable; however, the constant of motion is only violated when a classical trajectory leaves the billiard. This does not occur for the whispering-gallery trajectories which are confined by total internal reflection. Therefore, the adiabatic invariant curves are an exact description of the WG dynamics here.

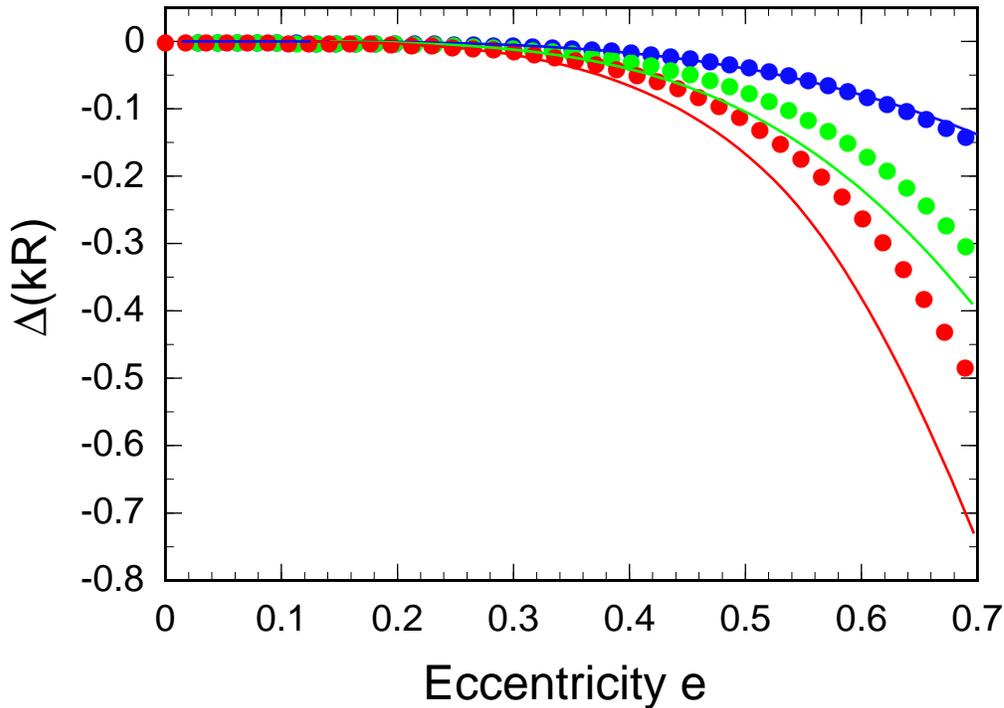


Figure 10.2: Shift in real part of kR , denoted here by ΔkR , versus deformation for three resonances of the quadrupole with refractive index $n = 2$. The wavenumbers and angular momentum quantum numbers at $e = 0$ are $kR = 12.1$, $m = 20$ (blue), $kR = 27.5$, $m = 44$ (light green) and $kR = 44.6$, $m = 71$ (red). The dots of the same color represent the corresponding semiclassical shift. The relative errors $\Delta(kR)/kR$ are 0.002, 0.004 and 0.005 with increasing m .

Having seen that the openness of the system does not invalidate the semiclassical treatment of WG modes, we now apply the same procedure to the quadrupole billiard. In Fig. 10.2, three WG resonances are followed with deformation. First note the universal red shift of the resonance positions for all the resonances, which is reproduced by the semiclassical approximation. All three resonances correspond roughly to the same classical trajectories in the circle, because the semiclassical quantization yields almost the same value of $\tilde{p} \approx 0.8$ [i.e. the same adiabatic curve, cf. Eq. (7.10)] for all of them. However, the shift in position, ΔkR is larger when m is larger. This tendency is correctly reflected in the semiclassical results, and one can immediately explain it using Eq. (7.13). The deformation is introduced at constant area, which necessarily increases the circumference L_0 of the billiard. In order for the same number of integer wavelengths to fit into L_0 as in the circle, the wavelength must increase, causing the red shift.

10.1.2 Dynamical localization

The overall accuracy of the semiclassical approximation is better than the figure suggests. It has to be kept in mind that we are only plotting the shifts in position, which make up only a fraction of the total kR , especially at high kR . The relative error in kR is smaller than 1% for all states, even at the largest deformation. Adjusting α to agree with the resonance position in the circle is not responsible for this accuracy, considering that according to Eq. (7.3) a change in α has less of an effect on the radial action than a change in the radial quantum number, so that kR changes only by a fraction no matter what α is chosen. Having pointed out this overall agreement, we now turn to a brief discussion of the observed discrepancies.

The agreement between semiclassics and exact results is best for smaller kR . This is surprising at first sight because it runs counter to the expectation that the semiclassical limit should be approached as the wavelength decreases. If we want to resolve this contradiction, it must be concluded that deviations of the whispering gallery trajectories from the adiabatic invariant curves become more important at high kR . After all, the adiabatic curves are only followed approximately, and for a finite number of reflections, while their validity is a prerequisite for the EBK quantization.

We interpret this as an indication for the existence of dynamical localization in the billiard, at least at high eccentricity where the WG trajectories no longer move on invariant curves. Of course, the deviation between semiclassics and exact results already develops at lower deformations, and the agreement should in this regime be improved if the exact invariant curves were used in the EBK formulas rather than the adiabatic invariant curves. However, once chaos has spread to the region around $\sin \chi \approx 0.8$, the adiabatic invariant curves are the best approximation we have left. The localization interpretation then addresses not the question why the agreement is so bad at high kR (which is related to the classical question of how good the adiabatic curves work), but why the agreement is so good at low kR .

To appreciate the concept of dynamical localization¹³⁷, recall subsection 7.3, where it was noted that the effect of classical diffusion on the quantization decreases with increasing \hbar and increases with the classical diffusion constant. Now D should be similar for all states considered, because they correspond to the same \tilde{p} ; but in the eikonal formulation, we identified $\hbar = 1/k$, so we conclude that diffusion will play a lesser role at large k , in agreement with Fig. 10.2. We are referring to a diffusion in $\sin \chi$, away from the original adiabatic curve. Via the relation $\sin \chi = m/(nkR)$ for the circle, this is equivalent to the admixture of some range of m to the quantum wave function ψ . Localization is the phenomenon that not all m corresponding to chaotic intervals of $\sin \chi$ contribute equally to ψ , i.e. that the m -distribution of the state is peaked at some \bar{m} and vanishes for m outside some interval around \bar{m} . This effect shows up in the angular momentum basis for our billiard problem, but also manifests itself in the real space wavefunction, which for localized whispering-gallery modes display the remnant of a caustic, even if the corresponding classical rays are already allowed to explore the center of the billiard after diffusing in phase space for a sufficiently long time.

Localization is not observed in all chaotic maps, a counterexample being the kicked top¹³⁷. However, one might expect localization to occur in convex billiards because of the slow diffusion in momentum. The kicked top lacks precisely this separation of a slow and a fast variable. This subject deserves further study, and we have not investigated it in detail during this thesis. Therefore, we now return to

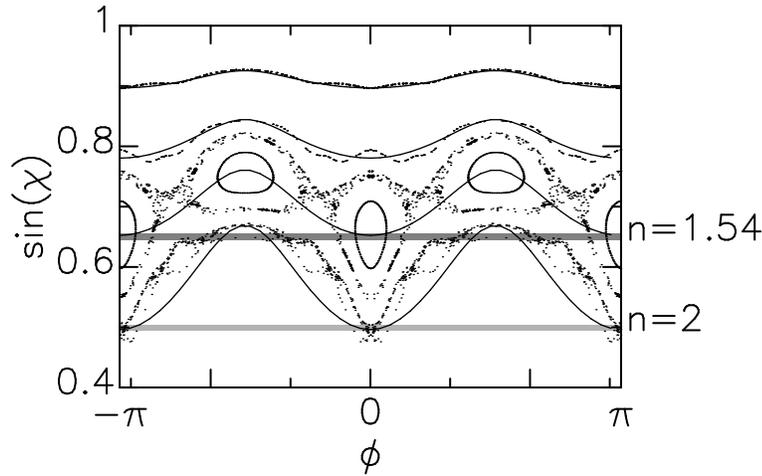


Figure 10.3: The starting condition for the ray escape simulation is given by the adiabatic invariant curve $\tilde{p}_{m,q}$. If tunneling and above-barrier (Fresnel) reflection are neglected, the classical escape condition is that the trajectory cross the line $\sin \chi_c = 1/n$. This defines a billiard with an escape window in phase space that must be reached by classical time evolution. This window is smeared out when the above wave effects are included.

the phenomena that are (unlike the localization problem) unique to open systems.

10.2 Quasibound states and ray dynamics

As a by-product of the semiclassical quantization, whose good relative accuracy was noted in the remarks accompanying Fig. 10.2, we also obtain information about the classical trajectories associated with the resonance under consideration. This is a result of Eqs. (7.10) and (7.11), where we first have to determine $\tilde{p}_{m,n}$ and then the quantized wavenumber. Denoting the analog of the radial quantum number by q to avoid confusion with the refractive index, $\tilde{p}_{m,q}$ specifies the adiabatic invariant curve, Eq. (7.4), along which the trajectories move for sufficiently long time to give rise to the resonant state.

Deviations from the adiabatic curve were neglected in determining the real parts of kR , but clearly one cannot neglect them in solving the problem posed at the beginning of this chapter: determining the resonance widths in the deformed system.

In the circle, we know that WG resonances are narrow due to the low tunneling escape rate. The basic idea that opens the connection to nonlinear dynamics is that at sufficiently large deformations, a new and competing escape mechanism becomes dominant, replacing tunneling as the process limiting the decay. The resonance lifetime at high deformation is limited by *classical ray escape*. It occurs when a ray starting on the adiabatic curve belonging to a WG mode diffuses downward in $\sin \chi$ until the condition for total internal reflection,

$$\sin \chi > \frac{1}{n}, \quad (10.3)$$

is violated. The real-space picture of this process was illustrated in Fig. 1.2, and the location of starting and escape conditions in the Poincaré section is shown in Fig. 10.3. As an implication of this argument, it is precisely the deviation of the trajectory from the adiabatic curve due to phase-space diffusion that determines the resonance lifetimes at high deformations. This does not constitute a contradiction

to the validity of the semiclassical quantization provided the escape times due to classical diffusion are still long enough to permit the adiabatic curve to yield an accurate semiclassical quantization. As a minimal criterion, this calls for at least one revolution around the boundary along the adiabatic curve. This is the condition that led us to include only WG modes in our discussion in the first place, cf. Eq. (10.1).

After this general statement of the program, we now elaborate on the attempt to make these ideas quantitative. In the circular dielectric, we defined a reflection probability per collision, Eq. (8.68), which can be expressed as a function of kR and $\sin \chi$ using $\sin \chi = m/(nkR)$, cf. Fig. 8.5. This is very valuable because it contains none of the quantum numbers of the circle explicitly. Having replaced angular momentum m by $\sin \chi$, we can generalize the escape law of the circle to the non-circular case. Our assumption is that at every collision with the boundary, the ray is reflected with a probability $p_0(\sin \chi, kR)$ and escapes with likelihood $1 - p_0$. By definition, this reproduces the correct results in the circle. In the deformed cavity, we want to apply the same rule, with $\sin \chi$ given by the *momentary* angle of incidence since χ is no longer a conserved quantity. In other words, we model the ray in the deformed dielectric as experiencing the interface locally like a circle at each collision.

This straightforward generalization of arguments from the circle allows us to define the decay time as an average over an ensemble of trajectories on the adiabatic curve $\tilde{p}_{m,q}$, of the time t needed by each orbit to escape. For each orbit, the escape time t can be obtained from a Monte-Carlo simulation, following the classical trajectory and producing at each collision with the boundary a random number between 0 and 1; if the latter is larger than p_0 , escape occurs. The path length L of the ray up to this event is related to the escape time by $L = ct/n$, and the decay time is

$$\tau = \left\langle \frac{nL}{c} \right\rangle, \quad (10.4)$$

where the average over different trajectories on the adiabatic curve (denoted by the angular brackets) is necessary because $\sin \chi$ is a function of position ϕ along this curve, so the starting conditions are inequivalent. We call this the *pseudoclassical* model for the decay time.

The model thus defined suffers from the approximation that coherence of any kind is not taken into account. This includes the possibility of coherence between successive tunneling events, because the simulation is purely sequential. It also includes the fact that the internal evolution of $\sin \chi$ (or m) does not necessarily follow the classical dynamics, e.g. as a consequence of dynamical localization. The only wave effect that is contained in the simulation is direct tunneling through the instantaneous effective potential barrier as derived from the angle of incidence. This is why we call the model pseudoclassical. Another approximation is that we take kR in the expression for p_0 to be a constant, thus neglecting both the shift in k and the variation in radius of curvature between collisions; however these two points could be corrected in future work without making any changes to our basic physical understanding, and their quantitative importance is likely to be small because neither quantity varies strongly.

Chapter 11

Comparison of ray model and exact solutions

11.1 Universal widths at large deformation, localization correction

With these caveats, we proceed to the results of our model simulation and examine their relation to the data obtained from numerical solution of the quasi-bound state problem¹³⁸. Without the need for actual simulations, our model already predicts that resonances whose semiclassical quantization yields approximately the same value of \tilde{p} should have roughly the *same* width at large deformation, with an accuracy that increases with kR . Besides the red shift of the real parts, this is the second property of the ARC that is *universal*. The reason for this prediction is that the reflection probability per collision, $p_0(\sin \chi, kR)$ becomes wavelength-independent for sufficiently large kR when $\sin \chi < 1/n$, as shown in Fig. 8.5. Then the escape is determined only by the refractive index and the classical ray dynamics, provided classical ray diffusion is faster than the tunneling decay rate.

Figure 11.1 shows the exact resonance widths for three resonances whose widths at zero deformation differ by many orders of magnitude. As we increase ϵ , the widths go through a roughly exponential increase until at large deformation they all coalesce to nearly the same decay rate within less than a factor 10. As shown in the inset, the agreement between the pseudoclassical model and the exact width for the resonance with $\sin \chi = 0.8$ and $m = 20$ is excellent at high deformations. This confirms our expectation of the universal resonance widths in at large asymmetry. The slight differences in width in the region $\epsilon > 0.11$ also deserve comment. Note that the order of the widths versus $\text{Re}(kR)$ there is reversed compared to $\epsilon = 0$, i.e. the resonance with the highest frequency, which had the smallest width in the circle, displays the largest width at high deformation. This can again be understood as evidence for dynamical localization, because it indicates that classical diffusion, which for all three resonances starts from almost the same adiabatic invariant curve, is more effective at shortening the lifetime for larger kR , i.e. smaller effective \hbar .

The case for localization is made again in the inset to Fig. 11.1, where we redraw one of the resonances with $\tilde{p} \approx 0.8$ with its pseudoclassical companion, and compare it to a resonance that is semiclassically located near $\tilde{p} \approx 0.9$. For the latter state, the pseudoclassical model *overestimates* the width by an order of magnitude. This could be explained by a stronger effect of dynamical localization on this state. Indeed, we know from our discussion of the billiards mapping that the diffusion constant decreases as $\sin \chi \rightarrow 1$, and therefore states are expected to become more localized in this limit.

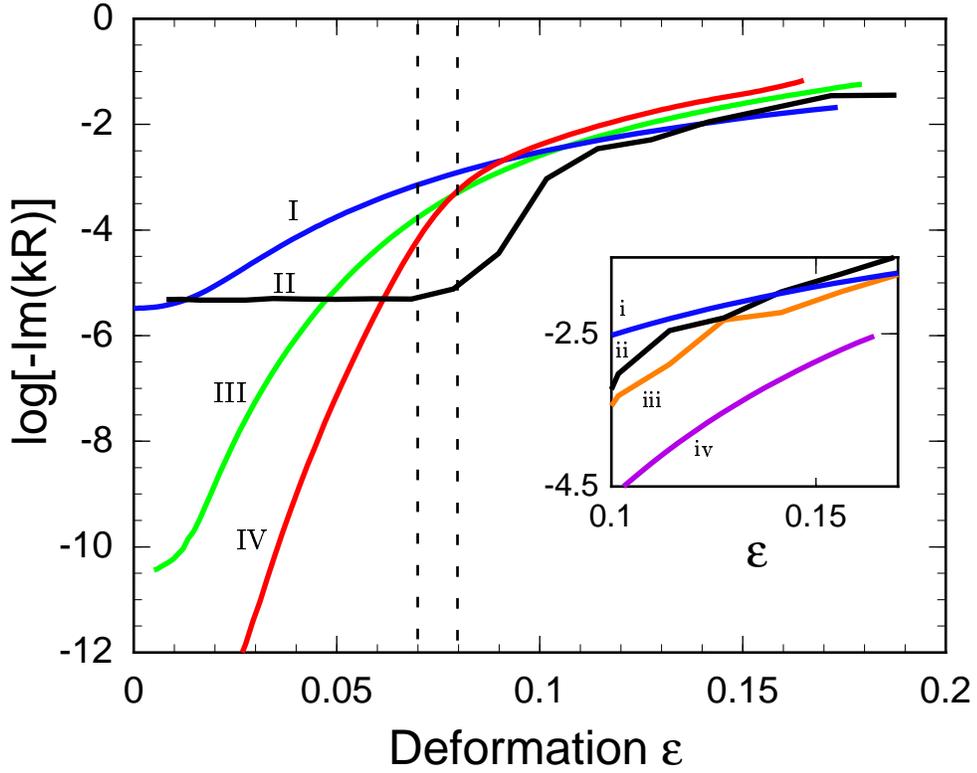


Figure 11.1: Logarithmic plot of imaginary part of resonance positions vs. deformation for quadrupole with refractive index $n = 2$. The quadrupole deformation parameter ϵ is used here instead of eccentricity to expand the region $\epsilon > 0.8$. The red, green and blue lines correspond to the same resonances whose real parts are displayed in Fig. 10.2. The vertical dashed lines delimit the ϵ -interval where the whispering-gallery trajectories around $\sin \chi \approx 0.8$ cease to move on invariant curves and classical diffusion to the critical $\sin \chi_c = 1/n$ begins. The black solid line is the escape rate, $1/(n\langle L \rangle)$, as obtained from the pseudoclassical model for the $m = 20$ state (blue line). In the inset, the blue and black lines have the same meaning as in the main figure, while the purple line shows the wave result for $kR = 33.2$, $m = 60$, and the brown line is the corresponding pseudoclassical result.

11.2 Chaos-assisted tunneling

The pseudoclassical model agrees well with the exact result for $\epsilon > 0.11$, but fails badly in the transition region between the circle and the strongly deformed cavity (the black and blue lines should agree but clearly do not). In this regime, the WG orbits are still supported by unbroken KAM curves while at the same time chaotic regions exist at lower $\sin \chi$. According to our model, the ray has an escape probability at each reflection, given by the below-barrier behavior of $p_0(\sin \chi, kR)$ since the KAM curves are bounded from below in $\sin \chi$. This *direct* escape attempt is, however, not the most efficient channel. An alternative process is for the ray to tunnel *in phase space* to a slightly lower $\sin \chi$ lying in the chaotic domain, from which point on it can then classically diffuse to the escape window in the SOS, $\sin \chi_c = 1/n$. This classically forbidden jump from a KAM curve into the chaotic sea has been called chaos-assisted tunneling in the study of enhanced tunnel splittings in a closed billiard system known as the annular billiard^{51,139,140}.

Before we explain the idea of chaos-assisted tunneling, it should be mentioned that the first studies of quantum tunneling in KAM phase spaces were performed on driven (i.e. time-dependent) systems^{142,143}. Geisel *et al.*¹⁴³ consider the standard map and its quantum equivalent. A particle is launched with an action J_0 located in

a chaotic layer bounded from below and from above by two KAM tori. Its probability to be found outside this bounded layer is suppressed exponentially with the distance $J - J_0$ in action, due to the fact that classically the crossing of the bounding KAM tori is forbidden.

11.2.1 The annular billiard

The annular billiard is a conservative system in which the effects of chaos-assisted tunneling can be studied with great efficiency. This system consists of two non-concentric circles acting as hard walls, with the ray motion taking place in the annular region between them. The annular billiard has a reflection symmetry and its eigenstates are thus classified as either even or odd, and both symmetry types are degenerate in the concentric billiard. This degeneracy is lifted when the inner circle is shifted off-center, even for states that semiclassically correspond to \tilde{p} large enough such that the classical trajectories never touch the inner obstacle. This splitting is a tunneling effect and can be understood in real space by noting that the whispering-gallery orbits in the pure circle have wavefunctions that penetrate into the classically forbidden region of the angular momentum barrier, $V_{eff}(r)$. If we now combine solutions with $e^{im\phi}$ and $e^{-im\phi}$ angular dependence to form either symmetric or antisymmetric wavefunctions under reflection at the billiard axis, then the antisymmetric ones will vanish on the reflection axis while the symmetric ones are maximal there. The inner circle has its point of closest approach to the boundary on the reflection axis, so that the symmetric eigenfunctions of the unperturbed circle have more overlap with the obstacle than the antisymmetric solutions. However, the splitting is found to be much larger than expected from this “direct” tunneling argument when there are chaotic regions in phase space. It was argued in Ref.⁵¹ and confirmed analytically in Ref.¹⁴⁰ that the most important alternative processes involve transitions not directly from m to $-m$, but first to the nearest edge of the chaotic sea followed by classical diffusion to the opposite edge near $-m$, and then another tunneling process to the regular orbit at $-m$.

11.2.2 Chaos-assisted tunneling in open systems

In our system, the *escape rates* of the wave solutions are much larger than expected from the direct process, and chaos-assisted tunneling is the likely explanation. Analytical progress in the deformed billiards is more difficult than in the annular billiard, so that a first closer investigation of the effect of chaos-assisted tunneling on resonance widths in open systems has been undertaken by Hackenbroich and the present author⁵⁶ using the more tractable annular billiard, but with a *penetrable* outer wall. Enhancement of resonance widths by several orders of magnitudes was found whenever coupling to the levels quantized in the chaotic sea was possible. The escape can therefore be modeled as a two-step process with tunneling from the regular WG region to the chaotic sea, followed by classical diffusion in the chaotic component until escape occurs. The additional permeable outer coating in this case causes the chaotic states to be well-isolated from each other while their widths are nonetheless much larger than those of the regular states. The final escape is again a tunneling step through the outer barrier, rather than simply the violation of total internal reflection.

In the deformed billiards we are considering in this thesis, the chaotic states form broad, strongly overlapping resonances in the absence of the above coating, so that the two-step escape process now is no longer limited by the availability of a

chaotic state that comes sufficiently close in energy to the regular state with which we start out. Instead, chaotic resonances have spectral weight at all energies in the neighborhood of the regular state, and the limiting factor is now the size of the coupling matrix elements. The shortest tunneling distance in phase space from the regular state to the edge of the chaotic domain should then dominate the escape.

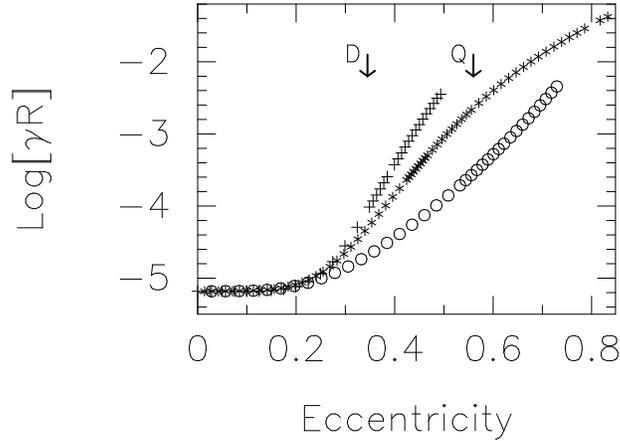


Figure 11.2: Logarithmic plot of the imaginary part (γR) of the resonance position kR as a function of deformation for the resonance $kR = 12.1$, $m = 20$ at refractive index $n = 2$ in the ellipse (circles), quadrupole (stars) and dipole (crosses). The arrows indicate the classical threshold deformation for the onset of diffusion from starting to escape condition, in the dipole (D) and quadrupole (Q).

As a further indication that chaos-assisted tunneling occurs in open billiards, we show in Fig. 11.2 the resonance widths of the resonance $kR = 12.1$, $m = 20$, followed as a function of deformation but for different shapes displaying different degrees of chaos. The width in the ellipse, where the internal dynamics is integrable, are significantly smaller at the same eccentricity than that in the quadrupole. The latter in turn is less chaotic than the dipole billiard at equal e , and accordingly we find the largest width for the dipolar deformation.

11.2.3 Tunneling without chaos in the ellipse

The width graph for the resonance in the ellipse is seen to be qualitatively similar to those of the chaotic billiards, despite the fact that the closed ellipse is an integrable system. In particular, a crossover from nearly constant width at small e to exponential broadening at large e occurs around the same “quantum” threshold e_q deformation (which is significantly smaller than the threshold for classical escape). Our pseudoclassical model cannot explain this behavior, because nothing dramatic happens to the invariant curves of the ellipse at e_q . We would expect the escape rate to be dominated by the tunneling escape from the points where the quantized invariant curve has its minima, and these minima do not suddenly change their dependence on e . Qualitatively, all resonance widths seem to follow a law of the form

$$\gamma = \gamma_0 [1 + \exp \alpha (e - e_q)], \quad (11.1)$$

with different slopes α , whereas the WKB tunneling rate from the minima of the invariant curves should yield a “threshold-less” expression of the form

$$\gamma = \gamma_0 \exp(\beta e) \quad (11.2)$$

for small e . At present, it is not clear if a perturbation expansion at higher orders in e in conjunction with the WKB tunneling rate is able to yield an expression displaying the observed threshold behavior.

One clue to an understanding of this threshold behavior, which does appear to be less sharply defined here than in the KAM billiards, might be the observation that tunneling between tori is in fact possible in the *open* ellipse since the separability of the wave equation is possible for Dirichlet boundary conditions but not for a partially transparent boundary as is considered here. Whereas the eigenfunctions of the ellipse with Dirichlet boundary conditions are simultaneous eigenfunctions of the operator describing the product of the two focal angular momenta, L_{12} , this is not true for the quasibound states of the open system. The quasibound states of the ellipse thus contain admixtures of different eigenfunctions of L_{12} , or in other words spread over some range of classical tori. This is in contrast to the circle, where L_0 (and hence L_{12}) is a good quantum number for the open as well as the closed system.

The first (constant) term in Eq. (11.1) can be attributed to the direct tunneling contained in our pseudoclassical model, and it should in fact have a small e dependence given by Eq. (11.2) with a small β . The second term in Eq. (11.1), on the other hand, can then be associated with cascades of tunneling processes between adjacent tori, for which the matrix element may be larger than for the direct process but which individually do not cover a large distance in phase space. At $e = e_q$ the number of participating tori is large enough to yield an escape channel that can compete with the direct process. This description is only a speculative scenario and needs to be quantified. However, it is intriguing to note the similarity between this picture and that of the quantum decay from a metastable well in the presence of dissipation¹⁴¹. There, one arrives at an escape law similar to Eq. (11.1) with e being replaced by the temperature, and the exponential being due to activated escape that competes with the quantum decay (γ_0).

All these considerations illustrate that chaos-assisted tunneling is harder to treat in the deformed billiard, because of the strong effect of the deformation on the regular states and their coupling among each other. This has to be subtracted in some way to isolate the genuine contribution due to chaotic diffusion. We leave the discussion of this interesting topic, because the work is still in progress.

To summarize, the resonance widths in the strongly deformed convex billiard show the predicted wavelength-independent coalescence to a value only determined by the refractive index and the semiclassical adiabatic curve, \tilde{p} , defining the starting condition for diffusion in the SOS. Corrections due to localization are observed at large deformations where the WG trajectories are already chaotic, and the importance of chaos-assisted tunneling is seen at intermediate deformations where chaos exists but has not reached \tilde{p} .

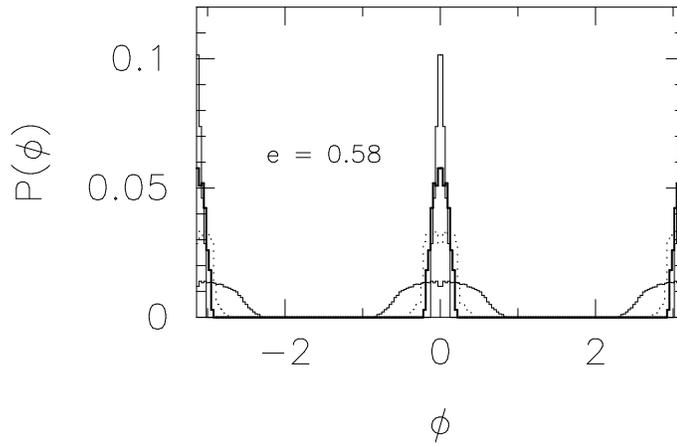


Figure 11.3: Classical escape directionality starting from $\tilde{p} \approx 0.85$ and $e = 0.58$. The broad solid histogram is predicted by the simple ergodic model, the dotted line results from the modified ergodic model (see text). The heavy line is obtained from the invariant curve model. The full ray simulation leads to the most sharply peaked distribution (solid line).

11.3 Emission directionality

As mentioned in the introduction, one of the questions of greatest interest in device applications of asymmetric resonant cavities is how to couple the emitted light into some other component, like an optical fiber. We have therefore used the ray-optic model to make predictions about the emission directionality of ARCs^{144,10,58}. The central feature that will be seen to cause directional emission is the fact that even chaotic trajectories do not move randomly when observed for short times. Short times come into play because to determine the escape position and orientation of a trajectory, one only needs to retrace its most recent history before the escape. We will arrive at this conclusion using two different classical models.

11.3.1 Ergodic model

As a first step to understanding the escape directionality distribution we formulate a simple model which should correctly represent the short-time ray dynamics near the critical angle. Significant contributions to this approach have been made by Atilla Mekis¹⁴⁶. Since motion in the chaotic region is ergodic (on that region) this suggests that as a starting point we can assume that the rays in the ensemble starting near the adiabatic curve \tilde{p}_{mq} will after some time fill the chaotic region above $\sin \chi_c$ approximately uniformly. To test this notion we see if the observed directionality can be reproduced by starting with a homogeneous distribution of starting conditions above $\sin \chi_c$ and iterating the map one step (allow each ray to collide once more with the boundary). The probability density $P(\phi)$ for escape in an interval $d\phi$ will then just be proportional to the area of the SOS mapped in one step from above to below $\sin \chi_c$ and into the interval $d\phi$. We call this set of assumptions the *ergodic model* for directionality. In order to recover the full symmetry of the billiard shape, we have to repeat this with an ensemble below $-\sin \chi_c$ (i.e. reversed starting momenta) and add the outcomes. The resulting approximation for the distribution $P(\phi)$ is shown in Fig. 11.3, and compared to the result of a simulation where an ensemble of rays is started near $\sin \chi = 0.85$ with homogenous distribution in ϕ , and their escape directions are recorded. For compatibility of the models, we for the moment leave out the contributions of tunneling and Fresnel reflection in the full simulation.

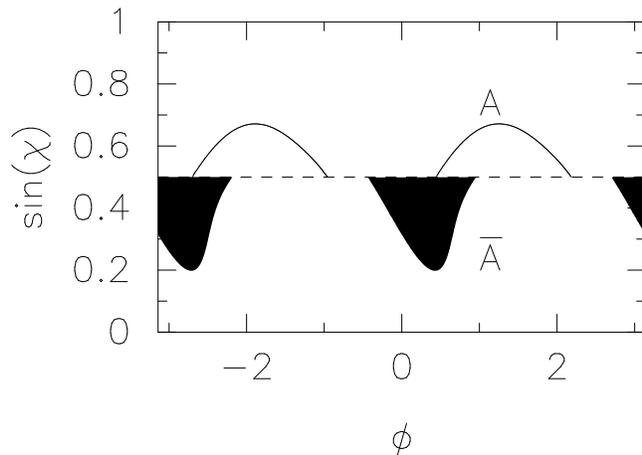


Figure 11.4: Illustration of the areas A and \bar{A} which end up below or come from above $\sin \chi_c$, respectively, in one mapping step.

Although the ergodic approximation overlaps well with the peaks in the true escape probability density, the $P(\phi)$ it predicts is clearly much broader than the exact ray-tracing distribution. Since the ergodic model includes the short-time effects of the varying curvature around the boundary, this indicates that the directionality of escape from ARCs is not simply obtained from a knowledge of curvature. Obviously the assumption of a uniform distribution of rays filling phase-space above $\sin \chi_c$ has missed some essential part of the physics. What has been missed is that even in the chaotic component there is a definite flow pattern in phase-space for short times. Due to this flow pattern, many initial conditions (above $\sin \chi_c$) which lead to escape in the ergodic model can only be reached if the *previous* reflection occurred with $\sin \chi < \sin \chi_c$, implying that the ray would already have escaped before getting to the assumed starting point in phase space. One can begin to take this into account within an extended ergodic model by assuming zero occupation probability for initial conditions above $\sin \chi_c$ which upon one-step iteration backwards are below $\sin \chi_c$ (and uniform probability for all other initial conditions). Fig. 11.3 shows that this extension of the ergodic model significantly improves the predicted $P(\phi)$ as compared to the true distribution.

The advantage of the ergodic model is that only one or (in the extended version) two mapping steps need be considered. For the simple version it is then possible to express $P(\phi)$ in terms of the one-step map, or in possibly the effective map derived in chapter 5. If we abbreviate $p \equiv \sin \chi$, the map can be described by the two functions $\bar{\phi}(\phi, p)$, $\bar{p}(\phi, p)$, giving the new position and momentum as a function of the old variables. The map is area-preserving, i.e. the Jacobian of the transformation $(\phi, p) \rightarrow (\bar{\phi}, \bar{p})$ is unity. One can alternatively specify the map by considering the old momentum and the *new* position as given so that the dependent variables are $\phi(\bar{\phi}, p)$ and $\bar{p}(\bar{\phi}, p)$. The differential probability of obtaining ϕ' after one mapping step applied to a homogenous starting distribution with $p > p_c$ is

$$P_1(\phi') = \frac{1}{|A|} \int_A d\phi dp \delta(\phi' - \bar{\phi}(\phi, p)), \quad (11.3)$$

where A is only the region above p_c that gets mapped below p_c , $|A|$ is its area, and $\bar{\phi}$ is the new position after one reflection. The image of A under the map, \bar{A} , has the same area $|\bar{A}| = |A|$ due to area preservation. These regions are shown in Fig. 11.4.

We make a transformation of variables in the integral from ϕ, p to $\bar{\phi}, \bar{p}$, yielding

$$P_1(\phi') = \frac{1}{|\bar{A}|} \int_{\bar{A}} d\bar{\phi} d\bar{p} \delta(\phi' - \bar{\phi}) \quad (11.4)$$

$$\equiv \frac{1}{|\bar{A}|} \int_{\min[\bar{p}(\phi')] }^{\max[\bar{p}(\phi')] } d\bar{p} \quad (11.5)$$

$$= \frac{1}{|\bar{A}|} (\max[\bar{p}(\phi')] - \min[\bar{p}(\phi')]). \quad (11.6)$$

Here, $\min[\bar{p}(\phi')]$ denotes the smallest \bar{p} within \bar{A} at the final angle $\bar{\phi} = \phi'$, analogously $\max[\bar{p}(\phi')]$. The boundary of \bar{A} is formed by the two curves $(\bar{\phi}, p_c)$ and $(\bar{\phi}, \bar{p}(\bar{\phi}, p_c))$. In our case, $\bar{p}(\bar{\phi}, p_c)$ is a unique function of $\bar{\phi}$, and as a result $\max[\bar{p}(\phi')] = p_c$, $\min[\bar{p}(\phi')] = \bar{p}(\phi', p_c)$. We therefore have the simple result that

$$P_1(\phi') = \begin{cases} \frac{1}{|\bar{A}|} (p_c - \bar{p}(\phi', p_c)) & \text{for } p_c > \bar{p}(\phi', p_c) \\ 0 & \text{otherwise.} \end{cases} \quad (11.7)$$

The distribution in Fig. 11.3 is obtained by forming the symmetrized function $P(\phi') = [P_1(\phi') + P_1(-\phi')]/2$.

For the extended ergodic model, we need the function $p(\phi, \bar{p})$. Then the lobes of $(\phi, p(\phi, p_c))$ above p_c delimit the region A . To decide whether a given $(\phi, p) \in A$ would have come from the region $p < p_c$ in the previous reflection, we simply invert the momentum p and ask whether the forward mapping yields a new momentum \bar{p} smaller *in magnitude* than p_c . This is true for a region B bounded above p_c by the curve $(\phi, -p(\phi, -p_c))$. The modified starting domain for the ergodic model is then $A' = A \setminus B$.

The ergodic model predicts anisotropic intensity in the nearfield, but not in the far-field. Since the ergodic models yield a final distribution of escaping trajectories that scatters widely in $\sin \chi$ (see \bar{A} in Fig. 11.4), any near-field directionality will be washed out in the far field after refraction is taken into account. On the other hand, we see in Fig. 9.2 that this is not the case. To provide an explanation for this persistence of directionality in the far field, we must be able to argue that in fact only $\sin \chi \approx \sin \chi_c$ occurs upon escape, thus causing the same degree of refraction for all escaping rays.

11.3.2 Diffusion between adiabatic curves

The solution to this problem is the slowness of diffusion in the direction of $p = \sin \chi$, as discussed in section 6.5. As explained there, the trajectory approximately follows an adiabatic curve \tilde{p} for intermediate times but eventually diffuses to smaller \tilde{p} . The first chance for classical escape (neglecting tunneling) therefore presents itself when the diffusion has reached the *tangent adiabatic curve* \tilde{p}_c which just touches the critical line for total internal reflection in the SOS, $\sin \chi_c = 1/n$. This is the situation depicted in Fig. 10.3 for the case $n = 2$ on which we focus for the moment.

The points of tangency are invariably the points of highest curvature of the billiard. This follows from Eq. (6.55) for the adiabatic curve, which has its minima at the points of maximum $\kappa(\phi)$. This is intuitively understandable, because it is to be expected that the highest radiation losses occur at the most sharply curved regions of the surface. In the ray picture, the high-curvature points are where the trajectory

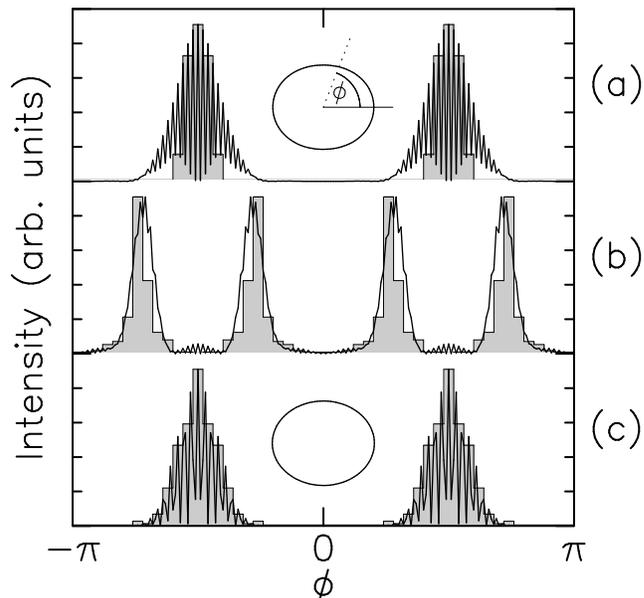


Figure 11.5: Far-field intensity distribution for the quadrupole (a,b) and the ellipse (c) at eccentricity $e = 0.6$, for a resonance with $m = 68$. The wave results are shown as solid lines, the pseudoclassical predictions as shaded histograms. The refractive index is $n = 2$ in (a), $n = 1.54$ in (b,c). The resonance positions are (a) $kR = 45.2$, (b) $kR = 48.1$, (c) $kR = 48.0$. The shapes of ellipse and quadrupole are shown in the insets.

has the highest likelihood of impinging with a small enough angle for escape, because it has to “turn the sharpest corner” there. We will see below, however, that this intuitive picture does not always hold.

The ray model allows us to predict furthermore that even the far-field emission pattern will be highly directional. The process of diffusion in p is slow, so that rays explore the adiabatic curves evenly (in ϕ) before moving further down in \tilde{p} . Therefore, a trajectory that has reached \tilde{p}_c will flow close to this tangent curve until it reaches the minima where escape then occurs. Consequently, escaping rays will predominantly have $\sin \chi$ near $1/n$. The escape itself to a good approximation follows Snell’s law,

$$\sin \chi_{out} = n \sin \chi, \quad (11.8)$$

where χ_{out} is the angle of the refracted ray with respect to the outward normal. This follows from the good agreement between Fresnel’s formula and p_0 shown in Fig. 8.5. Inserting here $\sin \chi \approx 1/n$, we conclude that those rays that leave the cavity do so almost tangent to the surface at the high curvature points.

For the quadrupole, the high-curvature points are at $\phi = 0$ and $\phi = \pi$, so that we expect the far-field emission to be highly peaked at $\phi = \pm\pi/2$. This is confirmed in Fig. 11.5(a). The classical directionality histograms are obtained from the pseudoclassical model, starting an ensemble of rays on the semiclassically quantized adiabatic curve \tilde{p}_{mq} , and recording the direction in which it escapes, using the probability function $p_0(\sin \chi, kR)$. This simulation therefore also contains the contribution of tunneling and above-barrier reflection, i.e. a ray can escape before reaching the tangent adiabatic curve and likewise need not escape the first time it crosses below $\sin \chi_c$. These wave effects, clearly did not invalidate the classical prediction just given. To appreciate the smoothing effect that tunneling and Fresnel reflection have, we plot a purely classical histogram for the same resonance in Fig. 11.6(a). These were obtained by taking p_0 to be a unit step function that is zero

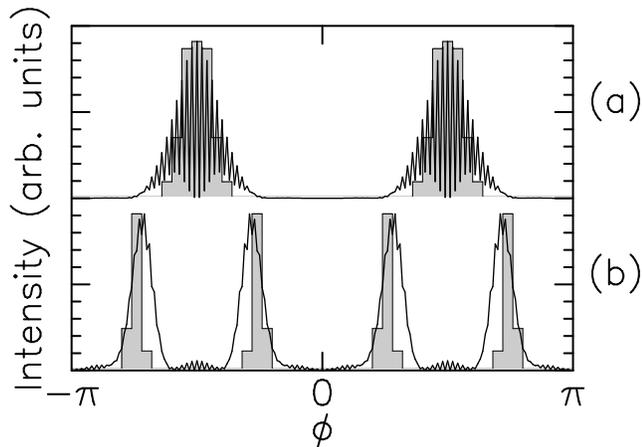


Figure 11.6: Far-field intensity distribution for the quadrupole at the same parameters as in Fig. 11.5(a,b). The histograms here show the result of a simulation neglecting above-barrier reflection and tunneling.

below $\sin \chi = 1/n$.

11.3.3 Dynamical eclipsing

An interesting effect occurs when there is island structure right at the points of maximum κ , intersecting the critical line for escape. This situation can be brought about in the quadrupole by changing the refractive index to $n = 1.54$, as shown in Fig. 10.3. The tangent adiabatic curve, which is also drawn in the figure, clearly does not describe the trajectories in this region of the SOS because the islands are transport barriers. Phase space diffusion must bypass them, and as mentioned in subsection 6.6.3 this occurs in the form of a rotation around the islands. As a result, escape now mainly occurs at those ϕ where a trajectory rotating around an island first crosses $\sin \chi_c$, i.e. on either side of the two islands intersecting the critical line (both sides of the island contribute because we have to take both rotation senses into account, corresponding to time-reversed pairs of orbits). In the near-field emission pattern, one should therefore observe a *suppression* of the intensity right at the high-curvature points. Since escape still follows Snell's law, and escaping rays still emanate roughly tangent to the surface, this suppression will propagate into the far field. This is confirmed in Fig. 11.5(b), where we now see four equally spaced peaks instead of two. We call this effect *dynamical eclipsing*, and it is an unambiguous fingerprint of the underlying classical phase space structure in the wave solution. This is noteworthy because the value of $kR = 48.1$ used here is by no means deep in the semiclassical regime.

To further emphasize the sensitivity of the wave equation to the classical phase space structure, we show in Fig. 11.5(c) the directionality of the identical resonance (same quantum numbers) as in (b), for the same index of refraction and same eccentricity, the only difference being that we now choose an elliptical deformation. As can be seen in the insets, the shapes of quadrupole and ellipse are barely distinguishable, but the dynamical eclipsing is nonetheless absent in (c). This conforms to the classical prediction because there is no island structure in the ellipse at $\sin \chi \approx 0.65$ as its dynamics is integrable.

The pseudoclassical model allows us to produce a directionality histogram for the ellipse, despite the fact that escape can only occur by tunneling since trajectories move on invariant curves that do not intersect the critical line. The result is seen

to agree well with the wave solution, even though we could make no purely classical prediction for the ellipse in the absence of diffusion. The reason for the intensity distribution in the ellipse is that the tunneling escape occurs mainly from the minima of its invariant curves, because that is where the tunnel reflectivity p_0 is smallest. An escaping ray will again emanate tangentially to the interface, because any other $\sin \chi_{out}$ would require the incident $\sin \chi$ to have been below $1/n$. Agreement for the dynamical eclipsing is slightly better in the pseudoclassical model than in the purely classical simulation, cf. Fig. 11.6(b).

11.3.4 Universal directionality

In the previous two subsections we have seen remarkable agreement with the ray simulations that is found without exception for all classically escaping WG modes we studied. This shows that the pseudoclassical, and even the classical, model is a good theory for the emission directionality, unaffected by the various approximations that appear to have such a strong effect on the width calculations.

The classical model implies that only the phase space flow near the critical line is of importance for the emission directionality, because the trajectory loses the memory of its starting position during the chaotic diffusion preceding the escape. In the absence of dynamical eclipsing, all that counts is that the tangent adiabatic curve be reached eventually, and the directionality is then prescribed. The same can be said for the flow around the islands if dynamical eclipsing occurs. As a consequence, the emission directionality is expected to be *the same* for all resonances whose semiclassical quantization involves adiabatic curves \tilde{p}_{mq} which are far enough above the critical line for escape, or in other words

$$\tilde{p}_{mq} > \tilde{p}_c. \quad (11.9)$$

This is shown in Fig. 11.7 for $n = 2$ where the tangent adiabatic curve is valid and escape thus should occur at the high-curvature points, and in Fig. 11.8 for $n = 1.54$ where we expect dynamical eclipsing.

The fact that the emission directionality is determined solely by the shape and the refractive index should work in favour of an experimental verification of our results. While dynamical eclipsing has not yet been observed, an experiment was recently conducted which confirms the emission from the high curvature points⁵⁸. This was done by creating a cylindrical stream of ethanol containing a lasing dye, which had an oval cross section due to the rectangular orifice at which it was produced. The far-field intensity was found to be peaked, with two maxima in agreement with our discussion above. An observation of importance for device fabrication is that the directionality is also largely independent of deformation beyond some transition region. This is illustrated in Fig. 11.9, showing essentially the same intensity distribution above $e = 0.3$. At $e = 0.3$, only tunneling escape is possible. As in the ellipse, we still have escape predominantly from the minima of the invariant curve on which the ray moves. The conclusion is that this configuration allows us to tune the resonance *width* over a large interval of practically exponential dependence on e , while the directionality stays unaffected. In particular, the directionality in the tunneling regime is correctly predicted by the pseudoclassical model.

In the case $n = 1.54$, dynamical eclipsing only occurs after the islands responsible for it have grown to sufficient size. Before that point, the emission looks similar to that of the billiard with $n = 2$. As shown in Fig. 11.10, the four-peak structure has fully developed at $e = 0.45$, again well before chaotic diffusion becomes possible.

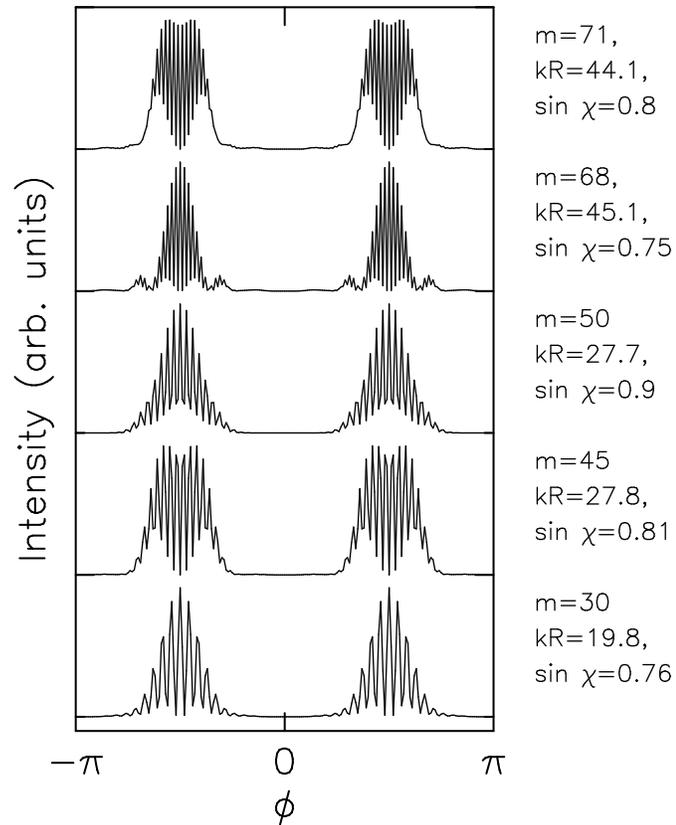


Figure 11.7: Emission directionality in the far-field of the quadrupole at eccentricity $e = 0.65$ for 5 different resonances with various kR and $\sin \chi$ (numbers given for the circle). The refractive index is $n = 2$.

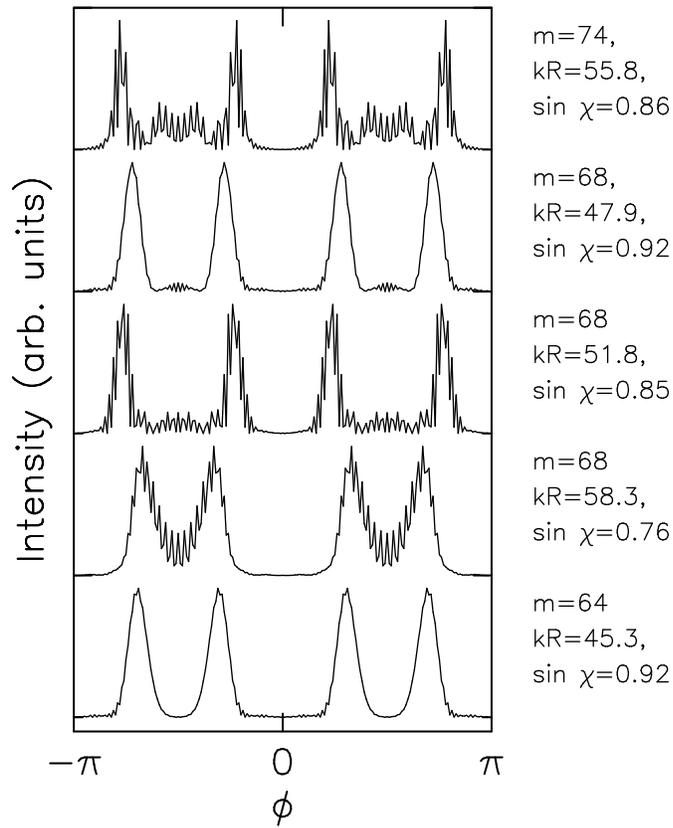


Figure 11.8: Far-field directionality for 5 different resonances of the quadrupole at eccentricity $e = 0.65$ and refractive index $n = 1.54$, displaying the peak splitting due to dynamical eclipsing. Note that the resonance with $\sin \chi = 0.76$ starts on an adiabatic curve very close to the islands intersecting the critical line $\sin \chi_c = 0.65$. Therefore tunneling into the island or direct tunnel escape could explain the incomplete suppression at $\phi = \pm\pi$.

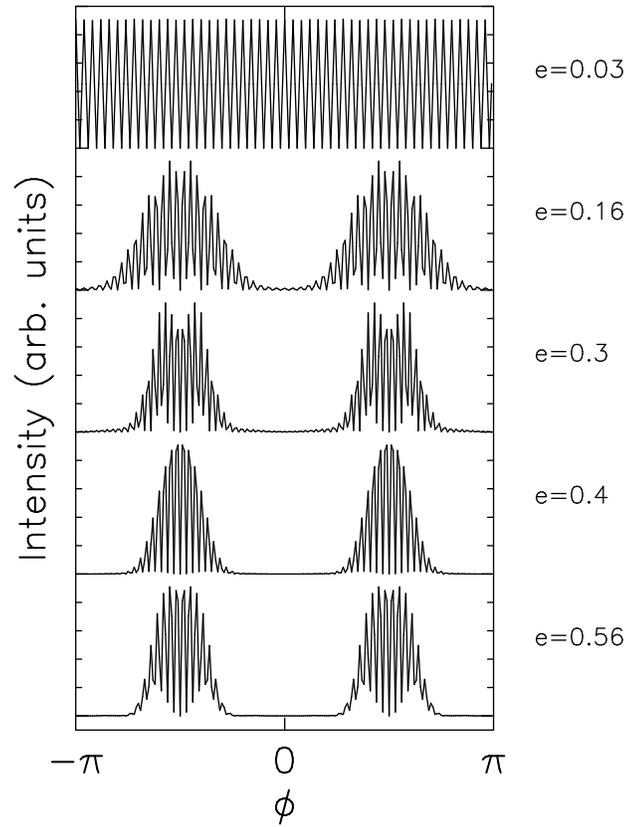


Figure 11.9: Far-field directionality in the quadrupole with increasing eccentricity e at $n = 2$ for the resonance with $m = 45$, $kR = 27.8$ also appearing in Fig. 11.7.

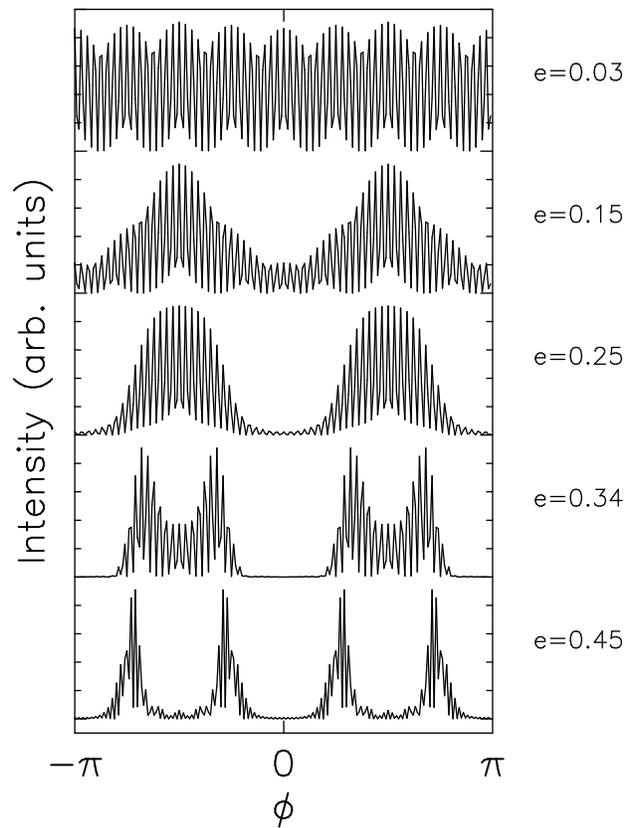


Figure 11.10: Far-field directionality in the quadrupole with increasing eccentricity e at $n = 1.54$ for the same resonance as in Fig. 11.5.

11.4 Directional lasing emission from dye jets

The classically predicted universal emission directionality has been verified experimentally by observing lasing dye jets, as will be described later. An important property of the experimental setup is the fact that many different modes take part in the laser action. We therefore have to ask which resonances will make a dominant contribution to the observed laser output under such multimode conditions. In particular, it has to be clarified if the classically escaping orbits which we studied will be relevant in the laser, given the fact that for convex cross sections there are always much longer-lived resonances corresponding to Lazutkin's caustics.

If the latter (regular) orbits were the dominant ones for lasing, dynamical eclipsing would not be observable because the lasing emission from states on invariant curves is always along tangents projecting outward from the high-curvature points, as we just demonstrated. Although a thorough analysis is still missing, we can give a qualitative argument that tells us which modes should become most important when the lasing condition is satisfied for more than one mode.

11.4.1 Multimode lasing

Lasing requires a gain medium and a cavity. The gain medium provides amplification of a light wave traveling in the cavity, depending on the pump power P supplied to it. When P exceeds the *lasing threshold* P_t , the gain exceeds the losses due to absorption, leakage from the cavity etc. Consider a given cavity mode with a loss rate $1/\tau$ and a number N of photons in it. In the limit of a clean resonator, τ is just the resonance lifetime discussed above. To maintain a steady-state laser action, the escape of photons from the cavity must be compensated precisely by the stimulated emission into the same mode. The latter is proportional to the number N_i of inverted atoms (or molecules) that interact with the mode, and to the intensity of the existing field. Therefore, we can write the stationarity condition as

$$0 = \frac{dN}{dt} = B N_i N - \frac{N}{\tau} \quad (11.10)$$

where B is the Einstein coefficient for induced emission. After canceling N we are left with

$$N_i = \frac{1}{B\tau} \quad (11.11)$$

which is independent of the pump power. The requirement of steady state therefore implies that the inversion N_i is *clamped* to a constant value as soon as P exceeds P_t .

For the cavities of interest here, one has to assume that many modes have spatial overlap with the gain medium, although their respective τ may vary widely. After the first mode starts to lase, we could stop increasing P and would thus obtain a single-mode laser. If P grows further, the original mode continues to lase with the same N_i as at threshold ($P = P_t$), but other modes may also satisfy the lasing condition that their modal loss be made up for by their modal gain. This is possible if the spatial overlap of the original mode and the new mode is incomplete, so that one has nodes where the other has antinodes. Since the interaction with the gain medium is suppressed in the neighborhood of field nodes, two such modes can interact with different atoms. The result is that the second mode can indeed lase, producing its own collection of inverted atoms N'_i . Let the threshold for this second mode be P'_t . Its loss is larger than that of the first mode $\tau' < \tau$, corresponding to

$P'_t > P_t$. The interesting observation here is that according to Eq. (11.11) we have

$$N'_i > N_i. \quad (11.12)$$

If we add the fact that the amount of pump energy converted into lasing emission grows with the inversion, this leads to the statement that *the lowest- τ lasing mode carries the largest emission energy*. The same can be said in the presence of more than two lasing modes.

The lasing spectra obtained from liquid spheres and jets containing a dye^{145,58} do indeed show multimode operation. While the longest-lived regular WG states are always among the lasing modes, one can now see how the emission directionality should be dominated by those states whose lifetime is long enough to meet the lasing condition but shortened due to classical escape.

It is then only a matter of achieving the required refractive index before dynamical eclipsing should be seen experimentally. In the absence of a suitable liquid for this purpose, a more immediate goal of an initial experiment is to test the universality of the emission directionality. This test has been performed successfully.

11.4.2 Observations on lasing jets

The experiment consists of a liquid dye jet that is optically pumped to induce lasing. Details of the setup are provided in Ref.⁵⁸. The refractive index of this liquid is approximately $n \approx 1.3$, which yields an escape condition above the phase space region where dynamical eclipsing occurs. A deformed cross section of this lasing dielectric column is achieved by forcing the liquid through a narrow rectangular opening of dimensions $1000\mu\text{m} \times 25\mu\text{m}$. The cross section oscillates as the liquid falls, with the long axis either parallel or perpendicular to the slit that creates the distortion. This deformation furthermore decays with the distance from the orifice until the cross section is nearly circular 2cm downstream, and in fact the multipole component with the longest decay time should be the quadrupolar one. We expect far-field peaks at $\pm 90^\circ$ from the long axis of the instantaneous cross section. This directionality would be washed out if different WG resonances had different directionality.

The lasing jet is imaged by two detectors with observation directions at right angles to each other, cf. Fig. 11.11. While one detector shows bright emission from the sides of the jet, the other image is dark, and vice versa. Moreover, the transition between brightness and darkness along the vertical direction is very sharp. The short intervals along the jet where no emission is seen in either direction correspond to circular cross section. One thus concludes that the directionality is present even for small deformations. This is the expected behavior, cf. the previous section. The fact that pronounced minima and maxima exists in the emission characteristic at right angles to each other is a consequence of the universal directionality of all lasing modes in this case. In a control experiment on a jet of circular cross section, produced by a circular orifice of radius $75\mu\text{m}$, the emission is seen to be equally bright in both detectors, as shown in Fig. 11.11.

To conclude this chapter, we note that the directionality can be predicted by inspection of the Poincaré section alone, without performing the pseudoclassical or even the classical simulations. This is because the anisotropy of emission follows from the local structure of the SOS near $\sin \chi_c$, unlike the resonance width which depends on the whole diffusive path taken by a trajectory starting from the semiclassical adiabatic curve. To produce a SOS, it generally suffices to follow a relatively

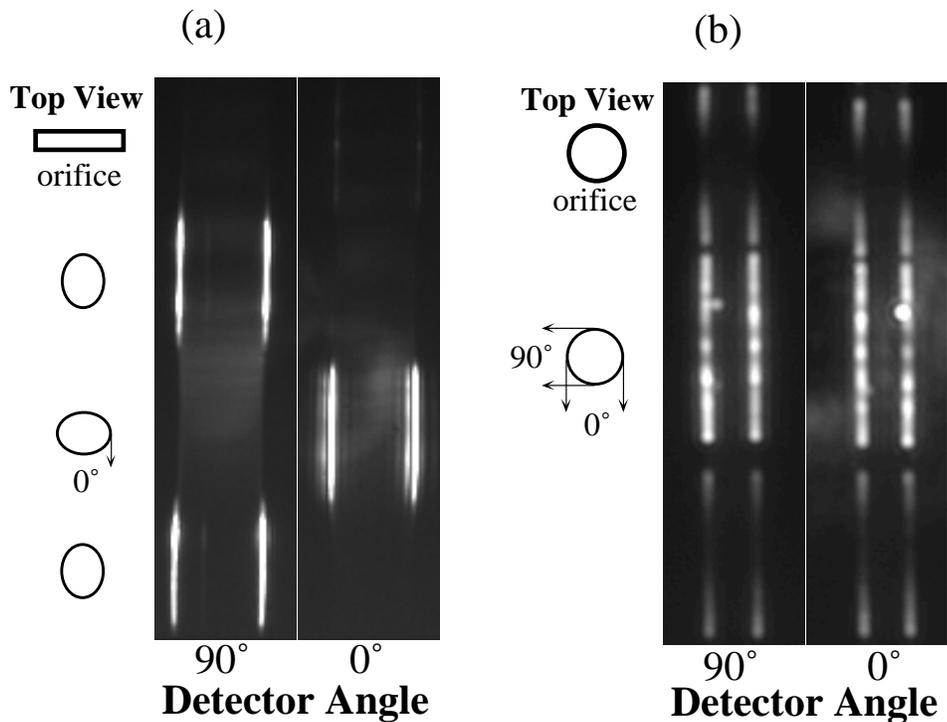


Figure 11.11: Total lasing intensity images of vertically flowing jets of Rhodamine B dye in ethanol. The jet with deformed cross section (left) is observed perpendicular and parallel to the slit from which the liquid originates, giving rise to the two intensity traces. On the righthand side is shown the same observation for a liquid column of circular cross section.

small number of trajectories, say 20, for about 500 reflections. The more detailed ray simulations require ensembles of 1000 trajectories which have to be followed for a potentially very long time. The subject of emission directionality is therefore an excellent example for the simplification that results when we can turn to an analysis of the ray dynamics in the form of the SOS, rather than performing lengthy ray tracing. Of course, even the pseudoclassical model is a far more efficient way of extracting information about widths and directionality (the latter with high accuracy) than numerical solutions of the wave equation.

Chapter 12

Thresholds and Intensity Distribution in Lasing Droplets

In this chapter we apply the ray-optics model for ARCs to the description of the WG modes of deformed liquid droplets and give an explanation for the observed anisotropy of lasing emission from droplets, shown in Fig. 12.1^{9,10}.

12.1 Experimental configuration

In the experiment, a stream of ethanol droplets containing Rhodamine-B dye with average radius $\approx 30\mu\text{m}$ is created at the vibrating orifice of a Berglund-Liu generator⁸. At the orifice the droplets are highly non-spherical and as they fall they undergo damped oscillations between oblate and prolate configurations (Fig. 12.1) driven by excess surface tension until they relax to highly spherical shape far downstream. The period of the shape oscillations is of order $50\mu\text{s}$, which is far longer than the lifetime of the WG resonances of $\approx 10\text{ns}$ ¹⁴⁷. Thus we may treat the different phases of the droplet oscillation as static examples of oblate, spherical or prolate micro-cavity lasers and analyze the angular emission intensity in terms of the theory of ARCs. Previously we have focused on cylindrical ARCs deformed perpendicular to their axes; here we apply the ray-optics model to dielectric spheres deformed so as to preserve azimuthal symmetry (the deformed droplets retain this symmetry as well as reflection symmetry through the equator to a good approximation).

As seen in Fig. 12.1, laser emission is fairly isotropic in the sphere, but gets suppressed near the poles of both the oblate and prolate droplets. Furthermore, the oblate shape is brightest around the equator whereas the prolate shape of the same deformation (ratio of long to short axes) emits most strongly from regions around $\theta \approx 30^\circ - 45^\circ$. These observations are stable over a range of prolate and oblate shapes (exceeding a certain degree of deformation) and are independent of the direction of the optical pumping. It is important to note that the three total-energy images are normalized individually with white referring to the maximum emission, which is of different magnitude in each shape.

12.2 The centrifugal billiard

12.2.1 Justification for a classical treatment

Since axial symmetry is preserved for the droplets, their instantaneous shape can be specified in spherical coordinates by $r_b(\theta)$, independent of azimuthal angle ϕ .

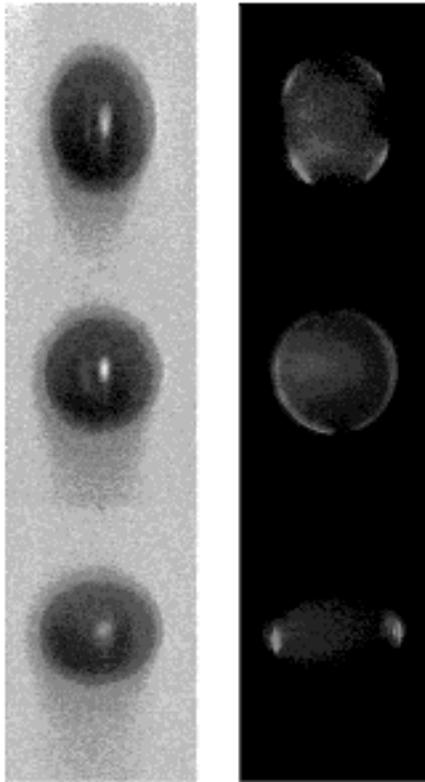


Figure 12.1: Shadow graphs (left) and simultaneous total-energy images (right) of three lasing droplets falling in air taken at different phases of oscillation: prolate (top), spherical (middle) and oblate (bottom). Light regions in (b) indicate lasing.

The damping is weakest for the low multipole components of the oscillation, so we include only the Legendre polynomials $P_0(\cos \theta)$, $P_2(\cos \theta)$, $P_4(\cos \theta)$ in an expansion of the shape. Odd orders do not appear because of the approximate symmetry $r_b(\theta) = r_b(-\theta)$ noted above. The particular shapes we use to model this behavior more realistically are

$$r(\theta) = 1 + \epsilon \left(\cos^2 \theta + \frac{3}{2} \cos^4 \theta \right) \quad (12.1)$$

for prolate deformations; an oblate shape with the same axis ratio [equal to $(5/2)\epsilon$] is obtained by replacing $\cos \theta$ with $\sin \theta$. For these shapes, we expect to find chaotic ray dynamics, because only the closed ellipsoid is integrable. That chaos results even for ellipsoidal shapes if the potential is not a hard wall, was observed by Brut and Arvieu⁷⁵, who also proposed a semiclassical treatment of such systems based on an adiabatic switching procedure. It uses only the known semiclassics of the sphere, which fixes the invariant curves associated with a quantum state; the deformation is then introduced into the classical dynamics by starting trajectories on the above invariant curve and changing the shape slowly on the scale of the time between reflections. After this process, the trajectories move on new invariant (KAM) curves, and the semiclassical energy shift can be obtained from the work done by the moving walls during each reflection. It is not clear at present how well this method can be extended into a regime of deformations where the relevant KAM curves break up themselves. To test this, the analogous algorithm has been programmed by the present author for the 2D convex billiard. Results of this method will however not

be included here because of their preliminary nature. In the context of lasing from droplets, we are in fact relieved of the requirement for a semiclassical quantization method if it is justified to assume multimode lasing from quasibound states with many possible quantum numbers, the only criterium being that their lifetimes exceed the lasing threshold. In this way we shall argue that the directionality can be understood by way of an average over all these states, which is then replaced by an average over the classical phase space.

12.2.2 Classical ray dynamics

The ray dynamics analysis is facilitated by the axial symmetry of the droplets which implies (in the language of particle trajectories) that the z component of angular momentum, L_z , is conserved. At any given L_z and total energy E , the equations of motion thus have only two degrees of freedom, just as in the deformed cylinder. This becomes explicit in cylindrical coordinates ρ , ϕ , z where one has

$$E = \frac{1}{2}m(\dot{\rho}^2 + \dot{z}^2) + \frac{L_z^2}{2m\rho^2}. \quad (12.2)$$

Let us look at the dynamics projected into the 2D (ρ, z) coordinate system. Each specular reflection causes a discontinuous change in $\dot{\rho}$ and \dot{z} ; however the angular velocity $\dot{\phi}$ remains unchanged because the normal to the surface of an axisymmetric cavity is always perpendicular to the ϕ direction. Thus a 3D specular reflection simply reverses the normal component of the 2D projected velocity $(\dot{\rho}, \dot{z})$ and reflections are also specular in the projected coordinates. Reflections occur whenever the trajectory $\rho(z)$ intersects the boundary curve $\rho_b(z)$. Between reflections the particle motion is free, $\dot{z} = \text{const}$, and Eq. (12.2) can be integrated to find $\rho(t)$. It can be shown that $\rho^2(z)$ describes a parabola whose vertex is the point of closest approach to the z -axis and whose intersections with the squared boundary curve $\rho_b^2(z)$ are the collision points. The curved trajectories in the z - ρ -plane between specular bounces [see the inset to Fig. 12.2(a)] are to be contrasted with the straight paths in conventional 2D billiards where the centrifugal potential $L_z^2/(2m\rho^2)$ is absent. For this reason we call this new class of systems *centrifugal billiards*.

To discuss the resulting dynamics we introduce dimensionless variables in Eq. (12.2) by setting $E = 1/2$ and $m = 1$. Then one has

$$1 = \dot{\rho}^2 + \dot{z}^2 + \frac{L_z^2}{\rho^2} \quad (12.3)$$

where $0 \leq L_z \leq \rho_b(z_{max})$ is the maximum distance from the z -axis is $\rho_b(z_{max})$. To simplify notation we assume that the droplets have their widest transverse cross-section in the equatorial plane, i.e. $z_{max} = 0$. Again the escape condition is simply $\sin \chi < \sin \chi_c$, where $\sin \chi$ is the angle of incidence with respect to the surface normal \mathbf{n} at the reflection point. This is not the same as the normal angle in the ρ - z -plane, as can be seen by considering a trajectory reflecting entirely in the equatorial plane at nonzero $\sin \chi$; its apparent angle of incidence in the ρ - z -plane will be zero. In our units $\cos \chi = \mathbf{n} \cdot \mathbf{v}/|v| = n_\rho \dot{\rho} + n_z \dot{z}$ since the total velocity is $v = \sqrt{2E/m} = 1$. The angle in the ρ - z -plane is then given by $\cos \chi_{\rho z} = \cos \chi / \sqrt{\dot{\rho}^2 + \dot{z}^2}$. It is convenient in the plotting of Poincaré sections to use as variables the polar angle θ and the 3D $\sin \chi$ at each reflection since in these coordinates the escape condition is still satisfied along a horizontal straight line.

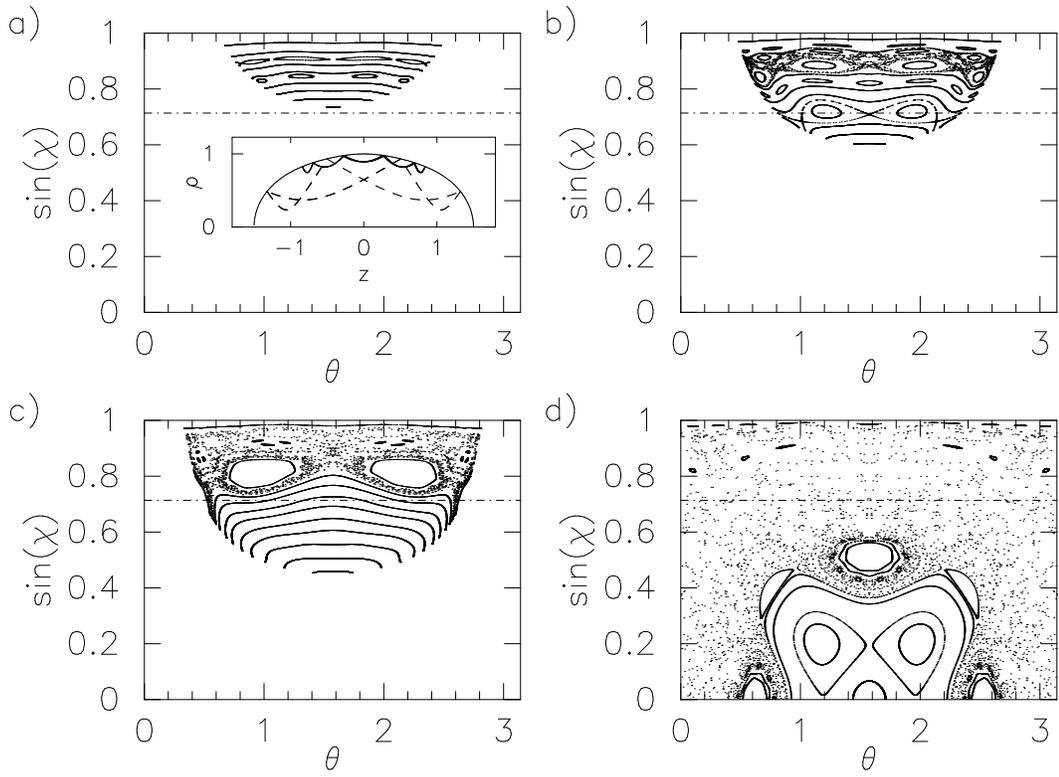


Figure 12.2: Poincaré surfaces of section for prolate droplets with $\epsilon = 0.2$ for $L_z = 0.735$ (a), $L_z = 0.6$ (b), $L_z = 0.45$ (c) and $L_z = 0$ (d). The dash-dotted lines denote $\sin \chi_c = 1/n = .735$ corresponding to the experimental value of $n = 1.36$ for the refractive index of the droplets in Fig. 12.1. Inset in (a) shows the droplet shape in the $z - \rho$ plane and special (periodic) trajectories for $L_z = 0.735$ (solid line) and $L_z = 0.2$ (dashed line).

At nonzero L_z certain regions of the SOS are forbidden due to the L_z angular momentum barrier (e.g. a ray reaching the pole ($\theta = 0$) must have $L_z = 0$). For the allowed bounce coordinates θ , $\sin \chi$ one finds the inequality $\sin \chi \geq L_z / \rho_b(z(\theta))$, where $z(\theta) = r_b(\theta) \sin \theta$; this relation delimits the empty spaces in the SOS's of Fig. 12.2(a-d) which were made for a prolate shape of fixed deformation ($\epsilon = 0.2$) and varying values of L_z . Before discussing ray escape in the deformed droplets it is important to note that as we proceed from higher to lower L_z in Fig. 12.2(a-d) in addition to the excluded regions of the SOS decreasing (because the angular momentum barrier becomes weaker) the degree of chaos grows rapidly. There is actually no visible chaos in Fig. 12.2(a) and a mostly chaotic SOS for $L_z = 0$ (Fig. 12.2(d)) *for a droplet of fixed deformation*. The reason for this is that high L_z trajectories are confined near the equator and a cross-section of the droplet at the equator is perfectly circular, i.e. high L_z orbits see an effective deformation which is much weaker than polar orbits ($L_z = 0$) which travel in the most deformed cross-section of the droplet. The effective deformation varies approximately as $\epsilon_{eff} = \epsilon \sqrt{1 - L_z^2 / \rho_b^2(0)}$ and tends to zero at the maximum allowed value of L_z . Thus as long as ϵ is large enough to induce classical Q-spoiling for the $L_z = 0$ orbits of interest, by looking at different L_z values for a fixed deformation one can study the classical Q-spoiling transition in a single ARC. We have illustrated this situation in Fig. 12.2(a-d).

Note that there is an absolute minimum allowed $\sin \chi \equiv \sin \chi_m$ which occurs at the equator ($\theta = \pi/2$) where ρ_b is maximal (i.e., $\sin \chi_m = L_z / \rho_b(0)$). This implies that classical ray escape is entirely forbidden due to the angular momentum

barrier for values of $L_z \geq \rho_b(0) \sin \chi_c$; such a case is shown in Fig. 12.2(a). As just noted these high L_z modes are confined to orbits near the plane of the equator [see also the inset to Fig. 12.2(a)]; since classical escape is forbidden for these modes we always expect to find high-Q WG modes in the equatorial region of axially-symmetric deformed microspheres. Since this follows simply from L_z conservation it will be true in both the oblate and prolate shapes.

Proceeding now to lower L_z in Fig. 12.2(b) we see that the angular momentum barrier has weakened enough that the allowed region of the SOS passes through $\sin \chi_c$ and rays with this value of L_z can escape. However as before WG modes will be associated with rays starting at large $\sin \chi \approx 0.9$ in this case. These rays are unable to reach $\sin \chi_c$ due to remaining KAM curves just as we saw earlier in our discussion of dielectric cylinders. Therefore we expect high Q WG modes for this value of L_z as well. This situation persists all the way to $L_z = 0$ for deformations less than roughly 5% of the radius, so we expect little Q-spoiling and approximately isotropic emission for smaller deformations than this.

However for the 50% deformation used in Fig. 12.2(a-d) reducing L_z a little more causes the appearance of regions of chaos which extend from high $\sin \chi$ across $\sin \chi_c$ allowing classical Q-spoiling of the WG modes. We expect all modes with L_z less than this value to have their Q rapidly degraded. As the Q of these modes decreases it will fall below the threshold Q-value to support lasing and these modes will go dark. But these low L_z modes are the only ones which can emit from the polar regions because of the angular momentum barrier for the high L_z modes. Therefore our model explains naturally why the polar regions are dark while the droplet still lases. The low L_z modes which could emit from the poles have too low Q to lase and the high Q modes which support lasing are confined away from the polar regions. This argument holds for both the oblate and prolate deformations in agreement with observations.

We are left with the question of why the emission profiles are nonetheless so *different* in prolate versus oblate shapes. To answer this question we must look at where the stable islands which block chaotic escape occur for the two types of deformations. The prolate shape corresponds to a stretching of the droplet in the vertical direction and a compression in the equatorial plane. Because it is compressed in the equatorial plane there exists a large stable island at $\theta = \pi/2$ corresponding to the two-bounce diametral orbit of the type we discussed in the 2D case in subsection 4.2. This island appears (distorted due to the L_z barrier) clearly in Fig. 12.2(c,d). As discussed above in subsection 11.3.3, a stable island intersecting the critical line will prevent the classical escape in the corresponding directions; thus in the case of Fig. 12.2(c) escape is blocked over an interval in polar angle centered at the equator.

In the oblate droplet the situation is reversed. The polar diameter is compressed and the equatorial diameter is stretched and the stable two-bounce orbit (if it is still stable) would appear at $\theta = 0, \pi$. But for most L_z the island around this orbit is unreachable and it has little effect on the dynamics. In Fig. 12.3(a) and (b) the SOS for the prolate and oblate shapes are compared at equal deformation and equal ratio $L_z/\rho_b(0)$. Indeed in the SOS for the oblate shape the regular island centered on $\theta = \pi/2$ and $\sin \chi_m$ is absent because equatorial orbits with low $\sin \chi$ are now unstable. One sees no effect of the stable islands at $\theta = 0, \pi$ due to the angular momentum barrier. There is still an island near the critical line at $\theta = 0$ and $\sin \chi = 0.6$ for this particular L_z , but its origin (a three-bounce orbit in the $\rho-z$ plane) as well as its effect are quite different from the prolate shape. In fact, all downward-diffusing trajectories first cross $\sin \chi_c$ in the vicinity of this island because

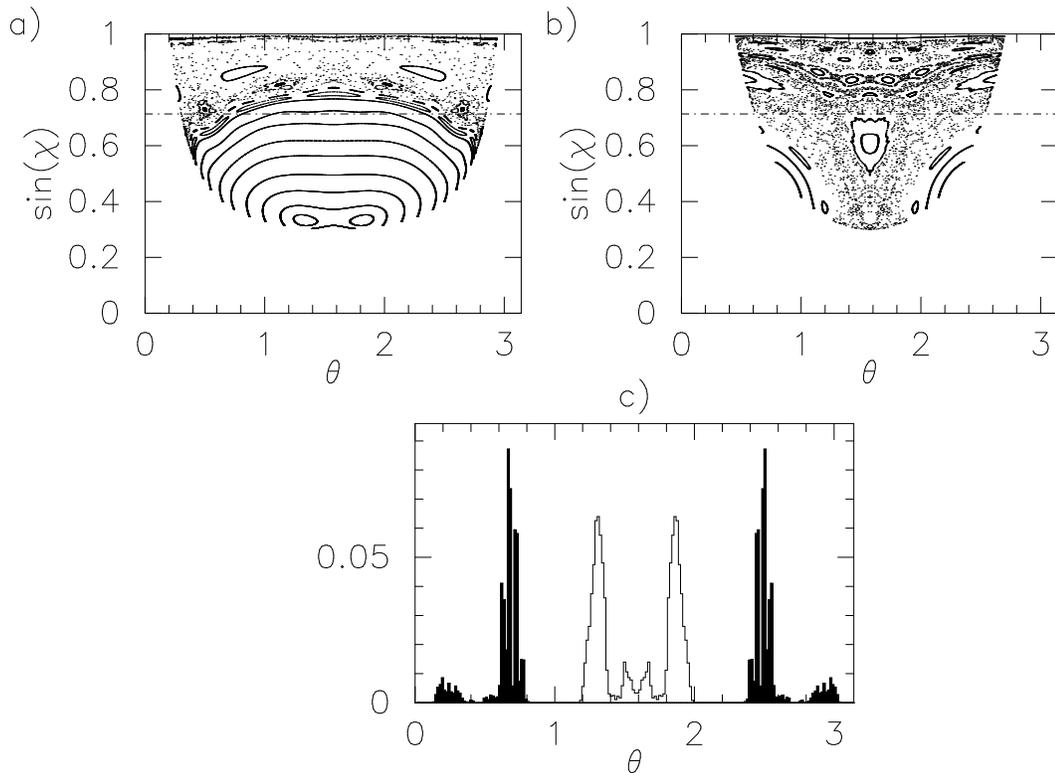


Figure 12.3: Prolate (a) and oblate (b) droplet SOS's at $\epsilon = 0.2$ and $L_z/\rho_b(0) = 0.3$. As in the previous Figure, the dash-dotted lines indicate $\sin \chi_c$. The escape directionality is shown in (c) for the prolate (filled histogram) and oblate (white) shape.

phase-space flow (see subsection 11.3.3) roughly follows a V-shaped curve connecting the three islands located at $\sin \chi = 0.825$ ($\theta = \pm 0.5$) and $\sin \chi = 0.6$ ($\theta = 0$). Hence escape is concentrated in the equatorial region. This remains true even if the island actually intersects the critical line (as is the case for higher L_z), thus blocking escape right at $\theta = 0$ – in that case escape still occurs very close to the island.

12.3 The prolate-oblate difference

Our explanation for the differing intensity profiles is as follows. First, it is reasonable to assume that all modes for which classical escape is completely forbidden will lase. As L_z decreases and classical escape begins to occur there will exist modes which are lower Q, but still high enough Q to lase. These modes will have highly directional emission in the polar angle (they will of course emit uniformly in the azimuthal angle). In the prolate case as L_z decreases the first regions of chaos which connect the WG orbits to the critical line do so only in small intervals around $\theta = 30^\circ$ and $\theta = 150^\circ$ (see Figs. 12.2(c) and 12.3(a)). These values of L_z will correspond to the lower Q lasing modes just discussed and will thus emit in small bright bands around these latitudes. In contrast, in the oblate droplets phase space flow curves touch the critical line near the equator (see Fig. 12.3(b)) and the lower Q lasing modes will emit around $\theta = 0$. However the total lasing intensity will be due to all the lasing modes and there will typically also be many high L_z modes of higher Q which can emit at the equator through the same processes as in the undeformed droplet. But unlike the undeformed case, these modes are competing with lower Q

lasing modes corresponding to lower L_z which therefore have higher gain and capture more of the pump energy. Since more of the pump energy goes into these highly anisotropic modes whose emission pattern differs strongly between the prolate and oblate shapes, these shapes show essentially the intensity profiles predicted by the ray-optics model for the intermediate values of L_z corresponding to slow classical escape and therefore low but above-threshold Q values.

To produce a numerical simulation of the cumulative directionality of all the lasing modes with classical escape, we consider ensembles of starting conditions homogeneously distributed in ϕ as in section 10.2. If the average Q of an ensemble starting at some $\sin \chi_0$ is above the lasing threshold, we record the resulting escape directionality; otherwise we discard that particular ray bundle. Repeating this for uniformly spaced $\sin \chi_0 \in [\sin \chi_c, 1]$ and $L_z/\rho(0) \in [0, 1]$, one obtains Fig. 12.3(c). It can be seen that maximum emission occurs near the equator for the oblate shape, and around 30° away from the poles in the prolate droplet. The emission peak of the oblate droplet is split because of the island shown in Fig. 12.3(b), but this structure is not seen in Fig. 12.1. Nevertheless, the ray theory is clearly able to account for the overall location of the emission maxima seen experimentally. The discrepancies in fine structure may be resolvable by modifying the parametrization in Eq. (12.1) which we chose to model the shape. More experiments are required to check if the above splitting does indeed occur in droplets at different oscillation stages.

Chapter 13

Conclusion

With the study of asymmetric resonant cavities, a fruitful extension of ideas from nonlinear dynamics and quantum chaos has been initiated, which holds immediate promise for real applications. To show that resonances in nonintegrable systems show the properties we predicted, exact numerical calculations were performed. Since these calculations are, however, themselves idealizations, it is highly desirable to test their validity in experiments.

The relative simplicity of our basic assumptions in the pseudoclassical model may be viewed as a shortcoming from the point of view of theoretical rigor, but efforts are under way to derive the results of the pseudoclassical model more rigorously. Moreover, the insights and ideas gained with the present approach have accelerated progress in the numerical work, as well in experimental activity. Since the latter two have shown encouraging agreement with our predictions, it will indeed be worth while to put the model on stronger foundations.

The main corrections to the pseudoclassical model are believed to be dynamical localization and chaos-assisted tunneling, as explained. The evidence for both effects is still rather indirect and comes solely from the numerical calculations. A study of chaos-assisted tunneling in the open annular billiard is a promising avenue for making analytical progress, due to the availability of an exact expression for the quantum map connecting successive boundary reflections. This type of expression has not been found for the general convex billiard, and it is likely that only approximate results can be obtained. The great advantage in the annular billiard is that both the inner and outer walls are perfect circles, so that in particular the value of $\sin \chi$ in the SOS (as defined for billiards above) is the exact classical analog of angular momentum with respect to the origin of the outer circle. In the deformed billiard, $\sin \chi$ is the tangential component of momentum at the boundary, which itself however changes orientation with respect to the radius vector to the billiard origin.

The latter is also a problem that one faces when trying to quantize the effective map derived here. If this could be done, we would be able to make analytical statements about the localization length in a general convex billiard. One possibility is to neglect the above problem and treat $\sin \chi$ as angular momentum. Then the effective map is still far more complex than the standard map, due to the implicit nature of the momentum mapping equation and the p -dependent kick strength appearing in it.

A very direct way of looking for localization is by using techniques based on the Wigner or Husimi function. Much preliminary work in this respect has been done by the author but was not included here due to its unfinished nature.

The assumption that the internal dynamics of the open billiard follows that of the closed system is not valid when applied to the low $\sin \chi$ resonances which are so short lived that they do not stay in the billiard long enough to follow an adiabatic curve for one period. Therefore it will be of interest to compare wave functions and Wigner functions of closed and open billiards to see how they differ in the admixture of low angular momentum components.

In the context of chaos-assisted tunneling, one might expect not only enhanced widths of the levels still confined to KAM tori at intermediate deformations, but in addition the enhanced tunnel splitting of symmetric and antisymmetric states. These states exist in our convex billiards for the same reasons as in the annular billiard, but there is preliminary evidence that their splitting is not enhanced while their width is. This is likely due to the fact that chaotic diffusion from one edge of the chaotic sea to the other, as required for the tunnel splitting, is preempted by the ray escape that occurs rapidly after entering the lower- $\sin \chi$ region of the SOS. This escape-mediated depletion of the processes required for enhanced level splitting deserves further study, as it further sets apart the quantum-chaology of the open system from that of the closed one.

Although we have managed to make contact with the experiments on lasing droplets on a qualitative basis, further investigation into the semiclassical theory for asymmetric spheroids, as well as the laser theory for these multimode systems, are required. Also, we have not performed any numerical calculations of the quasibound states. The additional complication arising here is that one can no longer choose the polarization such that it remains conserved, as soon as the shape becomes non-spherical. This means that in principle the wave function matching calculations have to include field components with two orthogonal polarizations with the associated matching conditions, thus doubling the number of unknowns. This can be done, but may not be necessary to capture the essential physics, which is still expected to result from the competition between classical ray diffusion, direct and chaos-assisted tunneling, and localization. These phenomena should be present in a similar way in the scalar wave equation, as is the case for the cylinders discussed mainly in this thesis.

There are also some interesting problems of classical mechanics that are worth pursuing. Having achieved satisfactory correspondence between the effective map and the real billiards, the former may find an application in determining the escape directionality knowing only the curvature. This could be done in combination with a refinement of the ergodic model, thus leading to analytical expressions that are valid for the case of dynamical eclipsing, too. It was also found by the author that if one plots the emission direction as a function of the precise starting value of ϕ on the adiabatic curve, a fractal function with intermediate continuous intervals is obtained. The continuous segments can be identified as being due to short escaping orbits that undergo almost no diffusion before leaving the billiard. Since their number increases with deformation, the said regular segments become larger and more numerous. It is easy to show that this leads to increased fine structure in the classical directionality histograms. This may be a feature that could be resolved in experiments at large wavenumbers so far inaccessible to our numerics.

First steps were taken during this thesis to understand the properties of nonintegrable open cavities that cannot be reduced to a two-dimensional problem. The system that was studied within the ray-optics framework consists of a dielectric cylinder of finite length with plane faces, one of which is tilted away from the normal to the cylinder axis. This tilted cap is an example for a truly three-dimensional

system undergoing a transition to chaos with tilt angle of the cap. It may provide a suitable entry point into the study of such systems because of its strong similarities to the problem of a charged particle in a quantum well subjected to a tilted magnetic field¹⁴⁸.

Overall, we have seen that the classical picture is able to make a number of predictions about the positions, widths and emission patterns of WG modes in ARCs, which are found to be in agreement with exact numerics as well as experiments. This is of great value in that it establishes the concepts that seem to capture the essential physics of these systems which are so difficult to treat from the wave point of view. It is to be hoped that this work will stimulate further developments both in the theory of semiclassics in mixed open systems and in microcavity optics.

Appendix A

Einstein-Brillouin-Keller quantization in closed systems

In this appendix, we give supplemental material concerning the relation between the EBK method and classical mechanics.

A.1 WKB approximation for eigenstates on a torus

The general WKB ansatz for a quantum wavefunction is of the form

$$\psi(\mathbf{r}) = A(\mathbf{r}) e^{iS(\mathbf{r})/\hbar}, \quad (\text{A.1})$$

where we use $S \equiv kS$ in this appendix for brevity. This can be inserted into the time-independent Schrödinger equation,

$$-\frac{\hbar^2}{2m} \nabla^2 \psi + V(\mathbf{r}) \psi = E \psi. \quad (\text{A.2})$$

If one neglects all terms of order \hbar or higher, one obtains a classical equation which does not contain \hbar :

$$\frac{1}{2m} (\nabla S(\mathbf{r}))^2 + V(\mathbf{r}) = E. \quad (\text{A.3})$$

But this is just the (time-independent) Hamilton-Jacobi equation introduced earlier in Eq. (3.76), with $\nabla S = \mathbf{p}$ because of Eq. (3.70). The phase S thus has the meaning of the classical action function and its gradient points in the direction of classical propagation. If one collects the terms of first order in \hbar , the resulting equation can be written in the form

$$\nabla \cdot (A^2 \nabla S/m) = 0, \quad (\text{A.4})$$

which is the *amplitude transport equation*¹⁰¹. This terminology becomes clear if we consider A^2 as a density and use $\nabla S = \mathbf{p} = m\dot{\mathbf{r}}$. Then this equation means that the probability density current is divergence free. A special solution of this equation is

$$A^2(\mathbf{r}) = \left| \det \left[\frac{\partial^2 S}{\partial \mathbf{r} \partial \mathbf{J}} \right] \right|, \quad (\text{A.5})$$

where \mathbf{J} are the action variables specifying the phase space torus. The proof given by Gutzwiller³⁸ starts with a time dependent formulation of the action, called Hamilton's principal function, as opposed to Hamilton's characteristic function which was introduced earlier. We will give a more direct proof that does not make the detour

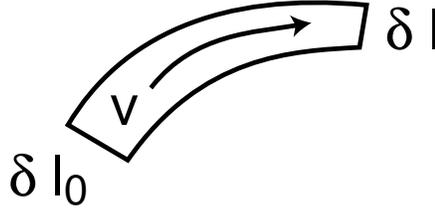


Figure A.1: A bundle of trajectories evolving in time.

via the principal function, but as in section 3.5.1 we will be restricted to two degrees of freedom, since the present derivation was devised only to provide a foundation for the applications in this thesis.

The first step is to rewrite Eq. (A.4) as

$$0 = \nabla \cdot (A^2 \dot{\mathbf{r}}) \quad (\text{A.6})$$

$$= \hat{\mathbf{z}} \cdot \nabla \times \mathbf{Q}, \quad (\text{A.7})$$

$$\mathbf{Q} = \begin{pmatrix} -A^2 \dot{y} \\ A^2 \dot{x} \\ 0 \end{pmatrix}. \quad (\text{A.8})$$

Here, $\hat{\mathbf{z}}$ is the unit vector perpendicular to the xy -plane. This curl expression is now integrated using Stokes' theorem. To define the integration domain, consider two line elements δl_0 , δl intersecting a bundle of trajectories, as shown in Fig. A.1. We keep δl_0 fixed and assume that δl is its image under the time evolution. Stokes' theorem then tells us that the line integral of \mathbf{Q} along δl is the same as that along δl_0 (there is no contribution along the flow lines, because \mathbf{Q} is perpendicular to $\dot{\mathbf{r}}$). The integrand of the line integral is

$$A^2 \begin{pmatrix} -\dot{y} \\ \dot{x} \\ 0 \end{pmatrix} \cdot \begin{pmatrix} dx \\ dy \end{pmatrix} = A^2 \det \left[\begin{pmatrix} \dot{x} & dx \\ \dot{y} & dy \end{pmatrix} \right] = A^2 \det(\dot{\mathbf{r}}, d\mathbf{r}). \quad (\text{A.9})$$

In the last expression, we transform both column vectors to angle coordinates, using

$$(\dot{\mathbf{r}}, d\mathbf{r}) = \left(\frac{\partial \mathbf{r}}{\partial \Phi} \right) (\dot{\Phi}, d\Phi). \quad (\text{A.10})$$

But the angle variables are just linear in time with frequencies $\vec{\omega}$. The fact that Φ is only translated parallel to the frequency vector $\vec{\omega}$ by the time evolution also means that displacements $d\Phi$ remain constant. Therefore, we have

$$\det[(\dot{\mathbf{r}}, d\mathbf{r})] = \det \left(\frac{\partial \mathbf{r}}{\partial \Phi} \right) \det(\vec{\omega}, d\Phi) = \text{const} \times \det \left(\frac{\partial \mathbf{r}}{\partial \Phi} \right). \quad (\text{A.11})$$

Now we shrink the length of δl_0 and the time interval to zero, so that the constancy of the line integral across δl implies the constancy of the integrand A.9 itself. Together with Eq. (A.11), this leads to

$$A^2 = \text{const} \times \det \left(\frac{\partial \Phi}{\partial \mathbf{r}} \right). \quad (\text{A.12})$$

Using the definition of S as the generating function for the transformation between \mathbf{r} and Φ , Eq. (3.71), one finally arrives at the solution Eq. (A.5), where the absolute value is taken to insure real A .

One can understand the meaning of A^2 in the following way: Assume we have distributed trajectories with a constant areal density σ on the torus, in the sense that using the angle variables, $\sigma d\Phi_1 d\Phi_2$ is the probability of finding trajectories in the surface element $d\Phi_1 d\Phi_2$ on the 2D torus. Then the motion just induces linear shifts in the coordinates Φ_i , and thus leaves the density constant, and in particular divergence-free. But in the configuration space \mathbf{r} , the density is obtained from the projection of the distribution,

$$A^2(\mathbf{r})d\mathbf{r} = \sigma \left| \det\left[\frac{\partial\Phi}{\partial\mathbf{r}}\right] \right| d\mathbf{r}, \quad (\text{A.13})$$

which reduces to Eq. (A.5) as we just showed (if we set $\sigma = 1$). This interpretation of A^2 as the classical probability density of a democratic ensemble on the torus means that A^2 will diverge at the caustics, cf. the remark on caustics at the end of subsection 3.7.1. Therefore, the WKB solutions break down near caustics.

The quantization conditions follow from the requirement that the WKB wavefunction be *single-valued* while satisfying the boundary conditions. In the billiard, one usually assumes Dirichlet boundary conditions, i.e. vanishing of the wavefunction at the hard walls. Other conditions will be admitted later, but it is clear that we cannot satisfy them in general with the ansatz of Eq. (A.1). The solution is to take a superposition of two (or more) terms of the WKB form, which can then achieve the necessary cancellations at the boundary by interference. The question is, how many such different terms are there at our disposal? From the discussion of section 3.7.2, we recall that in the integrable case we are considering here, there are two sheets of \mathbf{S} for a convex planar billiard, corresponding to the two possible directions of a trajectory through a given point that are not equivalent under time reversal. Using the notation of that section, we can write the ansatz⁹⁹

$$\psi(\mathbf{r}) = A_{BC}(\mathbf{r}) e^{iS_{BC}(\mathbf{r})/\hbar} + A_{CB}(\mathbf{r}) e^{iS_{CB}(\mathbf{r})/\hbar}. \quad (\text{A.14})$$

We know from the classical reflection law that

$$\nabla S_{BC} \cdot \mathbf{n} = -\nabla S_{CB} \cdot \mathbf{n} \quad (\text{A.15})$$

along the boundary, with \mathbf{n} being the normal vector. Also, we must have $S_{BC} = S_{CB}$ at the boundary because the boundary condition must be satisfied for a whole set of k which appears in the exponentials in ψ . These facts in conjunction with the boundary condition fix the relationship between A_{BC} and A_{CB} at the boundary, e.g. one has⁹⁹ $A_{BC} = -A_{CB}$ for Dirichlet and $A_{BC} = A_{CB}$ for Neumann boundary conditions (vanishing normal derivative). More generally, we write for the phase shift

$$A_{CB} = e^{-i\alpha} A_{BC}, \quad \alpha \in [0 \dots \pi]. \quad (\text{A.16})$$

At the caustics where the WKB ansatz becomes invalid, one has to derive connection formulas between solutions belonging to different sheets of \mathbf{S} . For general multidimensional systems, this was done by Maslov¹⁰². The result is analogous to the one-dimensional case, in that each crossing of a classical turning point in an effective potential that can be locally expanded with a linear term gives rise to a phase shift of $-\pi/2$. This is the shift of the outgoing wave relative to the incoming one, corresponding to a retardation.

The phase shifts due to the caustics and to the boundary have to be taken into account when we ask whether the wave function is single-valued. The meaning

of single-valuedness has been explained for billiards in section 7.1. Each sheet of S , i.e. S_{BC} as well as S_{CB} , is itself again multivalued, so that the value of S_{BC} (e.g.) at a given point can reflect the above phase advance or not. The possible values of S_{BC} at a point \mathbf{r} are given by the phase changes around all possible paths that can be patched together by following ray segments or moving along lines of constant phase, and computing the change in S as the integral of its gradient,

$$\Delta S/\hbar = \frac{1}{\hbar} \oint_{\Gamma} \nabla S(\mathbf{r}) \cdot d\mathbf{r} - \alpha - \mu \frac{\pi}{2}. \quad (\text{A.17})$$

The value of ΔS must be an integer multiple of $2\pi\hbar$ in order for each exponential in Eq. (A.1) to remain single valued despite the multivaluedness of S_{BC} and S_{CB} . We know that there are two fundamental inequivalent loops C_1 and C_2 , and they thus give rise to two quantization conditions for the actions, as written in Eq. (7.3).

Keller and Rubinow⁹⁹ explicitly used this method to obtain semiclassical expressions for both bouncing-ball and whispering-gallery orbits in the ellipse, obtaining good numerical agreement with exact calculations in both cases. The applicability of the EBK conditions to whispering-gallery orbits must be stressed here because we rely on it later on. Moreover, as we are about to find out, the Berry-Tabor formula (which is based on a periodic-orbit sum) *fails* for whispering gallery modes. Keller and Rubinow employed further approximations beyond the WKB ansatz to derive expressions for bouncing-ball and whispering-gallery orbits in the *general* convex billiard. In the whispering-gallery case they find for the quantized wave numbers with Dirichlet boundary conditions

$$k \approx \frac{2\pi m}{L_B} + \left(\frac{2\pi m}{L_B}\right)^{1/3} \left[\pi \left(n + \frac{3}{4}\right)\right]^{2/3} \frac{1}{L_B} \int_0^{L_B} \kappa^{2/3}(s) ds. \quad (\text{A.18})$$

Here, s is the arc length along the boundary of total length L_B , and κ is the curvature. The first term simply means that an eigenstate exists whenever the wavelength $2\pi/k$ fits into the circumference L_B an integer number of times m . The second term depends on the “radial” quantum number n . The formula becomes worse as n increases, but this is only a consequence of approximations made to estimate the actions on one hand, and further approximations to then solve the semiclassically exact implicit relations determining k . We shall derive later an expression based on the adiabatic curves which we leave in its implicit form and solve for k numerically. This is also how Keller and Rubinow obtained good results for the ellipse. First, we shall illustrate the method for a simple example which nonetheless exhibits all the physics we are going to encounter later.

A.2 Semiclassical quantization in the circle

This is the simplest example studied by Keller and Rubinow. We have obtained the actions in Eqs. (3.83) and (3.92). They are quantized for Dirichlet boundary conditions according to

$$J_\phi = 2\pi\hbar m, \quad (\text{A.19})$$

$$J_r = 2\pi\hbar \left(n + \frac{3}{4}\right). \quad (\text{A.20})$$

For von-Neumann boundary conditions, $3/4$ is replaced by $1/4$. Here m and n are the angular momentum and radial quantum numbers, and the phase in the radial

condition arises because the phase space contour in Fig. 3.2 (b) crosses the caustic once and goes through one reflection; the angular momentum quantization has no additional phases because the contour in Fig. 3.2 (a) remains on a single sheet of \mathcal{S} .

Now we insert the quantization condition for J_ϕ into that for J_r and arrive at

$$m \left[\sqrt{\frac{R^2}{r_0^2} - 1} + \arcsin \frac{r_0}{R} - \frac{\pi}{2} \right] = \pi \left(n + \frac{3}{4} \right). \quad (\text{A.21})$$

In the limit that the radius r_0 of the caustic becomes very small, $r_0 \ll R$, we can expand the square root and the arcsin, leading to

$$m \left[\frac{R}{r_0} - \frac{\pi}{2} \right] \approx \pi \left(n + \frac{3}{4} \right). \quad (\text{A.22})$$

Here, we use the definition of r_0 in Eq. (3.82), and furthermore convert from the energy E to the wavenumber k ,

$$r_0 = \frac{p_\phi}{\sqrt{2mE}} = \frac{m}{k}. \quad (\text{A.23})$$

Note that this is equivalent Eq. (7.12), with k is quantized according to Eq. (A.21). Substitution of r_0 into Eq. (A.22) leads to the approximation

$$kR \approx \pi \left(n + \frac{3}{4} + \frac{m}{2} \right). \quad (\text{A.24})$$

In the whispering-gallery limit, we can make use of the fact that the turning point r_0 approaches R , and write

$$\frac{r_0}{R} = \sin \chi = 1 - \eta. \quad (\text{A.25})$$

However, an expansion of the arcsin in this case is only possible in the non-analytic form

$$\arcsin(1 - \eta) \approx \frac{\pi}{2} - \sqrt{2\eta - \eta^2} \dots \quad (\text{A.26})$$

This causes the analogue of Eq. (A.18) in the circle to be a bad approximation for $n > 0$. Therefore, it is better to solve Eq. (A.21) numerically, which leads to good results. This is the approach we will follow from now on.

It should be stressed that the system of equations (7.3) not only fixes k to some quantized value, but also determines the other unknowns hidden in the action variables. In our two-dimensional case, there is only one more unknown, namely r_0 . It is the parameter that determines the classical actions, and since $r_0 = m/k$ it becomes quantized itself. Clearly, we could use $\sin \chi = r_0/R$ as the classical parameter, and the quantization will then provide $\sin \chi$ as a function of the quantum numbers. This connection between classical trajectories and a particular quantum state is of crucial importance in the subsequent analysis of phase space diffusion and its effects on resonance lifetimes.

A.3 Eikonal theory

For later applications we rewrite the WKB theory in the language of the optical Helmholtz equation,

$$\nabla^2 \psi + k^2 \epsilon(\mathbf{r}) \psi = 0, \quad (\text{A.27})$$

where ϵ is the dielectric constant. This becomes analogous to the Schrödinger equation if we set $\epsilon \equiv 2m(E - V)$ and $k = 1/\hbar$. Consequently, the WKB ansatz is

$$\psi(\mathbf{r}) = A(\mathbf{r}) e^{ikS(\mathbf{r})}, \quad (\text{A.28})$$

which implies that S now has the dimension of a length instead of an action. In this form, the WKB ansatz is known as the eikonal approximation of geometric optics¹²⁶, and ∇S now points in the direction of the rays. The analogy to quantum mechanics becomes a full equivalence only if we consider position-independent potentials V and dielectric constants ϵ , respectively. We will return to this issue later. For the study of billiards, we can restrict ourselves to precisely this special case. In particular, we can set $\epsilon = 1$ so that

$$(\nabla S)^2 = 1, \quad (\text{A.29})$$

i.e., ∇S is a unit vector in the ray direction. The quantization condition now reads

$$k J_i = 2\pi + \alpha_i + \mu_i \frac{\pi}{2}. \quad (\text{A.30})$$

Another way of stating the special conditions above is that in the plane hard-wall billiard, the quantum mechanical problem does not have \hbar and energy E as independent parameters. This is also called the scaling property¹⁰⁶.

Appendix B

Periodic-orbit theory for the density of states

From the fact that the actions of an integrable system are quantized, one should not conclude that the resulting tori parametrized by the J_i have rational winding number. To see this, note that we could vary the ratio J_i/J_j infinitesimally by changing α_i in the quantization condition, Eq. (7.3), thereby producing a small change in the winding number, which would in general lead to irrational w , corresponding to orbits that do not close on themselves. In the circle, this implies that it is incorrect to quantize the system in the naive way, requiring a ray path to close on itself and to have a length that is an integer multiple of a wavelength. For example, if we assume a trajectory with winding number $w = 1/3$ can be quantized, the resulting angle of incidence according to Eq. (3.107) is $\chi = \pi/6$, which can be inserted into Eq. (3.94), using

$$\sqrt{\frac{\beta}{J_\phi^2} - 1} = \sqrt{\frac{1}{\sin^2 \chi} - 1} = \sqrt{3}. \quad (\text{B.1})$$

Now we apply the quantization conditions, Eqs. (A.19) and (A.20), to obtain for the radial quantum number the condition

$$n = \pi \left(\frac{3m}{9 - \pi} - \frac{3}{4} \right). \quad (\text{B.2})$$

The righthand side cannot be an integer due to the fact that π is not a divisor of any integer.

Despite this remark, periodic orbits nonetheless do play an eminent role in all semiclassical theories that calculate the *density of states*^{38,107–109}.

This distinction between the EBK method for a single state and the density-of-states approaches (that periodic orbits have no special meaning in the former but are the only trajectories appearing in the latter) can be understood if we keep in mind that the integers in the quantization conditions are not the same as the integers characterizing a rational winding number. The relationship between the two is established via the Poisson summation formula,¹³⁷

$$\sum_{M=-\infty}^{\infty} e^{2\pi i Mx} = \sum_{n=0}^{\infty} \delta(x - n), \quad (\text{B.3})$$

which can easily be proved by writing the delta functions as a Fourier integral, performing the n -summation under the integral to get a geometric series, and applying the residue theorem to the latter.

After we showed how this formula connects winding numbers and quantum numbers, we will briefly explore the situation in chaotic systems where tori no longer exist but the density of states can still be expressed in terms of sums over periodic orbits.

B.1 Berry-Tabor formula for integrable systems

Formula B.3 is used in the density of states, which we can write with the Hamiltonian as a function of the actions, $H(\mathbf{J})$, as

$$\rho(E) = \sum_{\mathbf{n}} \delta(E - H(\hbar(\mathbf{n} + \beta))). \quad (\text{B.4})$$

Here, we lumped all phases in the quantization conditions (Maslov and boundary) into the vector β . This can also be written as an integral over the continuous action variables (N is the number of degrees of freedom),

$$\rho(E) = \int d\mathbf{J} \delta(E - H(\mathbf{J})) \sum_{\{n_i\}} \delta^N(\mathbf{J} - 2\pi \hbar(\mathbf{n} + \beta)). \quad (\text{B.5})$$

Now we use Eq. (B.3) to obtain

$$\rho(E) = \sum_{\mathbf{M}} e^{2\pi i \mathbf{M} \cdot \beta} \rho_{\mathbf{M}}(E), \quad (\text{B.6})$$

where we defined

$$\rho_{\mathbf{M}}(E) \equiv \left(\frac{2\pi}{\hbar}\right)^N \int d\mathbf{J} \delta(E - H(\mathbf{J})) e^{i \mathbf{M} \cdot \mathbf{J} / \hbar}. \quad (\text{B.7})$$

Since the conserved quantity $H = E$ is a function of the N actions J_i , we could equivalently parametrize the torus by only $N - 1$ of the J_i and use E as the N -th variable. This set can be orthogonalized while preserving E as the N -th variable, yielding a transformation of variables from \mathbf{J} to a set ξ . Here, $\xi_N = H$, whereas the other ξ_i do not appear in H . The Jacobian of this transformation is

$$\det\left[\frac{\partial \xi}{\partial \mathbf{J}}\right] = \det\left[\frac{\partial \xi_1}{\partial \mathbf{J}}, \frac{\partial \xi_2}{\partial \mathbf{J}}, \dots, \frac{\partial \xi_N}{\partial \mathbf{J}}\right]. \quad (\text{B.8})$$

From Hamilton's equation, Eq. (3.73), the N -th column is just the vector of angular frequencies

$$\frac{\partial \xi_N}{\partial \mathbf{J}} = \frac{\partial H}{\partial \mathbf{J}} = \vec{\omega}. \quad (\text{B.9})$$

Pulling the magnitude of this vector out of the determinant, we are left with the determinant of an orthogonal matrix which is unity. If we wish to apply this transformation in the integral in Eq. (B.7), we need the inverse of the above Jacobian, thus obtaining

$$\rho_{\mathbf{M}}(E) = \left(\frac{2\pi}{\hbar}\right)^N \int d\xi \frac{1}{|\vec{\omega}|} \delta(E - \xi_N) e^{i \mathbf{M} \cdot \mathbf{J}(\xi) / \hbar} \quad (\text{B.10})$$

$$= \left(\frac{2\pi}{\hbar}\right)^N \int d\xi_1 d\xi_2 \dots d\xi_{N-1} \frac{1}{|\vec{\omega}|} e^{i \mathbf{M} \cdot \mathbf{J}(\xi) / \hbar}, \quad (\text{B.11})$$

where the ξ_N -integral has been performed with the effect of setting $\xi_N = E$. Since we are considering the semiclassical limit in which \hbar can be considered small, it is

consistent with the previous expansion in powers of \hbar to evaluate the integral in Eq. (B.11) in the stationary phase approximation. The contribution arises when the ξ_i ($i < N$) are chosen so as to yield

$$\mathbf{M} \cdot \frac{\partial}{\partial \xi_i} \mathbf{J}(\xi) = 0. \quad (\text{B.12})$$

The vectors $\partial \mathbf{J} / \partial \xi_i$ ($i < N$) appear as rows in the inverse of the matrix in Eq. (B.8), so they are all orthogonal to the N -th row of that matrix, which is $\partial \mathbf{J} / \partial E$. We conclude from Eq. (B.12) that \mathbf{M} must be *parallel* to $\partial \mathbf{J} / \partial E$. But that means that any vector pointing in the direction of $\partial \mathbf{J} / \partial E$ must have rationally related components, because \mathbf{M} is a vector of integers.

This holds in particular for the angular frequency vector, because it satisfies the relation

$$0 = \frac{\partial H}{\partial \xi_i} = \frac{\partial H}{\partial \mathbf{J}} \frac{\partial \mathbf{J}}{\partial \xi_i} = \vec{\omega} \cdot \frac{\partial \mathbf{J}}{\partial \xi_i}, \quad (\text{B.13})$$

for $i < N$, where the first equality arises from the fact H depends only on ξ_N . As a consequence, the components of $\vec{\omega}$ satisfy

$$\omega_1 : \omega_2 \dots : \omega_N = M_1 : M_2 \dots : M_N. \quad (\text{B.14})$$

The stationary phase approximation has therefore selected precisely the tori with rational winding numbers, i.e. those that support periodic orbits. This means that the exponential in Eq. (B.11) can also be interpreted as the action along the periodic orbit where loop i is rounded M_i times:

$$e^{i \mathbf{M} \cdot \mathbf{J}(\xi) / \hbar} \equiv e^{i S(\mathbf{M})}. \quad (\text{B.15})$$

To obtain the density of states, the stationary contributions must be summed to obtain $\rho_{\mathbf{M}}(E)$, which is then inserted into Eq. (B.6). One arrives at the *Berry-Tabor formula*,

$$\begin{aligned} \rho(E) &= \rho_0(E) + \tilde{\rho}(E), \\ \tilde{\rho} &= \frac{2}{h^{(N+1)/2}} \sum_{\vec{\mu} > \mathbf{0}} \frac{|\vec{\mu}|^{(N-1)/2}}{|\vec{\omega}(\mathbf{J}) \mathbf{K}(\mathbf{J})|} \sum_{q=1}^{\infty} q^{(N-1)/2} \cos \left[\frac{q S(\vec{\mu})}{h} - \gamma(q, \vec{\mu}) \right]. \end{aligned} \quad (\text{B.16})$$

Here, we have decomposed the integer vectors

$$\mathbf{M} = q \vec{\mu}, \quad (\text{B.17})$$

where the elements of $\vec{\mu}$ are also integers that in addition have no common integer factor. The case $\vec{\mu} = \mathbf{0}$ gives rise to ρ_0 , which is the weakly E -dependent mean level density, semiclassically obtained by dividing the phase space volume of the energy shell by h^N . The remaining sum over $\vec{\mu}$ contained in $\tilde{\rho}$ reproduces the discrete level structure in the form of a superposition of oscillatory terms. $S(\vec{\mu})$ is the action along such a primitive periodic orbit, and q counts its repetitions. \mathbf{K} is the curvature of the energy surface in action space at the point $\mathbf{J}(\vec{\mu})$, and γ contains the phase shifts β and additional phases arising in the stationary phase approximation.

The first remark about this formula is that the ensuing periodic-orbit sum can be very slowly convergent in reproducing the series of delta-function peaks comprising the semiclassical density of states¹⁰⁹. The same information is obtained with full semiclassical accuracy from the EBK method, however restricted to one level at a time.

The slow convergence is understandable if we ask ourselves how the periodic orbits manage to reproduce peaks at the correct energies even though these correspond in general to irrational tori. The reason is the constructive interference of periodic orbits with winding numbers closer and closer to the irrational one, which approximate the EBK quantization conditions¹¹⁰. To achieve a sufficiently strong weight at the EBK energy level, many of such “rational approximants” have to be taken into account. But the better a rational orbit approximates an irrational winding number, the longer it gets. This is the origin of the convergence difficulties.

This trace formula for the spectral density is still useful even if the periodic-orbit sum is not carried to large accuracy, provided one has a physical reason for disposing of trajectories whose length exceeds some limit, i.e. in microstructures given by the phase coherence length. One should also note that $\rho(E)$ is always used in the sense of a distribution⁶⁵, i.e. under an integral with some function $A(E)$ whose spectral average is to be determined. The number of periodic orbits required then depends on the function $A(E)$.

Equation (B.16) was rederived from a scattering approach in Ref.⁶⁵ and tested for the circular billiard. The comparison with the exact spectrum is best done for the Fourier transform of $\rho(k)$, where $k = \sqrt{2mE}/\hbar$. This gives a *length spectrum* $\rho(l)$, and excellent agreement between the Berry-Tabor formula and the exact result was obtained. A notable exception are the whispering gallery orbits, which in the true length spectrum show up as peaks near integer multiples of the billiard circumference. The problem arises in the stationary phase approximation for these orbits, and we will encounter related difficulties later when the exact wavefunctions of the circle are used directly as a starting point for asymptotic approximations. The stationary phase approximation breaks down because the whispering-gallery orbits are never far away from the caustic; but at the caustic, many trajectories come close to being parallel to each other. Under such circumstances, the stationary points (B.12) coalesce. One can interpret this to mean that the curvature of the energy shell vanishes, $K(\mathbf{J}) = 0$, and causes the stationary-phase result in Eq. (B.16) to become undefined. Of course, the whispering-gallery modes are precisely the ones of greatest interest in our optics applications, so periodic-orbit theory has little appeal in this context.

We shall not dwell on this issue here for the simple reason that the EBK quantization conditions *do not* fail for whispering gallery orbits, as evidenced by the work of Keller and Rubinow described above. The WKB ansatz is also the foundation of a very general theory of semiclassical quantization discussed, e.g., by Shnirelman and Lazutkin⁶¹.

B.2 Gutzwiller trace formula for chaotic systems

Similar (or worse) convergence problems arise in the periodic-orbit sums for chaotic systems. We will omit a detailed discussion of this subject because it is not used in what follows; extensive treatments from various viewpoints can be found in Refs.^{65,38,107,108}. Gutzwiller’s trace formula for the density of states is

$$\rho(E) = \rho_0(E) + \frac{1}{\pi \hbar} \sum_p \sum_{q=1}^{\infty} T_p A_{q,p} \cos \left[\frac{q S_p}{\hbar} - \gamma(q, p) \right], \quad (\text{B.18})$$

where the mean density ρ_0 is a smooth function of energy, as in the Berry-Tabor formula. The form is quite similar to Eq. (B.16) in that p labels the periodic orbits

and q their repetitions, while S_p is the action (integrated along the periodic orbit) and γ some phase shift as before. The power of \hbar in the prefactor is different, however, and so are the pre-exponential factors; T_p is the period and $A_{q,p}$ the so-called stability amplitude of the periodic orbits, which have to be isolated in this formulation. For $N = 2$ degrees of freedom, an integrable system exhibits an $\hbar^{-3/2}$ prefactor in the density oscillations, while the contributions from periodic orbits in the chaotic system go as \hbar^{-1} . The latter is a smaller correction due to the smallness of Planck's constant.

The Gutzwiller formula is still the most versatile method even for the determination of individual quantized states, because its applicability is not restricted to a certain class of chaotic systems, as is our approach to be described shortly. The way in which one can hope to extract single-level information from Eq. (B.18) with only a finite number of periodic orbits is based on the following observation.

It is not necessary to reproduce the full density of states down to delta-function accuracy if we are looking for the positions of eigenvalues. As for a scattering resonance of the Breit-Wigner shape, where the exact resonance position can simply be read off from the peak location, we expect to find the semiclassical eigenvalues from the broadened approximates to the delta function density by simply noting the positions of the maxima. All we have to require then is that these broadened maxima are isolated from each other. This is a (as yet unproven) conjecture which was used in Refs.^{111–113}. It is implemented by using the *spectral staircase function* $N(E)$, which counts the number of levels up to the given energy E . It is the integral of the density of states, and it can therefore be approximated by taking as the integrand Eq. (B.18) with only a finite number of periodic orbits up to some classical period T^* . If we call this approximate staircase function $N_{T^*}(E)$, then it will have smoothed steps of unit height centered around those energies E_n where the broadened peaks in the density of states appear. The center point of the steps corresponds to a step height of $1/2$, so that we obtain the condition

$$N_{T^*}(E_n) = n + 1/2, \quad (\text{B.19})$$

where n is an integer. Of course E_n depends on T^* but should approach the correct semiclassical eigenvalue as this classical period is taken to infinity. The minimum value of T^* necessary to guarantee isolated peaks in the approximate density of states is given by the *Heisenberg time* T_H .

This time can be obtained by requiring Eq. (B.18) to have oscillations in energy on the scale of the mean level spacing $\Delta E = 1/\rho_0$, in order to produce peaks that are separated from each other. Now the argument of the cosine in Eq. (B.18) has an energy period given by

$$\Delta E = \frac{2\pi\hbar}{\partial S/\partial E}. \quad (\text{B.20})$$

The derivative of the action appearing here can be calculated for an orbit of period T as

$$\partial S/\partial E = \frac{\partial}{\partial E} \oint \mathbf{p} d\mathbf{q} = \oint \frac{\partial \mathbf{p}}{\partial E} d\mathbf{q} = \oint \left[\frac{\partial H}{\partial \mathbf{p}} \right]^{-1} d\mathbf{q} \quad (\text{B.21})$$

$$= \oint \frac{dt}{d\mathbf{q}} d\mathbf{q} = \oint dt = T. \quad (\text{B.22})$$

Equating the two expressions for ΔE , we then have for the classical period that is needed

$$T \equiv T_H = 2\pi\hbar \rho_0(E) \propto \hbar^{(-N+1)}. \quad (\text{B.23})$$

This Heisenberg time may not be prohibitively long if the effective h is sufficiently large. In any case, the need to consider potentially large numbers of periodic orbits leads to the question of whether an alternative approach yields useful results in a simpler way, especially taking into account the fact that we are interested in a region of phase space not far from the chaos border.

Bibliography

- [1] C. Weisbuch and B. Vinter, *Quantum Semiconductor Structures*, (Academic Press, Boston, 1991)
- [2] C. M. Marcus *et al.*, Phys. Rev. Lett. **69**, 506 (1992)
- [3] T. M. Fromhold *et al.*, Phys. Rev. Lett. **72**, 2608 (1994)
- [4] R. A. Jalabert, H. V. Baranger and A. D. Stone, Phys. Rev. Lett. **65**, 26442 (1990)
- [5] *Proc. 1994 Les Houches Summer School on Mesoscopic Quantum Physics*, E. Akkermans, G. Montambaux, J. L. Pichard and J.Zinn-Justin, eds. (Elsevier, Amsterdam 1995)
- [6] E. S. C. Ching, P. T. Leung and K. Young, in *Optical Processes in Microcavities*, R. K. Chang and A. J. Campillo, eds., (World Scientific, Singapore, 1996)
- [7] Md. M. Mazumder *et al.*, in *Optical processes in Microcavities*, R. K. Chang and A. J. Campillo, eds. (World Scientific, Singapore, 1996)
- [8] S. X. Qian *et al.*, Science **231**, 486 (1986)
- [9] A. Mekis, J. U. Nöckel, G. Chen, A. D. Stone and R. K. Chang, Phys. Rev. Lett. **75**, 2682 (1995)
- [10] J. U. Nöckel and A. D. Stone, in: *Optical Processes in Microcavities*, edited by R. K. Chang and A. J. Campillo (World Scientific, Singapore, 1996).
- [11] J. U. Nöckel, Phys. Rev. B **45**, 14225 (1992)
- [12] J. U. Nöckel, Phys. Rev. B **46**, 15348 (1992)
- [13] J. U. Nöckel, A. D. Stone and H. U. Baranger, Phys. Rev. B **48**, 17569 (1993)
- [14] J. U. Nöckel and A. D. Stone, Phys. Rev. B **50**, 17415 (1994).
- [15] J. U. Nöckel and A. D. Stone, Phys. Rev. B **51**, 17219 (1994).
- [16] H. M. Tseng *et al.*, Opt. Lett. **9**, 499 (1984)
- [17] S. L. McCall *et al.* , Appl. Phys. Lett. **60**, 289 (1992)
- [18] A. Serpengüzel, S. Arnold and G. Griffel, Opt. Lett. **20**, 654 (1995), and references therein.
- [19] M. Kuwata-Gonokami *et al.* , *QELS Technical Digest '95*, Paper QThL5, *Polymer microdisk and microcylinder lasers*; M. Kuwata-Gonokami *et al.*, preprint.

- [20] Y. Yamamoto and R.E. Slusher, *Physics Today* **46**, 66 (1993) and references therein
- [21] L. Collot V. Lefevre-Seguin, M. Raimond and S. Haroche, *Europhys. Lett.* **23**, 327 (1993).
- [22] S. C. Hill and R. K. Chang in *Proceedings of the 5th International Topsøe Summer School on Nonlinear Optics* (Aalborg, Denmark), O. Keller, ed. (Nova Science Publishers, New York, 1992)
- [23] A. F. J. Levi *et al.*, *Electron. Lett.* **28**, 1010 (1992)
- [24] A. J. Campillo, J. D. Eversole and H. B. Lin, in *Optical processes in Microcavities*, R. K. Chang and A. J. Campillo, eds. (World Scientific, Singapore, 1996)
- [25] S. D. Brorson, P. M. W. Skovgaard, in *Optical processes in Microcavities*, R. K. Chang and A. J. Campillo, eds. (World Scientific, Singapore, 1996)
- [26] R.E. Slusher *et al.*, *Appl. Phys. Lett.* **63**, 1310 (1993).
- [27] M. L. Gorodetsky, A. A. Savchenko and V. S. Ilchenko, *Opt. Lett.* **21**, 453 (1996)
- [28] Lord Rayleigh, *The Problem of the Whispering Gallery*, in “Scientific Papers” **5**, 617 (Cambridge University, Cambridge, England, 1912)
- [29] V. F. Lazutkin, *Math. USSR Izvestija* **7**, 185 (1973)
- [30] A. F. Levi *et al.* , *Appl. Phys. Lett.* **62**, 561 (1993)
- [31] P. A. Mello, in: *Proc. 1994 Les Houches Summer School on Mesoscopic Quantum Physics*, E. Akkermans, G. Montambaux, J. L. Pichard and J.Zinn-Justin, eds., 373 (Elsevier, Amsterdam 1995)
- [32] Y. V. Fyodorov, in: *Proc. 1994 Les Houches Summer School on Mesoscopic Quantum Physics*, E. Akkermans, G. Montambaux, J. L. Pichard and J.Zinn-Justin, eds., 373 (Elsevier, Amsterdam 1995)
- [33] M. C. .Gutzwiller, *Physica D* **7**, 341 (1983)
- [34] R. Blümel and U. Smilansky, *Phys. Rev. Lett.* **60**, 477 (1988)
- [35] P. Cvitanović and B. Eckhardt, *Phys. Rev. Lett.* **63**, 823 (1989)
- [36] C. H. Lewenkopf and H. A. Weidenmüller, *Ann. Phys.* **212**, 53 (1991)
- [37] P. Gaspard and S. A. Rice, *J. Chem. Phys.* **90**, 2242 (1988)
- [38] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York 1990)
- [39] B. Georgeot and R. E. Prange, *Phys. Rev. Lett.* **74**, 2851 (1995)
- [40] D. Weiss *et al.*, *Phys. Rev. Lett.* **70**, 4118 (1993)
- [41] M. V. Berry, *J. Phys. A* **12**, 2083 (1977)

- [42] E. Heller, Phys. Rev. Lett. **53**, 1515 (1984)
- [43] G. Casati *et al.*, in *Lecture Notes in Physics* **93**, 334 (1979)
- [44] D. R. Grempel, R. E. Prange and S. Fishman, Phys. Rev. A **29**, 1639 (1984)
- [45] J. Bayfield and P. M. Koch, Phys. Rev. Lett. **33**, 258 (1974)
- [46] R. Blümel, R. E. Prange and U. Smilanski, J. Chem. Phys. **84**, 2604 (1986)
- [47] G. Casati *et al.*, Phys. Rev. A **48**, R1613 (1993)
- [48] B. Li and M. Robnik, J. Phys. A: Math. Gen. **28**, 2799 (1995)
- [49] F. Borgonovi, G. Casati and B. Li, Phys. Rev. Lett. **77**, 4744 (1996)
- [50] K. M. Frahm and D. L. Shepelyansky, Phys. Rev. Lett. **78**, 1440 (1997)
- [51] O. Bohigas, S. Tomsovic and D. Ullmo, Phys. Rep. **223**, 44 (1993)
- [52] F. -M. Dittes, H. L. Harney and A. Müller, Phys. Rev. A **45**, 701 (1992)
- [53] H. Alt *et al.*, Phys. Rev. Lett. **74**, 62 (1995)
- [54] P. Hänggi, R. Utermann and T. Dittrich, Physica B **194-196**, 1013 (1994)
- [55] S. Frischat, unpublished
- [56] G. Hackenbroich and J. U. Nöckel, Europhys. Lett. **39**, 371 (1997).
- [57] J. A. Lock and E. A. Hovenac, J. Opt. Soc. Am. A **8**, 1541 (1991)
- [58] J. U. Nöckel, A. D. Stone, G. Chen, H. Grossman and R. K. Chang, Opt. Lett. **21**, 1609 (1996)
- [59] M. Robnik and M. V. Berry, J. Phys. A: Math. Gen. **18**, 1361 (1985)
- [60] B. Li and M. Robnik, Preprint CAMTP/95-3 July 1995
- [61] V. F. Lazutkin, *KAM Theory and Semiclassical Approximations to Eigenfunctions*, (Springer, New York, 1993)
- [62] L. D. Landau and E. M. Lifshitz, *Theoretical Mechanics* (Pergamon Press, Oxford, 1977)
- [63] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Massachusetts, 1950)
- [64] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1989)
- [65] U. Smilanski, in: *Proc. 1994 Les Houches Summer School on Mesoscopic Quantum Physics*, E. Akkermans, G. Montambaux, J. L. Pichard and J. Zinn-Justin, eds., 373 (Elsevier, Amsterdam 1995)
- [66] L. I. Schiff, *Quantum Mechanics*, McGraw-Hill, Singapore (1968)
- [67] A. M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge University Press, New York, 1988)

- [68] P. J. Richens and M. V. Berry, *Physica* **2D**, 495 (1981)
- [69] Desloge, *Classical Mechanics II*
- [70] S. J. Chang, R. Friedberg, *J. Math. Phys.* **29**, 1537 (1988)
- [71] A. Cayley, *Quart. J. Pure Appl. Math.* **2**, 31 (1958)
- [72] M. V. Berry, *Eur. J. Phys.* **2**, 91 (1981)
- [73] Ya. G. Sinai, *Introduction to Ergodic Theory* (Princeton University Press, Princeton, 1976)
- [74] E. Y. Amiran, 1991, cited as a preprint in: Lazutkin, KAM Theory and semi-classical approximations to eigenfunctions
- [75] F. Brut and R. Arvieu, *J. Phys. A: Math. Gen.* **26**, 4749 (1993)
- [76] E. Ott, *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge (1993)
- [77] G. D. Birkhoff, *Acta Math.* **50**, 359 (1927)
- [78] S. Aubry and P. Y. Le Daeron, *Physica* **8 D**, 381 (1983)
- [79] H. Poincaré, *Les Methodes Nouvelles de la Mechanique Celeste*, Gauthier-Villars, Paris (1892)
- [80] J. Moser, *Stable and Random Motions in Dynamical Systems*, Annals of Mathematical Studies 77, Princeton University Press (1973)
- [81] L. E. Reichl, *The Transition to Chaos*, Springer-Verlag (1992)
- [82] A. J. Lichtenberg and M. J. Lieberman, *Regular and Stochastic Motion* (Springer, New York, 1983)
- [83] S. Smale, *Bull. Am. Math. Soc.* **73**, 747 (1967)
- [84] A. N. Kolmogorov, *Dokl. Akad. Nauk. SSSR* **98**, 527 (1954)
- [85] V. I. Arnol'd, *Soviet Math. Dokl.* **2**, 501 (1961); *Russ. Math. Survey* **18**, 9 (1963)
- [86] J. Moser, *Nachr. Akad. Wiss. Göttingen II, Math. Phys. Kl.*, 1 (1962)
- [87] B. V. Chirikov, *J. Nucl. Energy Part C Plasma Physics* **1**, 253 (1960)
- [88] G. H. Hardy and E. M. Wright, *Introduction to the Theory of Numbers*, Clarendon, Oxford, 1979)
- [89] A. V. Prasad, *J. Lond. Math. Soc.* **23**, 169 (1948)
- [90] J. Greene, *J. Math. Phys.* **20**, 1183 (1979)
- [91] B. V. Chirikov, *Phys. Rep.* **52**, 263 (1979)
- [92] D.F. Escande and F. Doveil, *J. Stat. Phys.* **26**, 257 (1981)
- [93] J. N. Mather, *Ergod. Th. & Dynam. Sys.* **2**, 397 (1982)

- [94] G. Casati, I. Guarneri, D. L. Shepelyansky, *IEEE J. Quant. Electron.* **24**, 1420 (1988)
- [95] I. Dana, N. W. Murray and I. C. Percival, *Phys. Rev. Lett.* **62**, 233 (1989)
- [96] R. S. MacKay, J. D. Meiss, I. C. Percival, *Physica D* /bf 13, 55 (1984).
- [97] D. Bensimon and L. P. Kadanoff, *Physica* **13 D**, 82 (1984)
- [98] B. V. Chirikov, *Phys. Rep.* **52**, 263 (1979)
- [99] J. B. Keller and S. I. Rubinow, *Ann. Phys.* **9**, 24 (1960).
- [100] M. V. Berry and K. E. Mount, *Rep. Prog. Phys.* **35**, 315 (1972)
- [101] R. G. Littlejohn, in *Quantum Chaos*, G. Casati and B. Chirikov, eds., 343 (Cambridge University Press, Cambridge, 1995)
- [102] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximations in Quantum Mechanics* (Reidel, Boston, 1981)
- [103] A. Einstein, *Verh. Deut. Phys. Ges.* (1917)
- [104] L. Brillouin, *J. Phys. Radium* **7**, 353 (1926)
- [105] J. B. Keller, *Ann. Phys.* **4**, 180 (1958)
- [106] M. V. Berry, *Proc. R. Soc. London A* **400**, 229 (1985)
- [107] E. B. Bogomolny, *Nonlinearity*, 805 (1992)
- [108] R. Balian and C. Bloch, *Ann. Phys.* **60**, 401 (1970); *Ann. Phys.* **63**, 592 (1971); *Ann. Phys.* **64**, 271 (1971); *Ann. Phys.* **69**, 76 (1972); *Ann. Phys.* **85**, 514 (1974)
- [109] M. V. Berry and M. Tabor, *Proc. R. Soc. Lond. A* **349**, 101 (1976); *J. Phys. A* **10**, 371 (1972)
- [110] M. S. Child, *Semiclassical Mechanics with Molecular Applications*, 189 (Clarendon Press, Oxford, 1991)
- [111] E. Bogomolny and C. Schmit, *Nonlinearity* **6**, 523 (1993)
- [112] R. Aurich *et al.*, *Phys. Rev. Lett.* **68** 1629 (1992)
- [113] E. B. Bogomolny and J. P. Keating, *Phys. Rev. Lett.* **77**, 1472 (1996)
- [114] B. Eckhardt, G. Hose and E. Pollak, *Phys. Rev. A* **39**, 3776 (1989)
- [115] E. Pollak, *Chem. Phys.* **61**, 305 (1981)
- [116] W. H. Miller, *J. Chem. Phys.* **48**, 1651 (1968)
- [117] O. Bohigas, S. Tomsovic and D. Ullmo, *Phys. Rev. Lett.* **65**, 5 (1990)
- [118] S. Sridhar, *Phys. Rev. Lett.* **67**, 785 (1991)
- [119] J. Stein and H. J. Stöckmann, *Phys. Rev. Lett.* **68**, 2867 (1992)
- [120] H. D. Gräf *et al.*, *Phys. Rev. Lett.* **69**, 1296 (1992)

- [121] C. A. Kruehle *et al.*, *Physica D* **78**, 214 (1994)
- [122] M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation* (Academic, New York, 1969)
- [123] P. W. Barber and S. C. Hill, *Light Scattering by Particles: Computational Methods*, World Scientific, Singapore (1990)
- [124] B. R. Johnson, *J. Opt. Soc. Am.* **10**, 343 (1993)
- [125] I.S.Gradshcheyn,I.M.Ryzhik,Table of Integrals, Series and Products, Academic Press, New York (1980)
- [126] Morse and Feshbach, *Methods of Theoretical Physics I* (McGraw-Hill, New York, 1953) Methods of
- [127] P. C. Waterman, *Proc. IEEE* **53**, 805 (1965)
- [128] S. A. Shelkunoff, *Electromagnetic Waves* (D. von Nostrand, New York, 1943)
- [129] P. Barber and C. Yeh, *Appl. Opt.* **14**, 2864 (1975)
- [130] R. F. Millar, *Electron. Lett.* **5**, 412 (1969)
- [131] D. R. Wilton and R. Mittra, *IEEE Trans. Antennas Propag.* **AP- 20**, 310 (1972)
- [132] I. N. Vekua, *new Methods for Solving Elliptic Equations* (North Holland, Amsterdam, 1967)
- [133] R. Penrose, *Proc. Cambridge Phil. Soc.*, no. 52 (1956)
- [134] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in Fortran* (Cambridge University Press, Cambridge, 1992)
- [135] H. Bruus, unpublished
- [136] H. Bruus and A.D. Stone, *Phys. Rev.* **B50**, 18275 (1994).
- [137] F. Haake, *Quantum Signatures of Chaos*, Springer, Berlin, 1991)
- [138] J. U. Nöckel and A. D. Stone, *Nature* **385**, 45 (1997)
- [139] S. Tomsovic and D. Ullmo, *Phys. Rev. E* **50**, 145 (1995)
- [140] E. Doron and S. D. Frischat, *Phys. Rev. Lett.* **75**, 3661 (1995)
- [141] H. Grabert, P. Olschowski and U. Weiss, *Phys. Rev. B* **36**, 1931 (1987)
- [142] R. C. Brown and R. E. Wyatt, *Phys. Rev. Lett.* **57**, 1 (1986)
- [143] T. Geisel, G. Radons and J. Rubner, *Phys. Rev. Lett.* **65**, 2883 (1986)
- [144] J. U. Nöckel, A. D. Stone and R. K. Chang, *Opt. Lett.* **19**, 1693 (1994).
- [145] Lin *et al.*, *J. Opt. Soc. Am. B* **9**, 40 (1992)
- [146] A. Mekis, J. U. Nöckel and A. D. Stone, unpublished
- [147] J. Z. Zhang, D. H. Leach and R. K. Chang, *Opt. Lett.* **13**, 270 (1988)
- [148] D. L. Shepelyansky and A. D. Stone, *Phys. Rev. Lett.* **74**, 2098 (1995)