COURSE 3

SEMICLASSICAL MECHANICS OF REGULAR AND IRREGULAR MOTION

Michael BERRY *

Institute for Theoretical Physics, University of Utrecht
Utrecht, The Netherlands

*Permanent address: H.H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK.

G. Iooss, R.H.G. Helleman and R. Stora, eds.
Les Houches, Session XXXVI, 1981 – Comportement Chaotique des Systèmes Déterministes/ Chaotic Behaviour of Deterministic Systems

Contents

1. Introduction

After half a century, the relation between quantum mechanics and classical mechanics is still not fully understood. Uncertainty extends even to the foundations of the subject: it is not known in general whether or how it is possible to specify uniquely the quantum mechanics of a system whose classical mechanics is given. This fundamental "problem of quantization" will not be discussed here. Instead, I shall concentrate on rather simple systems – noninteracting particles moving in low-dimensional spaces under external forces – for which it is known how to specify the quantum mechanics, in terms of Schrödinger's equation involving well-defined operators. For these systems, I shall study the *semiclassical limit*, i.e. the behaviour of wave functions, energy levels, etc. as Planck's constant $\hbar$ tends to zero (by comparison with classical quantities having the same dimensions). This is not the same as the classical limit, for which $\hbar$ is precisely equal to zero because, in general, quantal functions are nonanalytic in $\hbar$ as $\hbar \to 0$. So the semiclassical limit cannot be related to the classical limit by perturbation theory, but has a rich and interesting structure of its own. Nevertheless there must be some kind of correspondence principle, according to which the semiclassical limit reflects the nature of the underlying classical motion. We shall learn that what really affects the semi-classical mechanics is whether the classical motion is regular (predictable, integrable) or irregular (unpredictable, chaotic, non-integrable).

Of course these problems have a wide variety of actual and potential applications in quantum mechanics and indeed more generally throughout physics and applied mathematics wherever short waves are involved. Examples are the vibration spectra of non-symmetrical molecules, modes of acoustic oscillation in rooms with typical shapes, and the optics of waveguides. I shall not discuss any of these, but shall consider semiclassical mechanics as worth studying for its own sake, in order to understand the connection between two important branches of theoretical physics.

Many different sorts of scientists and mathematicians have studied

semiclassical mechanics (or short-wave asymptotics) and the literature is correspondingly varied and scattered. I do not intend to attempt a complete survey, but shall concentrate on the qualitatively different sorts of semiclassical wave functions and spectra corresponding to qualitatively different sorts of classical motion. This is an aspect of quantum mechanics whose significance has been appreciated only recently, as advances in classical mechanics have unravelled the intricate varieties of predictability and chaos that orbits can display. To my knowledge this material has been reviewed only once before by Zaslavsky [118]; although Percival [1] and Duistermaat [2] have given reviews of aspects of it, from (respectively) physical and mathematical viewpoints. "Prehistoric" semiclassical mechanics was reviewed by Berry and Mount [3]. My treatment here will be nonrigorous, combining analytical and pictorial arguments, simple models and the results of computation, to bring out what I consider to be the essential points.

2. Regular and irregular classical motion

2.1. Two contrasting types of orbit

Here I discuss two contrasting types of Hamiltonian motion, that is motion governed by Newton's equations without dissipation. This is necessary in order to set the scene for the subsequent quantal treatment. More detailed expositions have been given by Ford [4], Arnol'd [5], Berry [6] and Helleman [7].

On the one hand, there is *regular motion*. This is exemplified by the one-dimensional oscillator (harmonic or anharmonic) which in its physical realization as a pendulum is the epitome of predictability ("as regular as clockwork"). Another example is the elliptical orbits of the planets when mutual perturbations are ignored. In systems with regular motion, trajectories with neighbouring initial conditions separate linearly.

On the other hand, there is *irregular motion*. This is exemplified by the (classical) motion of colliding molecules in a gas. If the molecules are confined to a plane, transformed into hard discs and all held fixed except one, then that one executes a motion with two freedoms idealising a pinball machine. Such motion is unpredictable in the sense that neighbouring trajectories separate exponentially, resulting in a sensitivity to initial conditions which has surprising consequences. For example, in order to ac-

curately predict the motion for n collisions (after which the angle of emergence is in error by, say, 90°), it is necessary to specify the initial position and momentum values to a number of digits D proportional to n, so that the capacity of any feasible calculator is exhausted for rather small numbers of collisions. (If the discs have radius r and mean separation l, a rough estimate for D when the variables are expressed in base-b notation is

$$D \approx n \log_b l/r, \tag{2.1}$$

as readers may verify. If $l/r = 10$, therefore, even specifying initial conditions with a relative error of one part in a million will ensure predictability for only six collisions.)

The distinction between regularity and irregularity is embodied in the geometry of typical trajectories in the system's phase space over infinitely long times. For N freedoms this is the $2N$-dimensional space q, p where $q = (q_1 \ldots q_n)$ are the coordinates and $p = (p_1 \ldots p_N)$ the momenta. There is a hierarchy of types of motion, increasingly chaotic, denoted by the terms integrable, quasi-integrable, ergodic, mixing, K-system, B-system, whose meaning is explained by Lebowitz and Penrose [8]. I shall not describe this hierarchy in detail, but give sufficient background for the later discussion of quantum mechanics.

For illustrative purposes it is not convenient to consider only smooth Hamiltonians, and I now introduce an important class of discontinuous systems with $N = 2$, namely the so-called *planar billiards*. These have Hamiltonian

$$\left.\begin{aligned} H(x, y; p_x, p_y) &= (p_x^2 + p_y^2)/2 \text{ inside a boundary } B \text{ in } q \text{ space} \\ &= \infty \text{ outside } B \end{aligned}\right\}. \tag{2.2}$$

Motion consists of straight segments joined by specular reflections at B. As we shall see, the nature of the orbits is very sensitive to the form of B (for an elementary review of billiard dynamics, see Berry [9]).

2.2. *Integrable systems*

The simplest situation, corresponding to the regular case, is fully integrable motion. Here there exist N constants of motion in the form of functions $C_i(q, p)$ $(1 \leq i \leq N)$ in phase space, assumed to be "in involution" (i.e. all Poisson brackets between pairs of C_i vanish). One of these constants is the Hamiltonian itself if this is time-independent. If the C_i are independent of

one another, their existence restricts motion to a surface Σ in phase space; Σ has dimensionality $N(=2N-N)$. Arnol'd [5] gives a clear formulation and proof of an important theorem, implicit in older literature, stating that if the C_i are "smooth enough", then

(i) Σ is an N-dimensional torus

(ii) the motion can be "integrated", i.e. the trajectories $q(t)$, $p(t)$ can be determined by elimination and integration. The proof (see also ref. [6]) involves constructing N nonsingular vector fields on Σ ("if a hairy Σ can be combed N ways without a singularity, then Σ is an N-torus"). (If the conditions of the theorem are not satisfied, and the vector fields do have singularities, then Σ need not have the topology of an N-torus, as will be shown in section 2.5 with a curious example.)

The simplest integrable systems are stationary harmonic or anharmonic oscillators with $N=1$ and mass μ, whose Hamiltonian is

$$H=\frac{p^2}{2\mu}+V(q), \tag{2.3}$$

where $V(q)$ (fig. 1a) is a potential well. The energy $E=H(q,p)$ is conserved, and since $N=1$ this one constant of motion suffices to make the system integrable. The "1-tori" on which motion occurs in phase space are simply the closed contours of H (fig. 1b).

All separable systems are integrable (they decouple into N one-dimensional systems). In particular, a particle moving in the plane ($N=2$) under a central potential $V(r)$ is integrable, the two constants of motion being E and the angular momentum L. Each choice of E and L labels a 2-torus Σ in the 4-dimensional phase space (fig. 2a). The trajectory winds

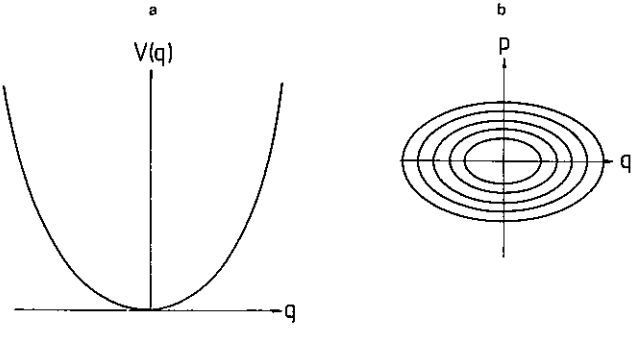
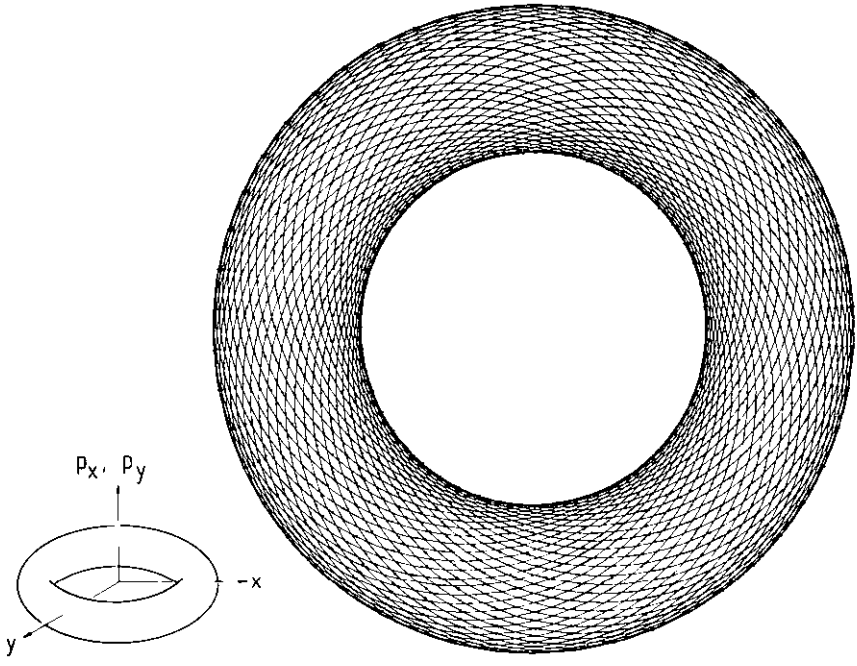


Fig. 1.



(a) Fig. 2. (b)

around Σ in a manner soon to be explained, and typically fills Σ as $t \to \infty$. When projected "down" p onto the coordinate plane x, y (fig. 2b), the trajectory is enveloped by a *caustic* in the form of two circles at the radii of closest approach and farthest recession from the centre of force. Caustics are the singularities of the projections of Σ. In the circular billiard (fig. 3) the outer circle corresponds to the boundary B and is not a caustic. In the

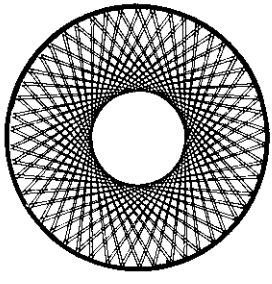


Fig. 3.

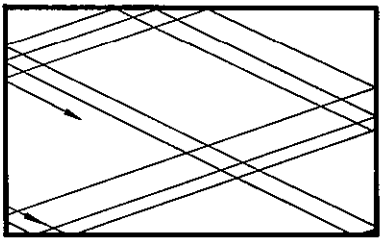


Fig. 4.

rectangular billiard (fig. 4), which is separable with p_x^2 and p_y^2 conversed, there are no caustics because the boundaries of the projection of Σ (a flat car tyre with four sheets above points q) coincide with B. Another separable billiard is the ellipse [9], whose orbits repeatedly touch an ellipse or hyperbola confocal with B.

The existence of tori throughout the phase space of an integrable system makes it natural to introduce an alternative set of coordinates and momenta known as *action-angle variables*, which will play an important part in the quantum mechanics of integrable systems. The *actions* $I=\{I_1 \ldots I_N\}$ are particular combinations of the C_i which label the tori Σ, defined in the most topologically natural way as

$$I_i \equiv \frac{1}{2\pi} \oint_{\gamma_i} p \cdot \mathrm{d}q, \tag{2.4}$$

where γ_i is the ith irreducible circuit of the torus. Now let I be the momenta of new phase-space variables. The conjugate coordinates are called the *angles* $\theta=\{\theta_1 \ldots \theta_N\}$. Any point q, p lies on a torus (fig. 5) labelled by

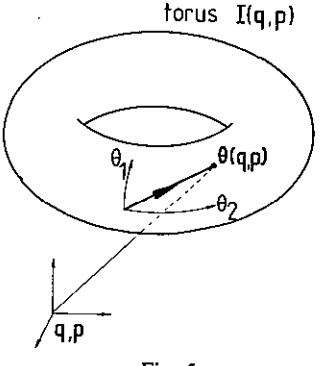


Fig. 5.

$I(q,p)$, and $\theta(q,p)$ locates the position of q, p on this torus.

The variables θ, I and q, p are related by a cononical transformation specified as follows. Let the torus Σ labelled by I be defined by a multivalued function with branches $p_j(q';I)$ in the phase (fig. 6). Then the generating function of the transformation is $S(q,I)$ with branches

$$S_j(q;I)=\int_{q_0}^{q} p_j(q';I)\cdot \mathrm{d}q', \tag{2.5}$$

where q_0 is an arbitrary point "beneath" Σ. This gives the transformation

$$\begin{aligned}&(q,p)\leftarrow S(q,I)\rightarrow(\theta,I)\\ &p=\nabla_q S,\ \theta=\nabla_I S.\end{aligned} \tag{2.6}$$

It follows from these definitions that each angle variable θ_i changes by 2π during the corresponding circuit γ_i of Σ, thus justifying the term "angle". The fact that S is locally single-valued implies from eq. (2.6) that tori have the "Lagrangian" property

$$\partial p_i/\partial q_j=\partial p_j/\partial q_i. \tag{2.7}$$

Because there are only N independent constants of motion, each member of the original set C_i must be expressible in terms of the N I_i's. In particular, the Hamiltonian is conserved and can be written in the new variables as

$$H=H(I). \tag{2.8}$$

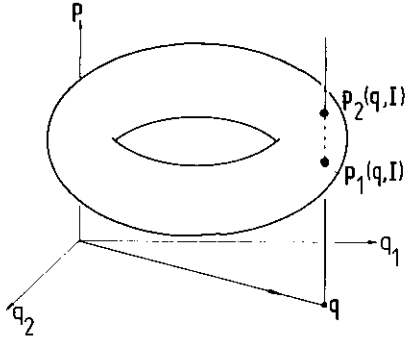


Fig. 6.

By one of Hamilton's equations, this means that the angle variables change at a constant rate, given by

$$\theta = \omega t + \theta_0, \tag{2.9}$$

where $\omega = \{\omega_1 \ldots \omega_N\}$ are the frequencies, given by

$$\omega(I) = \nabla_I H(I). \tag{2.10}$$

The components ω_i give the rate at which a trajectory winds round Σ in directions labelled by the angles θ_i. In the general case, all ω_i will be mutually incommensurable and the orbit will eventually fill Σ densely. But if all ω_i are rationally dependent, i.e. if

$$\omega = M\omega_0, \tag{2.11}$$

where $M = \{M_1 \ldots M_N\}$ is a lattice vector (i.e. with integer components), then each orbit on Σ will be *closed*. In fact closure will occur after M_1 circuits of θ_1, M_2 circuits of $\theta_2 \ldots$, M_N circuits of θ_N. These closed orbits are *nonisolated*, in the sense that each "rational" torus [i.e. satisfying eq. (2.11)] will contain infinitely many of them, related by parallel translation in θ (fig. 7 shows the case where $N=2$, $M_1=1$, $M_2=0$). In the typical case, where $\omega(I)$ varies smoothly with I, tori made of closed orbits are of measure zero but nevertheless densely distributed. Finally, note that θ provides an invariant measure on Σ, in the sense that a cloud of phase-space points with uniform density θ preserves density as the cloud moves under time-evolution governed by $H(I)$.

2.3. Chaotic systems

The integrable systems so far considered are rare; amongst Hamiltonian systems, "almost all" are nonintegrable in the sense that there are no

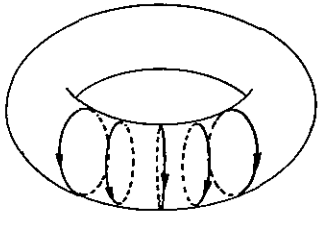


Fig. 7.

global constants of motion other than H. An important class of nonintegrable system is those with the property of *ergodicity:* almost all orbits eventually explore almost all points in a $2N-1$ dimensional energy surface, defined by $H(q,p)=$ constant, instead of being confined to an N-torus as with integrable systems. The case $N=1$ is degenerate, because $2N-1=N$ and orbits are trivially ergodic as well as integrable (cf. fig. 1b) - provided of course that it is time-independent. When $N>1$, ergodicity is usually associated with (but does not imply) motion which is chaotic or unpredictable as briefly described in section 2.1. The precise sense in which such motion is chaotic can be understood by dividing ("coarse-graining") phase space into cells ("Markov partitions") and studying the sequence ("Bernoulli shift") in which cells are explored; the sequences are often indistinguishable from random sequences (see for example Ornstein [105], and the elementary discussions in refs. [4] and [6]).

Are there any ergodic systems with $n=2$? Yes. It was shown by Sinai [10] that the billiard motion of a particle moving on the coordinate torus x, y (square with opposite sides identified) containing a circular reflecting obstacle with radius R (fig. 8) is ergodic if $R>0$. In this case ergodicity is a consequence of exponential chaos (unpredictability) which in turn results from the defocusing of particle beams that hit the disc (fig. 9). Another effect of defocusing is that no orbits form caustics in coordinate space and indeed there are no phase space tori whose projections would have them as singularities. Sinai's billiard is the simplest example of a system of two or more hard spheres or discs moving in a compact space and colliding elastically. For the case where the number of particles is very large, Sinai's proof at last established the ergodicity of the hard-sphere gas.

Another ergodic billiard is the "stadium" of Bunimovich [11,12], consisting of two semicircles with radius R joined by parallel straight lines with length L (fig. 10) with $L>0$. The semicircles cause a convergence of beams

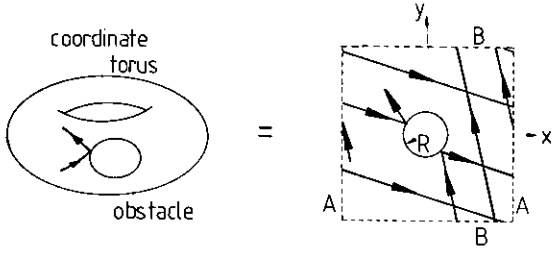


Fig. 8.

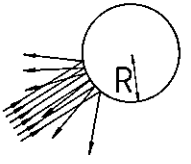


Fig. 9.

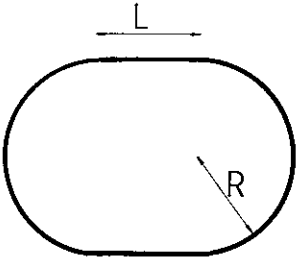


Fig. 10.

of trajectories but this is outweighed by the subsequent divergence before the next reflection. Again no caustics form, as the orbit in fig. 11 illustrates (cf. fig. 3).

In these ergodic billiards, a typical orbit passes through almost every point within B with almost every direction if followed long enough. The energy surface thus explored is 3-dimensional $(2N-1)$ in contrast to the 2-tori explored when there is an additional constant of motion. But not every orbit is typical: just as in integrable systems there is a dense-but-zero-measure set of closed orbits (each of which explores a 1-dimensional region in phase space). For integrable systems (with $N=2$) we have seen that these closed orbits are not isolated (fig. 7) but fill tori in a one-parameter family. And in some billiards (such as those of Sinai and Bunimovich) it is also possible to have one-parameter families of nonisolated closed orbits (fig. 12), although these do not fill tori. For ergodic billiards the nonisolated orbits drastically slow down the exploration of the energy surface [9], and are

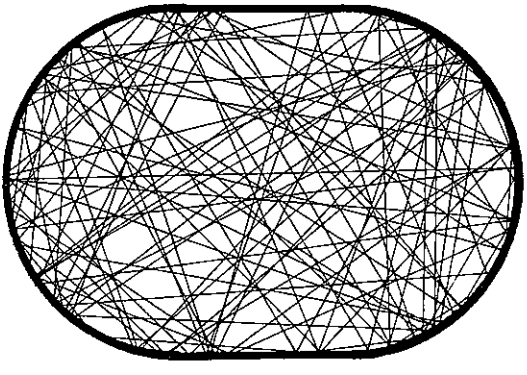


Fig. 11.

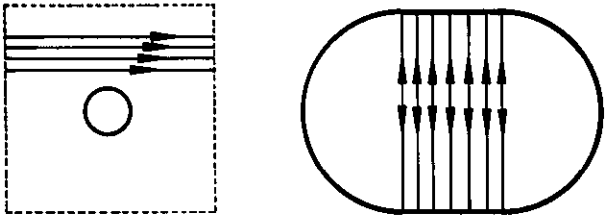


Fig. 12.

probably responsible for long tails in autocorrelation functions (Berry and Casati, unpublished). But vastly more numerous are the *isolated closed orbits*, some of which are shown in fig. 13. The distinction between isolated and nonisolated closed orbits has an important effect on the clustering of quantum-mechanical energy levels, to be explained in section 5.

These closed orbits, isolated and nonisolated alike, are unstable: the slightest error in position or momentum gives an orbit separating rapidly (exponentially, for isolated orbits) from the closed one and eventually exploring all the energy surface. By contrast, the closed orbits in an integrable system are only linearly unstable [cf. eq. (2.9)], and a perturbed orbit, far from filling the energy surface, merely fills a nearby torus. In this connection it is amusing to note that in the ideal roulette wheel and coin toss - two exemplars of unpredictable systems whose randomness arises largely from imperfect knowledge of initial conditions - the motion is in fact integrable (rigid disc rotating in or out of its plane), so that the instability is of a linear type and therefore much weaker than in, say, a pinball machine.

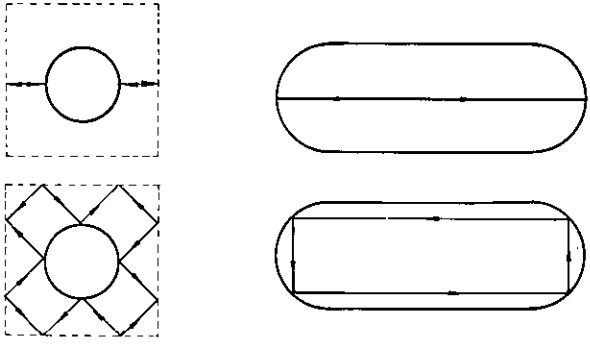


Fig. 13.

Another (non-billiard) ergodic system with $N=2$ is geodesic motion on a compact surface of everywhere negative curvature (which cannot be realized in Euclidean 3-space) [5]. No smooth Hamiltonian of the canonical type $H=$kinetic + potential has been proved to be ergodic. However, Gutzwiller [13] gives strong evidence that the "anisotropic Kepler problem", with

$$H=ap_x^2+bp_y^2-\frac{C}{(x^2+y^2)^{1/2}} \qquad (a\neq b) \tag{2.12}$$

is ergodic, and Percival (private communication) has reason to believe that orbits in the potential

$$H=\frac{p_x^2+p_y^2}{2}+x^2y^2 \tag{2.13}$$

explore their (unbounded) energy surfaces chaotically.

2.4. Quasi-integrable systems

These are neither integrable nor ergodic. Excluding billiards for the moment, they have smooth Hamiltonians of the form

$$H(q,p)=H_0(q,p)+\varepsilon H_1(q,p), \tag{2.14}$$

where H_0 is integrable and ε is a small parameter that turns on a generic perturbation H_1. When $\varepsilon=0$, all trajectories lie on tori filling phase space. What happens to these tori when $\varepsilon\neq 0$? The answer is given by the celebrated Kolmogorov [14], Arnol'd [15], Moser [16] (KAM) theorem (see also refs. [5,6]): under perturbation, most tori survive (albeit distorted). Therefore the motion is not ergodic. But for almost all H_1, some tori are destroyed, so the motion is not integrable either. The destroyed tori form a set of finite measure growing with ε. They are centred on those unperturbed tori whose frequencies ω_i are incommensurable, i.e. whose orbits are closed. The manner of their destruction is complicated but now fairly well understood [6]: of the continuous family of linearly unstable closed filling each rational torus, only a finite even number of closed orbits survives perturbation; half of these are unstable and half are stable. Motion near the unstable orbits is chaotic and fills regions with dimensionality $2N-1$. Near the stable closed orbits, most trajectories lie on "higher-order" tori, but these, like the original "parent" tori, have gaps near sites

where "rational" tori would be, and the whole structure repeats in microcosm, recursively down to infinitely fine scales.

The solar system is quasi-integrable [15]. Planets unperturbed by their neighbours move (integrably) in Kepler ellipses. Including the perturbation results in a Hamiltonian of type (2.14) [6]. Some of the effects described in the last paragraph can be seen in the asteroids, which correspond, in effect, to an ensemble of zero-mass "test particles". Most of them move in approximately elliptical orbits, in spite of being perturbed by Jupiter, but there are gaps in the asteroid belt where orbital motion would be commensurate with (resonant with) Jupiter's; these gaps correspond to the destroyed tori. There are similar gaps in Saturn's rings, which may correspond to resonant perturbation of the orbits of ring particles by the satellite Mimas.

The tori whose existence the KAM theorem guarantees can be discovered "experimentally" by the caustics their orbits envelop when projected "down" onto coordinate space (cf. fig. 2). An example is shown in fig. 14; this was computed by Noid and Marcus [23] for a particle in the (nonintegrable) field of two anharmonically coupled harmonic oscillators. How can we deduce that the surface Σ in phase space, whose projection displays the caustics, is a torus? By the following procedure, devised by Ozorio de Almeida and Hannay [17] as part of a detailed study of the ways 2-tori can be embedded in 4-space and projected onto 2-space; traverse all branches of the caustic, keeping two-sheeted regions on one's right; one

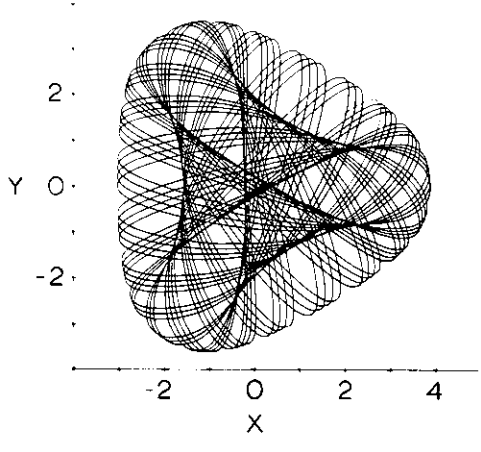


Fig. 14. Reprinted from ref. [77] with permission.

will rotate clockwise r times (not counting the necessary half-turns at cusps). The genus (number of handles) of the Σ which projects to give the caustic, is $1-r$. For a torus, the genus must be unity, so r must be zero; it is easy to confirm that the caustic of fig. 14 passes the torus test. Generically, the forms of caustics are classified by catastrophe theory, but because I discussed this aspect in detail in last year's Les Houches lectures [18] I shall not elaborate it here.

Returning now to billiards, it is necessary in quasi-integrable cases to be very careful in specifying what class of deformation of the boundary B is being considered. For example we have already seen that if the integrable circle billiard is perturbed to the stadium (fig. 10), all tori are instantly destroyed (no matter how small L/R is) because the system becomes ergodic. However, Lazutkin [19] proved for billiards a theorem analogous to the KAM theorem but more general (in that it is nonperturbative): if B is convex and smooth enough, some orbits (of positive measure) will envelop caustics – i.e. there will be some tori in phase space. The sufficient condition for smoothness is that B's radius of curvature as a function of arc length must possess 553 continuous derivatives! Such a large number is almost certainly not necessary, but to have no continuous derivative is insufficient as the stadium example shows. Berry [9] illustrates Lazatkin's theorem with computations indicating quasi-integrable behaviour (i.e. some tori, some chaos) for a family of (analytic) oval billiards, and Bennettin and Strelcyn [20] study a generalization of the stadium made from four circular arcs, which has discontinuous curvature but nevertheless displays quasi-integrability.

2.5. Pseudo-integrable systems

Even if N constants of motion exist, independent and in involution, so that each orbit restricted to an N-dimensional surface Σ in phase space, it is not always the case that Σ is a torus. For billiards whose boundary B is a polygon with angles which are rational multiples of π (apart from the rectangle, equilateral triangle, 30°-60°-90° triangle and 45°-45°-90° triangle, which are integrable), Σ is not a torus but a *multiply-handled sphere* (i.e. genus $g>1$) (for a reason to be explained, Arnol'd's theorem (section 2.2) does not apply). Richens and Berry [21] call these systems "pseudo-integrable"; they were previously studied by Zemlyakov and Katok [22].

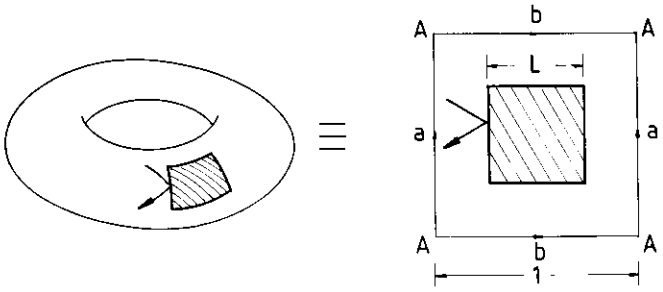


Fig. 15.

To illustrate this unexpected behaviour, consider the "square torus billiard" (fig. 15), which resembles Sinai's billiard except that the reflecting disc is replaced by a square. It is clear that the squares v_x^2 and v_y^2 of the velocity components are separately conserved, and that at most four directions are available to each trajectory $(v_x, v_y; -v_x, v_y; v_x, -v_y; -v_x, -v_y)$. Therefore the phase-space surface Σ consists of four sheets, each being a copy of the accessible xy area, connected at the central square according to the identifications in fig. 16. To construct Σ, these four sheets must be sewn together. It is simplest first to join the inner squares to get a "skeleton" for Σ (fig. 17). Then the outer tori must be sewn onto each "square" (fig. 18). Finally, the two remaining ("l, q") holes must be joined

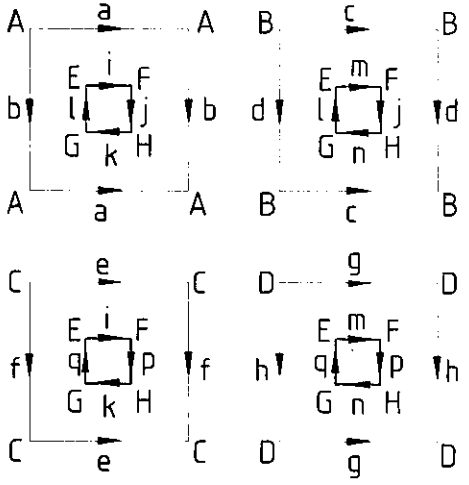


Fig. 16.

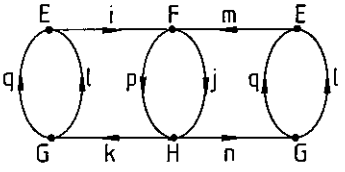


Fig. 17.

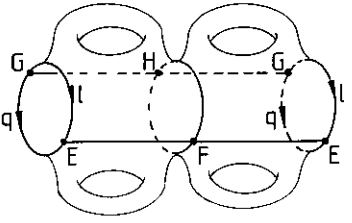


Fig. 18.

(fig. 19). It is clear from this construction that far from being a torus, Σ is a five-handled sphere.

Now, the constants v_x^2 and v_y^2 are smooth, independent and in involution. What then prevents Arnol'd's theorem (described in the first paragraph of section 2.2) from being applied? The answer, elaborated by Richens and Berry [21] is that certain vector fields on Σ, constructed from v_x^2 and v_y^2, possess singularities at the vertices of the reflecting square. Indeed it is possible to establish the topology of Σ from a study of these singularities. As an example, formulae given in ref. [21] yield the following equations for the genus g of Σ for billiards in a regular polygon with m sides:

$$g=\frac{(m-1)(m-2)}{2} \qquad (m \text{ odd})$$
$$g=\frac{(m-2)^2}{4} \qquad (m \text{ even}) \tag{2.15}$$

As expected, this gives $g=1$ for the triangle and square, which are integrable.

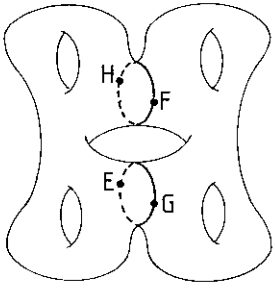


Fig. 19.

As discussed in ref. [22], rational billiards have a chaotic property resulting from the splitting of beams of trajectories hitting polygon vertices. According to Richens and Hannay (private communication) it is convenient to study this computationally in terms of the one-dimensional map produced by successive intersections of an orbit with a curve cutting Σ (for a torus this map is a simple rotation, but if $g>1$ it is a "shuffling transformation").

For polygons with angles that are irrational multiples of π, each reflection of a typical trajectory produces a segment moving in a new direction, so these systems are not pseudo-integrable (unless we accept surfaces with $g=\infty$). Are they ergodic? Computations by Casati and Ford [24] suggest that the answer is yes, but that exploration of the three-dimensional energy surface is very slow and without exponential separation of orbits; to my knowledge nobody has proved a theorem demonstrating (or disproving) the ergodicity for irrational polygons.

2.6. *Discrete area-preserving maps of the plane*

The distinction between regular and irregular motion is very clearly exhibited by discrete maps M in the phase plane whose variables are a coordinate q and a momentum p. Under M (fig. 20), each point (q_{n+1}, p_{n+1}) is a deterministic function of its antecedent (q_n, p_n):

$$\begin{aligned} M:\ q_{n+1} &= q_{n+1}(q_n, p_n) \\ p_{n+1} &= p_{n+1}(q_n, p_n) \end{aligned} \tag{2.16}$$

Area-preservation is ensured by requiring

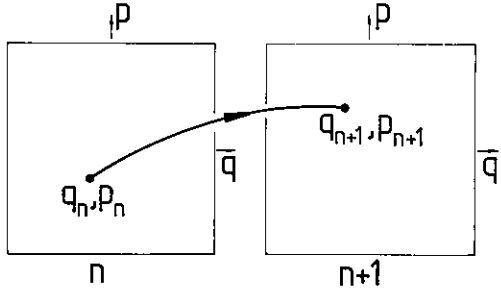


Fig. 20.

$$\det \begin{pmatrix} \dfrac{\partial q_{n+1}}{\partial q_n} & \dfrac{\partial q_{n+1}}{\partial p_n} \\ \dfrac{\partial p_{n+1}}{\partial q_n} & \dfrac{\partial p_{n+1}}{\partial p_n} \end{pmatrix} = \frac{\partial(q_{n+1}, p_{n+1})}{\partial(q_n, p_n)} = 1. \tag{2.17}$$

M can be regarded simply as an abstract dynamical system whose evolution takes place in discrete time steps according to the "Hamilton equations" (2.16) with (2.17) replacing Liouville's theorem on conservation of phase-space volume. But these maps also arise from continuous time dynamical systems, in two ways.

Firstly, one may have $N=2$ and a time-independent Hamiltonian. In this case q and p are coordinates on a "surface of section" (see e.g. ref. [6]) through the three-dimensional energy surface. M is determined by the succession of points q_n, p_n in which a trajectory intersects the surface of section. For billiards, a convenient surface of section is given by successive bounces at B, and variables for which M is area-preserving (see e.g. ref. [9]) are $q=$(arc length round B) and $p=$(cosine of angle made by emerging trajectory with forward tangent to B).

Secondly, one may have $N=1$ and a Hamiltonian periodic in time, i.e.

$$H(q, p, t+T) = H(q, p, t). \tag{2.18}$$

Then M is defined by a "stroboscopic phase portrait", i.e. by jumps of points q, p between snapshots of the motion at intervals T, i.e.

$$q_n \equiv q(nT), \qquad p_n \equiv p(nT), \tag{2.19}$$

and the motion between snapshots is ignored.

In terms of M, regularity or irregularity depends on the manner in which iterates (q_n, p_n) of some initial point (q_0, p_0) are distributed as $n \to \infty$. There are three possibilities. Firstly, iterates may lie on a zero-dimensional set in the plane (fig. 21), by forming a closed orbit, i.e. a *fixed point* of some finite power N of M:

$$q_{n+N} = q_n, \qquad P_{n+N} = p_n. \tag{2.20}$$

Secondly, iterates may fill a one-dimensional set in the plane (fig. 22), a so-called *invariant curve* which maps onto itself although its individual points do not. An example is the twist map which in polar coordinates is

$$M: \; r_{n+1} = r_n, \qquad \theta_{n+1} = \theta_{n+1} + 2\pi\alpha(r_n). \tag{2.21}$$

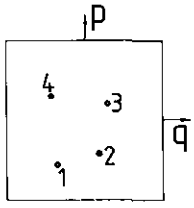


Fig. 21.

Here the invariant curves are circles. If the "rotation number" α is irrational, iterates eventually fill the circle, while if $\alpha = M/N$ (M, N mutually prime finite integers) the iterates form a closed orbit with period N. And thirdly, iterates may fill a two-dimensional set in the plane (fig. 23), a so-called *chaotic area*. I shall not discuss more exotic possibilities, such as orbits filling a "fractal" set with intermediate dimension, because there is no firm evidence that they occur in area-preserving maps (although they do occur - as "strange attractors" (see ref. [25] and other papers in the same volume) - in area-contracting maps corresponding to dissipative systems).

Integrable maps are those like fig. 22, where all points lie on invariant curves. These curves are the analogues of tori, and in the "surface of section" interpretation are precisely sections of 2-tori. Their existence throughout the phase plane implies a "constant of motion" in the form of a function $f(q, p)$ whose contours are the invariant curves and whose value is unchanged under M (for a twist map, $f = q^2 + p^2$).

Ergodic maps are those where almost every initial point explores almost all the phase plane, so that no finite measure is covered with invariant curves. An example is "Arnol'd's cat" [100], for which phase space is the unit torus and M the linear map corresponding to hyperbolic shear:

$$M: \begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} q_n \\ p_n \end{pmatrix}. \tag{2.22}$$

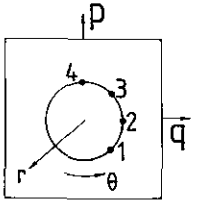


Fig. 22.

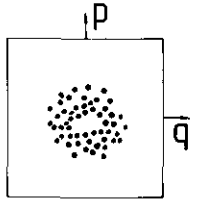


Fig. 23.

In the infinite plane this is trivially simple, but on the torus it winds area elements round and round in an irrational direction so that initially simple shapes quickly become unrecognizable (fig. 24). (Even after one hyperbolic mapping it can be hard to recognize shapes, and this idea is the basis of a Victorian amusement called the "anamorphic picture" (fig. 25), which can be comprehended only when viewed after an "inverse mapping" consisting of reflection from the exterior of a convex cylinder). There are closed orbits in Arnol'd's cat – they are executed by every point with rational

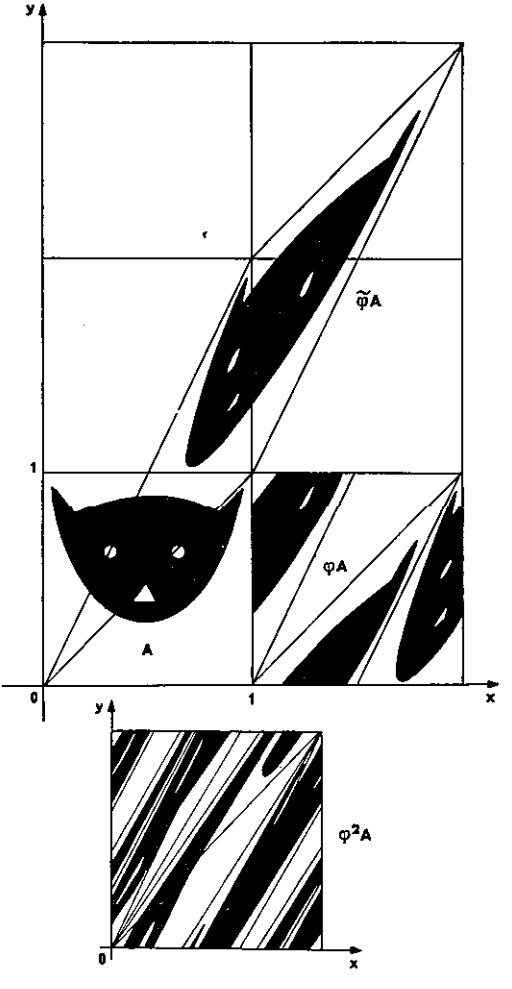


Fig. 24. Reprinted from ref. [100] with permission.

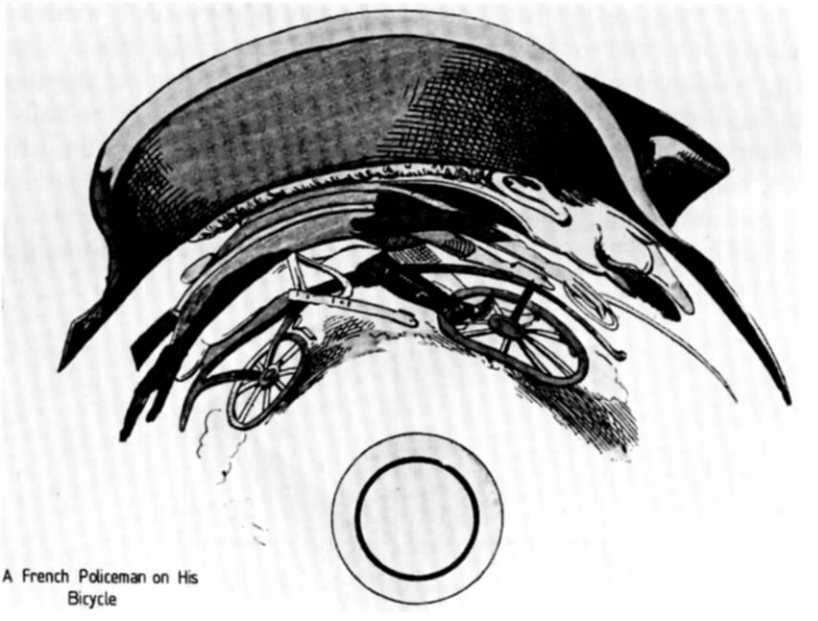


Fig. 25.

q, p - but they are of measure zero and all isolated and unstable.

A quasi-integrable map, with a hierarchy of fixed points, invariant curves (tori) and chaotic areas, is generated by

$$H(q, p, t) = \frac{p^2}{2\mu} + TV(q) \sum_{n=-\infty}^{\infty} \delta(t - nT), \tag{2.23}$$

corresponding to a particle with mass μ moving freely along a line between impulses at intervals T whose magnitude depends on the coordinate q. From Hamilton's equations, M is easily found to be

$$\begin{aligned} M: q_{n+1} &= q_n + \frac{Tp_n}{\mu} \\ p_{n+1} &= p_n - T\frac{\partial V}{\partial q}(q_{n+1}) \end{aligned} \tag{2.24}$$

The choice $V(q) = Aq^4/4$ (anharmonic potential well), followed by scaling, gives

$$M: q_{n+1} = q_n + p_n$$
$$p_{n+1} = p_n - q_{n+1}^3 \tag{2.25}$$

Figure 26 shows the iterates of a number of initial points in the qp plane. Fixed points, invariant curves, and a continuous chaotic area are all clearly visible. Points outside the chain of eight stable "islands" escape rapidly [as $\exp(\exp 3n)$] to infinity under M.

3. Wave functions and Wigner functions

3.1. Maslov's association between phase-space surfaces and semiclassical wave functions

Now we turn to the main work of these lectures, which is to study how the different types of underlying classical motion manifest themselves in quantum mechanics. We begin by considering the morphologies of eigenfunc-

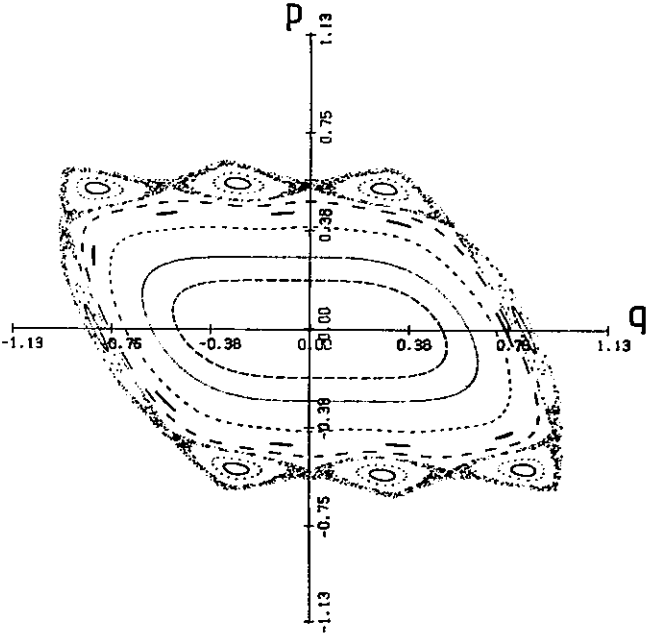


Fig. 26.

tions $\psi(q)$ of time-independent Hamiltonian operators whose analogous classical Hamiltonians generate regular or irregular bounded motion. The result will be strong evidence that the morphologies corresponding to the two sorts of motion are indeed very different. The suggestion that this might be the case goes back to Einstein [36], and was revided in a modern form by Percival [37].

I begin by describing an important semiclassical concept developed in full geometric detail in the 1960s by Maslov [26] but implicit in earlier work by Van Vleck [27] and Keller [28]. Maslov's ideas have been reviewed by Kravtsov [29] and (briefly) by me [18] in the context of propagating waves, by Duistermaat [2] and Guillemin and Sternberg [30] and Voros [44] from the mathematical viewpoint, and by Percival [1] and Eckmann and Sénéor [31] using simple examples.

The concept is an association between wave functons $\psi(q)$ and N-dimensional surfaces Σ in the $2N$-dimensional phase space q, p. ψ need not be an eigenstate of the Hamiltonian and Σ need not be one of the tori discussed in section 2. To start with, the association between ψ and Σ is purely geometric; dynamics is introduced later. Locally (fig. 27), Σ can be written as a function $p(q)$, and corresponds to an N-parameter ensemble of states (points) in classical phase space. We define a density on Σ in which these states are distributed uniformly in some coordinate $Q = \{Q_1 \ldots Q_N\}$. A convenient definition of Q can be accomplished if we imagine an N-parameter family of surfaces filling phase space near Σ, labelled by $P = \{P_1 \ldots P_N\}$, and then regard Q, P as alternative phase space variables in a canonical transformation from q, p. This is specified by a generating

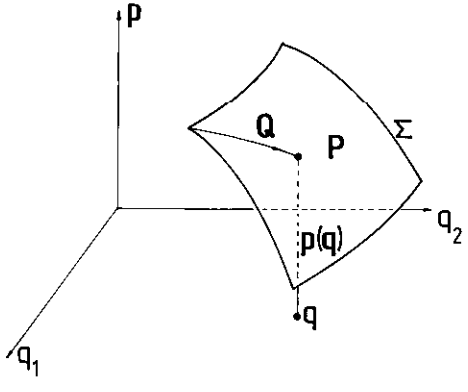


Fig. 27.

function $S(q;P)$ by

$$(q,p) \leftarrow S(q;P) \rightarrow (Q,P)$$
$$p = \nabla_q S,\ Q = \nabla_p S \tag{3.1}$$

(cf. the transformation (2.6) to action–angle variables, which is a special case). These equations imply that Σ has the "Lagrangian" property (2.7), and this will henceforth be assumed.

We want to associate with Σ a wave

$$\psi(q) = a(q)\mathrm{e}^{ib(q)}. \tag{3.2}$$

Because there is as yet no dynamics, ψ is unrestricted by any wave equation, and we are free to choose any association with Σ. There is, however, a natural choice, based on two physical principles, one for $a(q)$ and one for $b(q)$.

For the amplitude $a(q)$ we require that the wave intensity $|\omega|^2$ is proportional to the density of points in classical q-space. This is obtained by projecting the density on Σ and using eq. (3.1), in terms of the Jacobian determinant

$$a^2(q) = K\left|\frac{\mathrm{d}Q}{\mathrm{d}q}\right| = K \det\left|\frac{\partial^2 S}{\partial q_i \partial P_j}\right|, \tag{3.3}$$

where K is a constant.

For the phase $b(q)$ we use de Broglie's rule relating the classical momentum $p(q)$ to the wave vector $k(q)$ of a locally plane wave, so that, in terms of the phase difference between two points separated by δq,

$$b(q+\delta q) - b(q) = \nabla b \cdot \delta q = k \cdot \delta q = \frac{p \cdot \delta q}{\hbar}, \tag{3.4}$$

thereby introducing Planck's constant $\hbar$. Thus

$$\nabla b(q) = p(q)/\hbar \tag{3.5}$$

and

$$b(q) = \frac{1}{\hbar}\int_{q_0}^{q} p(q') \cdot \mathrm{d}q' = \frac{S(q;P)}{\hbar}, \tag{3.6}$$

again using eq. (3.1), where q_0 is a fixed position where S is defined to vanish. This phase is uniquely defined locally, independent of the integration path between q_0 and q, because the Lagrangian property (2.7) makes

S a (locally) single-valued function of q.

In this way we construct the local plane wave

$$\psi(q) = K \left| \det \frac{\partial^2 S(q;P)}{\partial q_i \partial q_j} \right|^{1/2} \exp\left(\frac{i}{\hbar} S(q;P)\right) \tag{3.7}$$

associated with the surface Σ labelled by P. Why should this association be a useful one? The answer lies in regarding ψ as an initial quantum state and letting it evolve according to the Schrödinger equation under dynamics governed by a Hamiltonian H. After a time t, ψ will have evolved into a new wave ψ'. In addition, Σ will have evolved into a new surface Σ' by virtue of the classical Hamiltonian motion of each of its points. Solving the time-dependent Schrödinger equation asymptotically, i.e. to lowest order in $\hbar$, it can be shown (see e.g. ref. [27] or Dirac [32]) that the ψ' can be constructed from Σ' by precisely the recipe (3.7). Therefore the construction persists in time, at least at the level of semiclassical approximations, and so represents a natural association between evolving quantal waves and N-parameter families of classical orbits (i.e. evolving classical surfaces). As $t \to \infty$, the association breaks down in a most interesting way, to be discussed in section 6. Now we pursue the purely geometric aspects of the association.

3.2. *Globalization*

The procedure based on fig. 27 and eq. (3.7) is well defined only if $p(q)$ is singlevalued. But what if Σ is curved (fig. 28) in such a way that a fibre drawn "upwards" from q intersects it at several momenta $p_i(q)$? It is natural to invoke the principle of superposition and extend the association

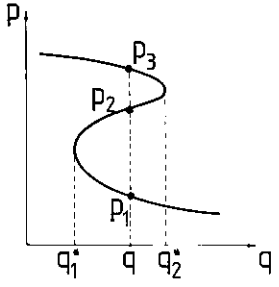


Fig. 28.

by writing $\psi(q)$ as a sum of terms like eq. (3.7), one for each branch $p_i(q)$. This raises the question of how these contributions are to be joined at caustics, where the projection of Σ "down" onto q is singular (e.g. at the points q_1^* and q_2^* on fig. 28). There are two difficulties. Firstly, it is not clear how to relate the phases of the different contributions. Secondly, eq. (3.7) has the undesirable property that it becomes infinite on caustics, because of the divergence of the "projection" Jacobian $\mathrm{d}Q/\mathrm{d}q$ [eq. (3.3)] determining the amplitude $a(q)$.

Both difficulties can be overcome by Maslov's ingenious procedure [26] of requiring the association between waves and phase-space surfaces to hold for momentum as well as position coordinates. Therefore the semiclassical momentum wave function $\bar{\psi}(p)$ will be given by a formula like eq. (3.7) with Σ specified by the function $q(p;P)$ and the generating function $\bar{S}(p;P)$ constructed from the line integral $\int q\cdot\mathrm{d}p$ rather than $\int p\cdot\mathrm{d}q$. The beauty of this method is that Σ is a smooth surface and so cannot have points where the projection is singular in both q and p. So when eq. (3.7) gives problems near q-caustics its momentum analogue is well-behaved. But $\psi(q)$ and $\bar{\psi}(p)$ are related by Fourier transformation, so that Maslov was able to employ the momentum analogue of eq. (3.7) to obtain representation of $\psi(q)$ in terms of an *oscillatory integral* which is well-behaved at caustics. I shall not discuss here the form of this wave close to caustics (this aspect was reviewed at Les Houches in 1980 [18] and by Berry and Upstill [33] under the heading "catastrophe optics") except to state that near a typical caustic (a smooth surface with dimensionality $N-1$ in q-space), eq. (3.7) is replaced by an Airy function. When q is not close to a caustic, the Fourier integral for $\psi(q)$ can be evaluated by the method of stationary phase, because when $\hbar\to 0$ the integrand is a rapidly-oscillating function of its variables p. The result is that the representation of ψ as a sum of terms like eq. (3.7) is regained, but with a definite phase relationship, which I now describe, between the terms corresponding to the branches that join on the caustic.

In constructing the correct superposition of terms (3.7), the phases S_i must be chosen to be the branches of a single action function $S(q)$ on Σ, so that the value of S on a caustic must be the same irrespective of the branch on which the caustic is approached. Corresponding to fig. 28, for example, S must be chosen as follows:

$$\text{branch 1: } S=\int_{q_1^*}^{q} p_1(q')\mathrm{d}q', \tag{3.8}$$

$$\text{branch 2: } S=\int_{q_1^*}^{q} p_2(q')\mathrm{d}q',$$

$$\text{branch 3: } S=-\int_{q}^{q_2^*} p_3(q')\mathrm{d}q'+\int_{q_1^*}^{q_2^*} p_2(q')\mathrm{d}q'. \qquad \text{(3.8 cont'd)}$$

In addition, the modulus signs must be removed from the determinantal amplitude factor, so that during passage through a typical caustic (smooth $N-1$ dimensional surface), where $|\mathrm{d}q|$ has a simple zero, the determinant changes phase by π and its square root by $\pi/2$. The only problem lies in deciding the sign of this change. The solution is that the phase change of ψ is $-\pi/2$ if the caustic is transversed with the (locally) two-sheeted region of Σ on the right. This is the well-known "phase advance of $\pi/2$ on crossing a caustic" [30]. Thus in fig. 28 the phase changes by $-\pi/2$ between p_3 and p_2 and by a further $+\pi/2$ (back to its original value) between p_2 and p_1.

The final step in generalizing the construction (3.7) is to consider the case where Σ is a closed surface in the form of an N-torus. This is not the only topology compatible with the Lagrangian conditions (2.7): for $N=2$, Hayden and Zeeman [34] have constructed a Lagrangian Klein bottle, but its quantal (or even classical) meaning is still obscure.

For an N-torus it is possible to return to any original point q in N essentially different ways, corresponding to the irreducible circuits γ_i of Σ. There is no global action function S which is single valued on Σ, because around γ_i there must be a change $\Delta_i S$ given by

$$\Delta_i S=\oint_{\gamma_i} \boldsymbol{p}\cdot\mathrm{d}\boldsymbol{q}, \qquad (3.9)$$

and equal to the sum of the areas of the projections of γ_i onto the N qp planes. During such a circuit, the phase of the wave constructed by eq. (3.7) will change by $\Delta_i S/\hbar$ plus a multiple α_i of $-\pi/2$ equal to the number of caustics encountered during γ_i. But the wave function $\psi(\boldsymbol{q})$ must be single valued under continuation, implying

$$\frac{1}{\hbar}\oint_{\gamma_i} \boldsymbol{p_i}\cdot\mathrm{d}\boldsymbol{q_i}-\tfrac{1}{2}\alpha_i\pi=2m_i\pi \qquad (1\leq i\leq N), \qquad (3.10)$$

i.e.

$$\frac{1}{2\pi}\oint_{\gamma_i} \boldsymbol{p_i}\cdot\mathrm{d}\boldsymbol{q_i}=(m_i+\tfrac{1}{4}\alpha_i)\hbar, \qquad (3.11)$$

where the m_i are integers. This is a set of N *quantum conditions* on the geometry of Σ. A typical torus will not satisfy these conditions, but recall that Σ is embedded in an N-parameter family labelled by $\boldsymbol{P}$, and if (as is true semiclassically) $\hbar$ is small compared with $\Delta_i S$ it will always be possible close to any given Σ to find a surface for which eq. (3.11) is satisfied and ψ is a single-valued semiclassical wave function. The numbers α_i, often called the "Maslov indices", depend on how Σ is embedded in phase space; a general topological discussion of α_i was given by Arnol'd [35] (see also refs. [30,1,2]). The Maslov indices embody the solution in N dimensions of what in one dimension is the "connection problem" [3,62] for WKB solutions of Schrödinger's equation.

The wave ψ thus obtained, based on a quantized torus Σ, will not in general be an eigenfunction of the Hamiltonian H, because it will change with time as Σ deforms under H. Only when Σ is an *invariant torus of motion* under H as described in section 2 will the surface remain fixed as its individual points q, p wind around it, and then ψ will be an energy eigenfunction, labelled by N quantum numbers m_i. The energy spectrum implied in this case by the conditions (3.11) will be studied in section 4.

It is obvious that this method of constructing semiclassical energy eigenfunctions can succeed only if tori exist, i.e. if the motion is regular as in integrable systems or throughout most of the phase space in quasi-integrable systems. For irregular and in particular ergodic motion, no tori exist and Maslov's method fails completely. At present there exists no asymptotic theory for the eigenfunctions corresponding to irregular motion. It is possible, however, to make conjectures in this case, on the basis of a general quantal phase-space formalism now to be described.

3.3. Wigner's function and the semiclassical eigenfunction hypothesis

In 1932, Wigner [38], introduced a phase-space distribution function $W(q,p)$ corresponding to a quantum state $\psi(q)$. This is defined by the N-fold integral

$$W(q,p)=\frac{1}{(2\pi\hbar)^N}\int\ldots\int \mathrm{d}^N X\,\exp(-\mathrm{i}p\cdot X/\hbar)\psi^*(q-X/2)\psi(q+X/2), \tag{3.12}$$

namely the Fourier transform of the product of ψ and ψ^* at positions separated by X. W is a quantal generalization of the classical density of

points in phase space. It is also possible to generalize other classical functions to get phase-space representations of quantal operators, and to generalize Liouville's equation to get a phase-space representation of the Schrödinger equation. Groenewold [39], Moyal [40], Takabayasi [41] and Baker [42] have shown that the resulting phase-space picture gives a complete representation of quantum mechanics, alternative to more familiar approaches in terms of wave functions, Hilbert space operators, or functional integrals. Wigner's picture is peculiarly well suited to the present problem, because it is in phase-space that the distinction between classical regular and irregular motion manifests itself most clearly, and so one can hope that the analogous quantal distinctions will reveal themselves with corresponding clarity in the form of $W(q,p)$. This idea was first expressed by Nordholm and Rice [43] and developed by Voros [44], Berry [45] and Berry and Balazs [46]. Other semiclassical aspects of Wigner's function were studied by Balazs and Zipfel [47], Heller [48] and Korsch [49].

Despite appearances, $W(q,p)$ as defined by eq. (3.12) has complete formal symmetry in q and p. It contains all the information about the quantum state. In particular, the coordinate probability density is obtained by projection "down" p onto q:

$$|\psi(q)|^2 = \int \ldots \int \mathrm{d}^N p \, W(q,p), \tag{3.13}$$

and the momentum probability density is obtained by projection "across" q onto p:

$$|\bar{\psi}(p)|^2 = \int \ldots \int \mathrm{d}^N q \, W(q,p). \tag{3.14}$$

W can also be employed more generally, to describe statistical mixtures of pure states ψ_i with weights a_i, by adding terms of the pure-state form (3.12), one for each state ψ_i in the mixture, with coefficients a_i, to get a phase-space representation of the quantal density matrix.

It is natural to ask what $W(q,p)$ looks like for a semiclassical state of the form (3.7), associated with a surface Σ. To find out, simply substitute eq. (3.7) into eq. (3.12), to get

$$W(q,p) = \frac{K^2}{(2\pi\hbar)^N} \int \ldots \int \mathrm{d}^N X \left| \det\left(\frac{\partial^2 S(q+X/2;P)}{\partial q_i \partial P_j} \right) \det\left(\frac{\partial^2 S(q-X/2;P)}{\partial q_i \partial P_j} \right) \right|^{1/2}$$

$$\times \exp\left(\frac{i}{\hbar}\left[\int_{q-X/2}^{q+X/2} \boldsymbol{p}(\boldsymbol{q}', \boldsymbol{P})\cdot \mathrm{d}\boldsymbol{q}' - \boldsymbol{p}\cdot\boldsymbol{X}\right]\right). \tag{3.15}$$

As $\hbar \to 0$, the integrand oscillates rapidly and is dominated by the region near $\boldsymbol{X}=0$. Expanding the phase for small $\boldsymbol{X}$ gives

$$\int_{q-X/2}^{q+X/2} \boldsymbol{p}(\boldsymbol{q}', \boldsymbol{P})\cdot \mathrm{d}\boldsymbol{q}' \approx \boldsymbol{X}\cdot\boldsymbol{p}(\boldsymbol{q}). \tag{3.16}$$

On setting $\boldsymbol{X}=0$ in the determinants, the integral can be evaluated to give the purely classical result

$$W(\boldsymbol{q},\boldsymbol{p}) \approx K^2 \left|\det \frac{\partial^2 S(\boldsymbol{q};\boldsymbol{P})}{\partial q_i \partial q_j}\right| \delta[\boldsymbol{p}-\boldsymbol{p}(\boldsymbol{q},\boldsymbol{P})]. \tag{3.17}$$

In this approximation, therefore, W is nonzero only on the surface Σ employed in the construction of ψ. This satisfying result can be expressed in a more illuminating way if we define Σ as the memeber $\boldsymbol{P}^*$ of the N-parameter family of surfaces labelled by $\boldsymbol{P}$, and use the fact that

$$\left|\det \frac{\partial^2 S(\boldsymbol{q};\boldsymbol{P})}{\partial q_i \partial P_j}\right| = \left|\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{P}}\right| \tag{3.18}$$

to change variables in the delta function in eq. (3.17). This gives

$$W(\boldsymbol{q},\boldsymbol{p}) \approx K^2 \delta[\boldsymbol{P}(\boldsymbol{q},\boldsymbol{p}) - \boldsymbol{P}^*], \tag{3.19}$$

where $\boldsymbol{P}(\boldsymbol{q},\boldsymbol{p})$ is the label of the particular surface that passes through $\boldsymbol{q}$, $\boldsymbol{p}$. It is clear from this representation that W is of uniform strength in the variable $\boldsymbol{Q}$ conjugate to $\boldsymbol{P}$.

It is possible to go much deeper into the asymptotics of W and give a full stationary-phase evaluation of eq. (3.15). This reveals ([45,17,50,46,119]) that in the semiclassical limit the purely classical W given by eq. (3.19) softens its delta function and develops an intricate fringe pattern (whose details depend on the geometry of Σ) in phase space near Σ. Moreover this more refined semiclassical W projects down onto q by eq. (3.13) to give a probability density with the correct non-diverging behaviour on caustics, and this approach leads to an infinite hierarchy of nonlinear identities between the caustic wave functions classified by catastrophe theory (Berry and Wright [51], summarized in ref. [18]). But for our present purposes the crude approximation eq. (3.19) is sufficient.

We wish to apply eq. (3.19) to the energy eigenstates of a system (integrable or quasi-integrable) for which some orbits trace out phase-space

tori. Such tori must satisfy the quantization condition (3.11) which according to eq. (2.4) restricts the action variables to

$$I_m = (m + \alpha/4)\hbar, \tag{3.20}$$

where $m \equiv \{m_1 \ldots m_N\}$ is the set of quantum numbers and $\alpha \equiv \{\alpha_1 \ldots \alpha_N\}$ are the Maslov indices. Then the Wigner function W_m representing the state labelled by m is given by eq. (3.19) as the following (correctly normalized) expression:

$$W_m(q,p) \approx \frac{1}{(2\pi)^N} \delta[I(q,p) - I_m], \tag{3.21}$$

where $I(q,p)$ is the action of the torus passing through q, p. The N-dimensional delta function implies that the Wigner function for an eigenstate is concentrated on the region that an orbit explores over infinite time – i.e. on the torus.

It is natural to extend this idea to cases where the motion is irregular and so not confined to tori. The resulting "semiclassical eigenfunction hypothesis" can be expressed as follows:

> Each semiclassical eigenstate has a Wigner function concentrated on the region explored by a typical orbit over infinite times.

The stipulation "typical" excludes the measure-zero closed orbits exploring one-dimension regions, which (except for certain degenerate cases) are too small to support quantum states (see also section 5.3). It was realized by Berry [52] and Voros [53] that this plausible hypothesis has powerful predictive force, as will now be explained.

First of all, let us apply the hypothesis to the extreme case of an ergodic system, whose orbits fill whole energy surfaces in phase space. Each quantum state corresponds to one energy surface, selected by a quantum condition. What these eigenenergies are is unknown, because nobody has so far discovered how to associate a wave with an energy surface in such a way that quantization follows from single-valuedness (the structure of the spectrum, and implicit quantum conditions for a particular ergodic system, will be discussed in sections 4 and 5). For an ergodic system, then, the hypothesis gives for the correctly normalized Wigner function representing an eigenstate with energy E,

$$W(q,p) \approx \frac{\delta[E - H(q,P)]}{\int \ldots \int \mathrm{d}^N q \; \mathrm{d}^N p \, \delta[E - H(q,p)]}. \tag{3.22}$$

In contrast to eq. (3.21), this is a one-dimensional delta function, reflecting the fact that W is spread over a much larger region of phase space.

The prediction of different morphologies for W is supported by computations of Hutchinson and Wyatt [54] for a Hamiltonian with $N=2$ corresponding to motion of a particle in a potential (of "Hénon-Heiles" type) giving predominantly regular motion at low energies and irregular motion at high energies.

3.4. Regular and irregular quantum states

By employing the "integrable" or "ergodic" Wigner functions (3.21) and (3.22), together with the definition (3.12), it is possible to obtain morphological information about two aspects of $\psi(q)$: its local average strength and its pattern of local oscillations.

Consider first the probability density $|\psi(q)|^2$, obtained according to eq. (3.13) by projecting $W(q,p)$ "down" p. For a system with tori, eq. (3.21) simply gives the particular case of eq. (3.3) appropriate for this form of Σ, namely (fig. 29) the sum over branches

$$|\psi(q)|^2 \approx \frac{1}{(2\pi)^N} \sum_i \left| \frac{\mathrm{d}\theta}{\mathrm{d}q}(q,p_i(q)) \right| . \tag{3.23}$$

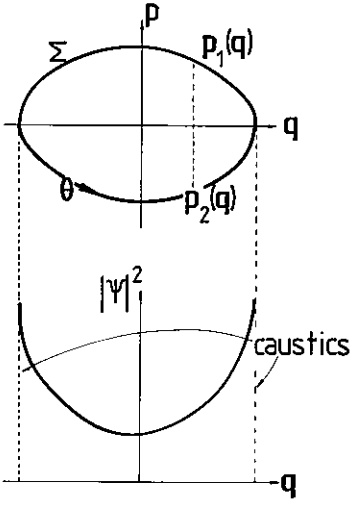


Fig. 29.

As already discussed, there are caustics at the singularities of the projection of the torus onto q, i.e. on local boundaries of the region explored by the orbit in q-space. Since the torus is usually locally parabolic near a caustic, we can extract the analytic form of the divergence of the approximate intensity $|\psi(q)|^2$ by taking a coordinate x perpendicular to the caustic and evaluating

$$|\psi(x)|^2 \sim \int \mathrm{d}p\,\delta(x-p^2) \sim \frac{\Theta(x)}{x^{1/2}}, \tag{3.24}$$

where Θ denotes the unit step function. In general the caustic has cusps, swallowtails, umbilics and other morphologies classified by catastrophe theory as singularities of projections [18,17] and becomes very complicated as N increases.

On the other hand, for an ergodic system (3.22) gives

$$|\psi(q)|^2 \approx \int \ldots \int \mathrm{d}^N p\,\delta[E-H(q,p)], \tag{3.25}$$

i.e. the projection of the energy surface. In ref. [52] I evaluated this for the case of a Hamiltonian

$$H(q,p) = \frac{|p|^2}{2\mu} + V(q), \tag{3.26}$$

with $V(q)$ corresponding to ergodic motion (e.g. the $N=2$ stadium billiard potential discussed in section 2.3), to get the result

$$|\psi(q)|^2 \sim [E-V(q)]^{N/2-1}\Theta[E-V(q)], \tag{3.27}$$

where the step function confines $|\psi|^2$ to the classically allowed part of q-space. In the trivial case $N=1$, when the ergodic system is also integrable, this reproduces the caustic divergence (3.24) on the classical boundary. But when $N>1$ the result shows that $|\psi(q)|^2$ does not diverge on the boundaries of the classical region: instead, it displays what I call *anticaustics*. Geometrically, this unexpected result is made plausible by a one-dimensional analogy in which instead of projecting a closed curve as in fig. 29 we project the patch of phase space enclosed by the curve (fig. 30). In section 6 we shall see evidence for the existence of anticaustics.

If the dynamical system consists of geodesic motion on a compact Riemannian manifold M, and is such that the geodesics are ergodic (e.g. if M has everywhere negative curvature), then there are no q-space boundaries. The semiclassical hypothesis predicts that for the corresponding

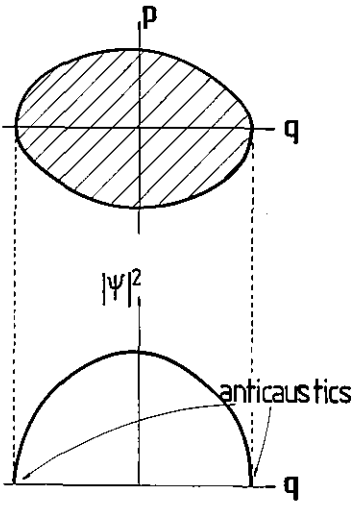


Fig. 30.

"quantal" problem, i.e. determination of eigenvalues and eigenfunctions $\psi_j(q)$ of the Laplace operator, $|\psi_j|^2$ should spread over the whole q space. That this is the case has been proved by Shnirelman [55], who shows that, for any smooth function $a(q)$,

$$\int_M \int dq\, a(q)|\psi_j(q)|^2 = \int_M \int dq\, a(q) / \int_M \int dq \qquad \text{as } j \to \infty \tag{3.28}$$

for almost all j, where j labels increasing eigenenergies.

At this point we should pause to ask exactly what sort of semiclassical |wave function|^2 is being calculated by this procedure involving the purely classical Wigner functions (3.21) and (3.22). After all, we know that the exact $|\psi|^2$ will have oscillations and moreover will behave smoothly across classical boundaries. The probable answer [which I cannot prove except for simple cases but which is supported in Shnirelman's [55] special case by the result (3.28)] is that what is being calculated is the limit as $\hbar \to 0$ of a probability density $|\psi|^2_{\text{sm}}$ which has been *smoothed* over a distance Δ in each direction in q space, where Δ vanishes as $\hbar \to 0$ more slowly than $\hbar$, so that oscillatory detail on the scale of the de Broglie wavelength is smoothed away. Mathematically, for any function $f(q)$ this procedure can be defined by

$$f(q) \to f_{\text{sm}}(q) \equiv \frac{1}{\Delta^N} \prod_{i=1}^{N} \int_{q_i - \Delta/2}^{q_i + \Delta/2} dQ_i f(Q), \tag{3.29}$$

where

$$\lim_{\hbar\to 0} \Delta = 0 \qquad \text{but } \lim_{\hbar\to 0} \frac{\hbar}{\Delta} = 0. \tag{3.30}$$

The vanishing of Δ ensures that the smoothed functions are "semiclassically sharp". On this interpretation, the classical Wigner functions (3.21) and (3.22) should be considered as the result W_{sm} of smoothing over q.

In view of this, it might be thought that eqs. (3.21) and (3.22) destroy all information about the oscillations of $\psi(q)$. But this is, surprisingly, not the case. Such information can in fact be obtained [52,53] from the *local autocorrelation function* of ψ, defined as

$$C(X;q) = [\psi(q+X/2)\ \psi^*(q-X/2)]_{\mathrm{sm}} / |\psi(q)|^2_{\mathrm{sm}}, \tag{3.31}$$

where the bars denote local semiclassical averaging as in (3.29). C contains information about the oscillations on the scale X near q, with irrelevant overall phases averaged away. For example, if (in one dimension),

$$\psi(q) = \exp\left[i\left(\frac{p_1 q}{\hbar} + \alpha\right)\right] \cos\left(\frac{p_2 q}{\hbar} + \beta\right) \tag{3.32}$$

then

$$|\psi(q)|^2_{\mathrm{sm}} = \tfrac{1}{2} \tag{3.33}$$

and

$$C(X) = \exp(\mathrm{i}p_1 X/\hbar)\cos(p_2 X/\hbar). \tag{3.34}$$

From the definition (3.12), C is given in terms of W_{sm} by

$$C(X;q) = \int \ldots \int \mathrm{d}^N p\, W_{\mathrm{sm}}(q,p) \exp(\mathrm{i}p\cdot X/\hbar) / |\psi(q)|^2_{\mathrm{sm}}. \tag{3.35}$$

For states associated with tori, eq. (3.21) gives

$$C(X;q) \sim \sum_i \left|\frac{\mathrm{d}\theta}{\mathrm{d}q}(q,p_i(q))\right| \exp\left(\frac{\mathrm{i}}{\hbar} p_i(q)\cdot X\right), \tag{3.36}$$

whose simple interpretation is of local de Broglie waves interfering at q. Each wave [of the type (3.7)] has a wave vector $p_i(q)/\hbar$ corresponding to one of the branches of Σ lying "over" q. For a torus there can be only a finite number of branches and so ψ is the local superposition of finitely many plane waves, which gives an anisotropic pattern of interference fringes.

Now contrast this with what happens for an ergodic system with Hamiltonian (3.26). The Wigner function (3.25) and formula (3.35) for the autocorrelation function give the following result, obtained in refs. [52] and [53]:

$$C(X;q) \approx \frac{\Gamma(\frac{1}{2}\mathrm{N}) J_{N/2-1}(|X|\{2\mu[E-V(q)]\}^{1/2}/\hbar)}{\{|X|\{2\mu[E-V(q)]\}^{1/2}/\hbar\}^{N/2-1}}, \tag{3.37}$$

where J denotes standard Bessel functions. Here, in contrast to eq. (3.36), C depends only on the length $|X|$ of the vector X, so that the local oscillations of ψ are, on the average, isotropic. This is because they result from the interference of de Broglie waves with the same wavelength but all possible directions, corresponding for the exploration by a trajectory of all momenta with the same energy $(p^2/2\mu)(q)+V(q)$, rather than the finite set $p_i(q)$ in the case of tori.

How can a function have locally isotropic autocorrelations about every point q? Only be being in some sense a *random* function of q. The randomness of $\psi(q)$ is governed by $W(q,p)$, which corresponds to a local spectral function specifying the distribution of local wave vectors $k=p/\hbar$ combining to interfere at q. For an ergodic Hamiltonian of type (3.26) the set of wave vectors consists of those with all possible directions but the same length – for $N=2$ this would be called a "ring spectrum".

To define the statistics of a random function ψ, it is necessary to specify more than the bilinear averages $|\psi|^2$ and C, in order to describe the fluctuations about these averages. In ref. [52] I suggested that the large number of interfering de Broglie waves would have random phases. This would make ψ a complex Gaussian random function of q, with local intensity fluctuations governed by the probability distribution

$$P(q,|\psi|^2) = (|\psi|^2_{\mathrm{sm}})^{-1} \exp(-|\psi|^2/|\psi|^2_{\mathrm{sm}}), \tag{3.38}$$

where $|\psi|^2_{\mathrm{sm}}$ is the "torus-projection" formula (3.27). A convenient measure of the fluctuations is the set of moments I_n, given by

$$I_n \equiv |\psi|^{2n}_{\mathrm{sm}} = n!. \tag{3.39}$$

These grow with n, but not as rapidly as moments associated with waves dominated by caustics, which (as I have shown for the optical case in a study [56,33] of the statistics of twinkling starlight) diverge as $\hbar \to 0$ according to power laws determined by catastrophe theory. Ozorio [50] has made a preliminary study suggesting that the Gaussian random nature of

a wave function is preserved under "metaplectic transformations", i.e. linear phase-space-transformations. For a detailed semiclassical study of the moments (3.39), see ref. [120].

Evidently the simple idea underlying the semiclassical eigenfunction hypothesis has led to dramatic predictions about the morphology of wave functions. It implies that as $\hbar \to 0$ wave functions separate into two universality classes, associated with regular and irregular classical motion. In the regular case, ψ is associated with tori, and has vivid anisotropic interference oscillations rising to high intensities $|\psi|^2$ on caustics. In the irregular case, ψ is associated with chaotic regions in phase space (e.g. the whole energy surface in the case of ergodic systems), and has a random pattern of oscillations (isotropic for ergodic systems) with anticaustics at classical boundaries. I emphasize that these universality classes are *emergent* properties as $\hbar \to 0$; away from the semiclassical limit, it may often be impossible to unambiguously categorize a state as being regular or irregular.

For the regular states, this description merely reformulates what is already known. For irregular states it predicts an unfamiliar structure. In section 6 I shall display some irregular wave functions in one dimension, generated by a time-dependent Hamiltonian. Here I show (fig. 31) the nodal lines of a high-lying eigenstate (the 157th) of the Laplace operator with the condition that ψ vanish on the boundary of a stadium (cf. fig. 10) and on its diameters as computed by McDonald and Kaufman [57]. Recall from section 2.3 that the stadium is classically ergodic, so that the quantal eigenstates should be irregular. And it is clear that the nodal lines do wander irregularly with no systematic well-defined direction. Compare this with fig. 32, showing nodal lines for an eigenfunction of a circle (with radial quantum number 9 and angular quantum number 28), which form a regular pattern as expected because the circle is an integrable billiard (cf.

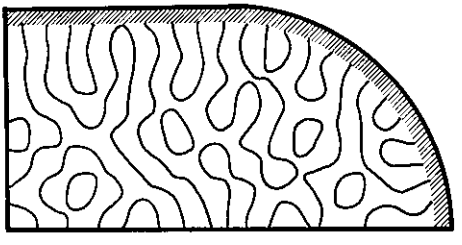


Fig. 31. Reprinted from ref. [57] with permission.

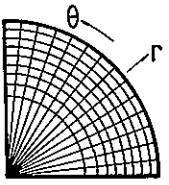


Fig. 32.

section 2.2). A similar contrast between the nodal patterns of states corresponding to regular and irregular classical motion was found by Stratt et al. [58]; their computations were for a smooth potential where some orbits filled neither tori nor whole energy surface, and indeed they found, roughly corresponding to these, states whose nodal pattern could not be classified as clearly regular or clearly irregular.

3.5. Crossing of nodal lines when $N=2$

The most striking difference between the nodal patterns for the irregular and regular wave function in figs. 31 and 32 is that in the irregular case the nodal lines do not intersect (except at the stadium boundary), while in the regular case they do. Is this a general property, which could be used to distinguish regular from irregular states in systems with $N=2$? The answer is no.

To see why, consider first the theorem published by Uhlenbeck [59] in 1976, which states that it is a generic property of eigenfunctions that their nodal lines do not intersect. "Generic" means here that the property fails to hold only for Hamiltonians forming a set of zero measure among all Hamiltonians. The idea is very simple. Suppose two nodal lines do cross. Near such a crossing the graph of a (real) eigenfunction $\psi(x, y)$ must be saddle-shaped, with contours in the form of hyperbolae near the nodal lines (it is easy to prove from the Schrödinger equation – with or without a potential but with isotropic dependence of H on the momentum components – that the nodal lines cross at right angles). At the crossing, not only ψ but its gradients $\partial\psi/\partial x$ and $\partial\psi/\partial y$ vanish. This gives three conditions on two coordinates xy and so generically no point can be found at which they are all satisfied. It could be objected that $\psi(x, y)$ is not a generic function because it is restricted by being a solution of the wave equation; but this relates ψ to its second derivatives and so does not spoil any genericity argument involving the gradient.

The form of this argument suggests (cf. ref. [57]) that node-crossing *will* occur generically in a family of Hamiltonians depending on a single parameter A, i.e. it is possible that for isolated values of A the nodes of a given eigenfunction may cross. If this is correct, then nodal lines behave differently from energy levels, whose crossing (degeneracy) requires two parameters, as we shall see in section 5.

There is no implication that "generic" here means "being associated with irregular classical motion", and indeed it cannot mean this, because the KAM theorem (section 2.4) establishes the persistence of tori in large classes of nonintegrable systems. Therefore even among regular states, associated with tori, the overwhelming majority will not show nodal line crossings.

So what is the special property allowing nodal lines to cross? It is not the existence of some tori throughout some but not all phase space, as in quasi-integrable systems. Nor is it the existence of tori throughout the whole phase space, as in integrable systems. In fact (as realized by Pechukas [60] in 1972) it is *separability* of the Schrödinger equation in an orthogonal coordinate system. This is not implied by integrability: even at the classical level the canonical transformation to action-angle variables does not imply that separation can be effected by coordinate transformation alone - see ref. [17] for a discussion of nonseparable tori.

If the system is separable, with orthogonal coordinates ζ, η, then eigenfunctions can be found with the special form

$$\psi(\zeta,\eta) = F(\zeta)\,G(\eta) \tag{3.40}$$

for which the nodal lines do cross. But, though separability allows nodal lines to cross, even this is not sufficient to *guarantee* that they do so. Because if the spectrum has *degeneracies* it is possible to add solutions of the form (3.40) to get new eigenfunctions whose nodal lines do not cross. Courant and Hilbert [61] give a dramatic example of this using the following eigenfunctions of a square with side π:

$$\psi_1(x,y) = \sin 2rx \sin y; \qquad \psi_2(x,y) = \sin x \sin 2ry. \tag{3.41}$$

The combination

$$\psi(x,y) = \psi_1(x,y) + (1+\varepsilon)\,\psi_2(x,y) \tag{3.42}$$

is also an eigenfunction. Figure 33a shows the nodal lines for $r = 12$ and $\varepsilon = 0$; there are 10 crossings, dividing the square into 12 regions. Figure 33b

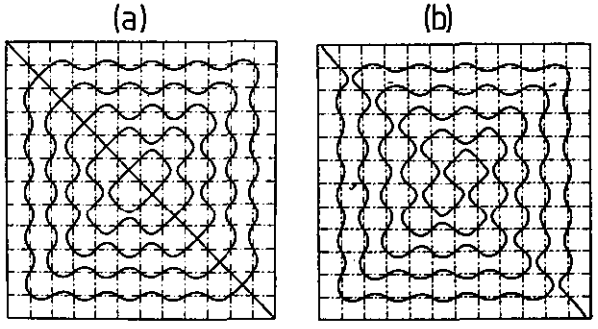


Fig. 33. Reprinted from ref. [61] with permission.

shows the nodal structure for a small value of ε; all the crossings have been destroyed, and the single nodal line divides the square into just two regions.

It is clear that the crossing of nodal lines is a very fragile property indeed, and certainly not as robust as the tori on which the existence of regular wave functions depend. Therefore although irregular wave functions must surely have no node crossings (except at isolated values of parameters in H), no implication about the nature of the associated classical motion can be drawn from the observation that nodal lines do not cross.

4. Eigenvalues: spectra on the finest scales

4.1. Quantum conditions

Having discussed the morphologies of semiclassical quantum states associated with regular and irregular motion, it is now necessary to consider how these classical distinctions affect quantal energy spectra. We shall find that these connections are quite subtle and depend on the scale on which the spectrum is studied. In my opinion the importance of understanding the rich variety of spectra on different scales (ranging from the locating of individual eigenvalues to the average density of states) has not hitherto been sufficiently appreciated.

On the finest scale, the most complete description of the spectrum would be explicit exact analytical formulae generating the energies E of all

eigenstates. Obviously, this ideal cannot be achieved except in very special cases, and so we seek semiclassical formulae, giving the energies with an accuracy that increases as $\hbar \to 0$. This more limited programme has been carried out only for quantum states associated with regular motion, i.e. with phase-space tori. The resulting semiclassical quantum condition is known by various names in different sects of physics or mathematics: Bohr, Sommerfeld, Einstein [36], Brillouin, Keller [28], Maslov [26]... I shall refer to the procedure simply as "torus quantization".

According to arguments given in sections 3.2 and 3.3, a quantal energy state labelled by quantum numbers $\boldsymbol{m} \equiv \{m_1 \ldots m_N\}$ is associated with a torus whose actions $\boldsymbol{I}_m$ are given by eq. (3.20) [cf. also eq. (3.11)]. The energy E_m of this state can be found in terms of the Hamiltonian (2.8) written as a function of the actions, and is simply

$$E_m = H(\boldsymbol{I}_m). \tag{4.1}$$

This explicit formulae is exact as $\hbar \to 0$. For fixed E this corresponds to highly excited states, but the formula often gives results accurate to a few per cent even for the ground state (e.g. when $N=1$, in which case (4.1) is the familiar WKB level formula [3]); in a few cases (harmonic oscillator, Coulomb potential, rectangular boxes) it gives all levels exactly. For integrable systems, the whole phase space is filled with tori and eq. (4.1) approximates all the levels (for applications to quantum billiards, see Keller and Rubinow [113]). For nonintegrable systems which are quasi-integrable (section 2.4), the KAM theorem guarantees the existence of tori filling some regions of phase space, and so eq. (4.1) can be employed to find a finite proportion of the levels, provided the actions and energies of the tori can be determined. This is a difficult problem of classical mechanics, which has been tackeld analytically by perturbation and iteration methods, and numerically by studying caustics and the Poincaré surface of section. By such techniques, Marcus and his colleagues [23,63–65], Percival and Pomphrey [66], Jaffé and Reinhardt [67], and Chapman, Garrett and Miller [68], obtained, for some chemically interesting quasi-integrable systems with $N=2$ and 3, eigenvalues in very good agreement with "exact" computations.

What determines energy levels if the classical motion is irregular? In quasi-integrable systems with irregular regions whose phase-space volume is not large in comparison with h^N (see refs. [45,6] and section 5.1), the levels can be found from eq. (4.1) using low-order perturbation theory to

approximate the motion in terms of tori even though the actual motion is not entirely confined to tori. But when large regions of phase space are filled by truly chaotic motion, torus quantization must fail. In the extreme case of an ergodic system, there are no tori, no actions and so no quantum numbers, and eq. (4.1) cannot give any of the levels. Nobody has yet succeeded in finding an explicit quantum condition, analogous to eq. (4.1), giving the levels of the "irregular spectrum" (to use Percival's terminology [37]) associated with chaotic trajectories.

It is, however, always possible to write *implicit* quantum conditions, by the conventional procedure of expanding the unknown eigenstate $|\psi>$ of the Hamiltonian $\hat{H}$ in terms of a complete set $|\phi_m>$ of basis states labelled by the N-fold index $\boldsymbol{m}$. Then Schrödinger's equation,

$$\hat{H}|\psi> = E|\psi>, \tag{4.2}$$

leads to the following secular determinant:

$$\det_{m,n}\{<\phi_m|\hat{H}|\phi_n> - E\delta_{m,n}\} = 0. \tag{4.3}$$

Although formally exact, this procedure, involving matrix elements labelled by $2N$ integers, is not suitable for understanding the semiclassical limit.

For general Hamiltonians I do not know how to improve on eq. (4.3), but for billiards with $N=2$ it is possible to derive a determinantal eigenvalue condition [69] which is much more compact in that each element is labelled by two integers instead of four. Moreover this new determinant is rapidly convergent and well suited to semiclassical studies of the spectrum. I shall illustrate the procedure by outlining its application to Sinai's billiard (section 2.3 and fig. 8); full details can be found in ref. [70]. The motivation for studying Sinai's billiard is that the classical motion is ergodic so that this is a system with no tori at all, so one knows from the outset that torus quantization is meaningless.

Because of periodicity with respect to the coordinate torus in Sinai's billiard, this system can be represented in terms of wave propagation amongst hard discs centred on the points $\varrho = \{\varrho_1, \varrho_2\}$ of the unit square lattice (fig. 34). Eigenvalues E are determined by Schrödinger's equation,

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + k^2\psi = 0, \qquad k^2 \equiv \frac{2\mu E}{\hbar^2}, \tag{4.4}$$

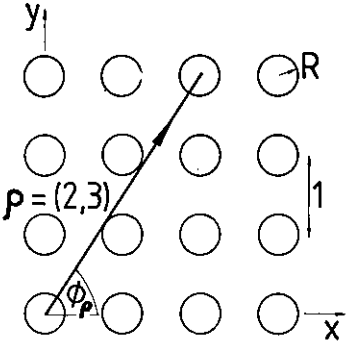


Fig. 34.

with boundary conditions

$$\begin{aligned} &\psi(r+\varrho)=\psi(r) \qquad \text{(torus periodicity)},\\ &\psi(|r|=R)=0 \qquad \text{(hard discs)}, \end{aligned} \tag{4.5}$$

where $r \equiv \{x, y\}$.

Thus the problem is transformed into one concerning waves in a periodic structure, and so the methods of solid-state electron band-structure theory can be brought to bear. Of several possible techniques (developed primarily for soft potentials and low energies) the one most suited to this problem (which has hard potentials and high – semiclassical – energies) is that devised by Korringa, Kohn and Rostoker (KKR) [71], where Green's identity plus periodicity is used to rewrite the Schrödinger equation as an integral equation round the boundary of one disc. In this equation the unknown function (normal derivative of ψ) is expanded as an angular Fourier series, i.e. in an angular momentum representation. After some (nontrivial) manipulation [70] this leads to the following determinantal condition for the energies E:

$$\det_{l,l'}[\delta_{ll'}+\sin\eta_l(E)\mathrm{e}^{\mathrm{i}\eta_l(E)}S_{l-l'}(E)]=0 \qquad (-\infty<l,l'<+\infty). \tag{4.6}$$

In this equation, η_l is the lth scattering phase shift from a single disc, given in terms of real Bessel functions J and Y [72] by

$$\tan\eta_l(E)=J_l(kR)/Y_l(kR) \tag{4.7}$$

and $S_{l-l'}$ are the "structure constants", given in terms of Hankel functions $H^{(1)}$ [72] by the lattice sum

$$S_l(E) = -\mathrm{i} \sum_{\varrho}{}' H_l^{(1)}(k\varrho)\mathrm{e}^{\mathrm{i}l\phi_\varrho}, \tag{4.8}$$

where the prime denotes exclusion of $\varrho = 0$ and where ϱ, ϕ_ϱ are the plane polar coordinates of ϱ (fig. 34).

In physical terms, eq. (4.6) is the condition that the wave scattered from a single disc interferes constructively with the waves scattered from and amongst all the other discs. It differs in two essential ways from the secular determinant (4.3): E appears non-linearly (as k in the Bessel functions) rather than linearly, and each matrix element is labelled by two integers rather than four.

The formally exact equation (4.6) involves an infinite determinant, but this can be truncated according to a simple semiclassical rule, as follows. When $\hbar$ is small, the argument kR of the Bessel functions in the phase shifts eq. (4.7) is large. Then as $|l|$ increases from zero the Bessel functions oscillate until $|l| \approx kR$. For $|l| > kR$, J_l gets exponentially small and Y_l exponentially large, so that the phase shifts can be set equal to zero and the determinant truncated at $l = kR$ [the full justification [70] of this procedure involves intricate asymptotics on $S_l(E)$]. Semiclassically, then, the size of the determinant is given by $l_{\max}$, defined as

$$l_{\max} = kR = \frac{2\mu E}{\hbar} = \frac{2\pi R}{\lambda}$$

$$= \frac{\text{perimeter of billiard boundary}}{\text{de Broglie wavelength of state being studied}}. \tag{4.9}$$

My opinion is that the determinant (4.6), truncated in this way and approximated by replacing the Bessel functions by their "Debye" asymptotic forms [72], is the semiclassical quantum condition for Sinai's billiard, analogous to the torus quantization rule (4.1). Of course eq. (4.6) is an implicit equation, and more complicated than eq. (4.1). Moreover the work of finding the levels increases as $\hbar \to 0$ (because $l_{\max} \to \infty$), but less rapidly than in any other method I know (a further discussion of this point will be given in section 5.4). A determinant analogous to eq. (4.6) can be written down [69] for any billiard (integrable, quasi-integrable or pseudointegrable as well as ergodic). Its size is always given by eq. (4.9).

4.2. Degeneracies

We shall learn that there is a great deal more structure in the spectrum than

can be immediately appreciated from a listing of the energies of all the eigenstates. This structure is embodied in correlations and clusterings of the levels. On the finest scale, such correlations concern neighbouring levels, and the first question to ask is: under what circumstances do these coincide? In other words: when do degeneracies occur?

If the Hamiltonian operator $\hat{H}$ has any symmetry, this may produce degeneracies, whose nature can be studied using group theory. It is not my intention to consider degeneracies of this type, and so when $\hat{H}$ does have symmetry I shall consider only states which all have the same symmetry class. This procedure is equivalent to considering all the states in a suitably "desymmetrized" Hamiltonian. For example, in the case of Sinai's billiard there will be degeneracies between states related by reflection about the axes $x=0$ or $y=0$ (fig. 8) or the diagonal $x=y$. We can eliminate these by considering, for example, only states which are odd under these reflections, and so effectively studying the modes of vibration of a membrane whose boundary is the billiard in fig. 35. Henceforth I shall consider all Hamiltonians as having been desymmetrized in this way.

For a typical (generic) such Hamiltonian it seems clear that degeneracy is infinitely improbable. But in a one-parameter family of Hamiltonians:

$$\hat{H}=H(\hat{q},\hat{p};A), \tag{4.10}$$

it might be expected that levels E can degenerate for isolated values of the parameter A. The reason would be that the jth level $E_j(A)$ is a curve in E, A space, and the crossing of curves in the plane (fig. 36) is a geometric occurrence which is stable under perturbation. But the surprising fact is that this picture is not correct: for typical systems with real eigenfunctions (the only ones considered here), it is necessary to vary two parameters, not one, in order to make two levels degenerate. This is the content of a theorem due originally to Von Neumann and Wigner [73] and Teller [74] and later

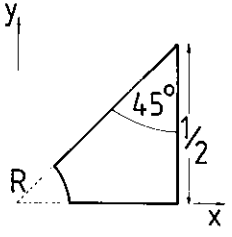


Fig. 35.

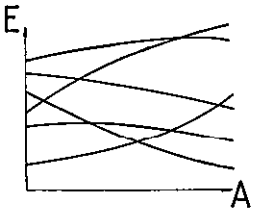


Fig. 36.

generalized by Arnol'd (ref. [5], appendix 10). The proof is based on a simple argument which I now give.

Consider Hamiltonians $\hat{H}$ close to a Hamiltonian $\hat{H}^*$ at which two orthogonal states $|u>$ and $|v>$ have the same energy E^*, i.e.

$$\hat{H}^*|u> = E^*|u>,\ \hat{H}^*|v> = E^*|v>,\ <u|v> = 0. \tag{4.11}$$

We wish to study the eigenvalues $|\psi>$ of $\hat{H}$, with energy E. Define

$$\hat{H} \equiv \hat{H}^* + \Delta\hat{H},\ E \equiv E^* + \Delta E, \tag{4.12}$$

and consider $\Delta\hat{H}$ and ΔE as small. Then the Schrödinger equation,

$$\hat{H}|\psi> = E|\psi>, \tag{4.13}$$

can be studied by degenerate perturbation theory. $|\psi>$ can be approximated as a normalized linear combination of $|u>$ and $|v>$, namely

$$|\psi> \approx \cos\chi|u> + \sin\chi|v>. \tag{4.14}$$

When substituted into eq. (4.13) using eqs. (4.11) and (4.12), this gives two homogeneous linear equations for $\cos\chi$ and $\sin\chi$, whose consistency gives a condition on the eigenvalue deviations $\Delta E_\pm$ which can be solved to give

$$\Delta E_\pm = \{<u|\Delta\hat{H}|u> + <v|\Delta\hat{H}|v> \pm [(<u|\Delta\hat{H}|u> - <v|\Delta\hat{H}|v>)^2 + 4(<u|\Delta\hat{H}|v>)^2]^{1/2}\}. \tag{4.15}$$

The discriminant is a sum of squares, so that coincidence of the eigenvalues ($\Delta E_+ = \Delta E_-$) requires the (real) matrix elements to satisfy two conditions and not one. Generically, this can be accomplished by varying two parameters in $\hat{H}$, as asserted.

Let the two parameters be A and B. Then eq. (4.15) implies that the connection of eigenvalues surface $E = E_\pm(A, B)$ in the space E, A, B takes the form of a double cone (diabolo) (fig. 37) whose sheets are joined at the "diabolical point" E^*, A^*, B^*, where A^*, B^* are the parameters for which

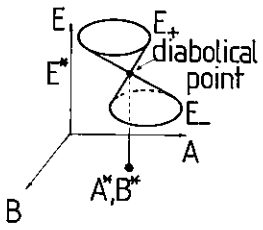


Fig. 37.

the degeneracy occurs. If only one parameter A is varied, the diabolical point will almost surely be missed and the curves $E_{\pm}(A)$ will avoid each other like branches of a hyperbola obtained by slicing the cone rather than crossing. So, instead of fig. 36 we expect curves like fig. 38, with near-degeneracies rather than actual crossings.

In testing the resulting picture of the spectrum, it is plausible to assume that a one-parameter family of non-symmetric classically ergodic systems will be generic in the quantal sense, and so will produce no level crossings. Such a family is the desymmetrized Sinai billiard (fig. 35) with the hard-disc radius R acting as the parameter. The levels $E_j(R)$ were computed "exactly" [70] using a simple modification of the determinant (4.6) which selects only states with the required symmetry. Figure 39 shows the spectrum. There are many near-degeneracies but careful examination shows that no two levels actually cross, so this test is successful.

The conclusion is that in a generic two-parameter family of Hamiltonian systems (e.g. Sinai's billiard with the discs replaced by ellipses of variable eccentricity and radius), we expect degeneracies at diabolical points A^*, B^* as in fig. 37. But how can be sure in any practical case that there really is a degeneracy at A^*, B^*, and not merely a near-touching of blunted cones, analogous to the near-crossing of curves in fig. 38? The answer lies in a little-known theorem, derived in ref. [70] by extending the algebra leading to eq. (4.15), concerning the two (real) eigenfunctions $\psi_{\pm}(q; A, B)$ whose energies degenerate at A^*, B^*: when taken round a circuit C in parameter

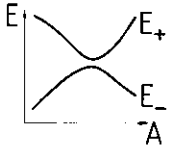


Fig. 38.

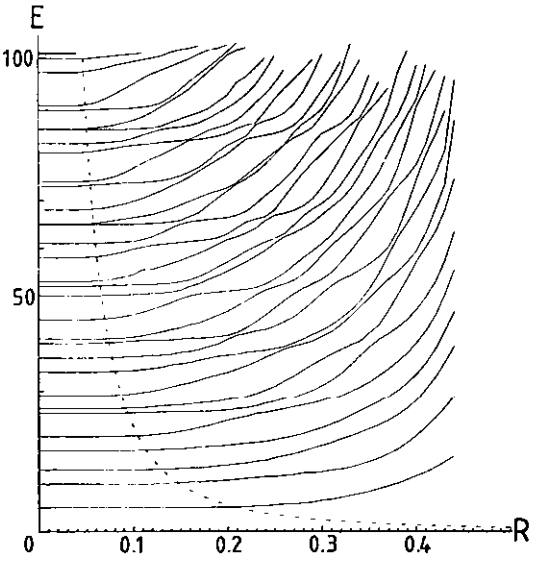


Fig. 39.

space, then, if (and only if) C encloses a degeneracy, ψ_+ and ψ_- will change sign. The theorem is mentioned by Arnol'd (appendix 10 of ref. [5]) and its converse was proved by Longuet-Higgins [75]. It is of course well known that wave functions must be single-valued under continuation with respect to their variables (e.g. q), but this theorem shows, surprisingly, that they need not be single-valued under continuation with respect to parameters in the Hamiltonian.

For systems whose levels are given by the torus quantization formula (4.1), degeneracies do not obey the generic two-parameter rule, but typically occur on varying only one parameter A. Therefore the picture of level crossings given by fig. 36, which we discarded in favour of fig. 38, is reinstated for the special class of systems with torus quantization. To see why this is so, regard the state with quantum numbers $\boldsymbol{m}^*$ as arising from the intersection of the $N-1$ dimensional hypersurface $E=H(\boldsymbol{I}_{\boldsymbol{m}^*};A)$, in the N-dimensional space of quantum numbers $\boldsymbol{m}$ (fig. 40), with the lattice point $\boldsymbol{m}^*$. Typically (i.e. for a fixed value of A) this hypersurface will not intersect any other lattice point, so the state $\boldsymbol{m}^*$ will be non-degenerate. But on varying A the hypersurface through $\boldsymbol{m}^*$ will smoothly change its orientation and there will typically be values A^* where it cuts another lattice

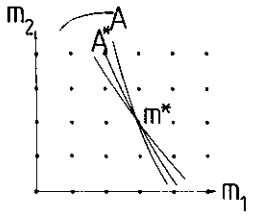


Fig. 40.

point, so that for such a value A^* there is a pair of degenerate states.

To summarize so far, we can say that, in the space of all Hamiltonians, those Hamiltonians with degenerate levels form a set of codimension two (connected cones), while in the space of Hamiltonians with all or part of the spectrum given by torus quantization, those Hamiltonians with degenerate levels form a set of codimension one (crossing curves). For the class of separable systems (all of which are integrable and so have torus-filled phase spaces), it is simple to show exactly that degenerate levels form a set of codimension one as torus quantization predicts (even through torus quantization is usually approximate rather than exact in these cases). But for nonseparable systems, even those which are integrable, the one-parameter degeneracies that torus quantization predicts will probably be split by the multi-dimensional analogue of barrier penetration, and I expect the exact spectra for this class of systems to have two-parameter degeneracies. On varying only one parameter, they should exhibit avoided crossings with exceedingly narrow separations, of order $\exp(-\text{constant}/\hbar)$, to be compared with the mean level separation which will be shown in section 5.1 to be of order $\hbar^N$. Semiclassically, if the spectrum is studied only to an accuracy described by power-law asymptotics in $\hbar$, such systems (i.e. all those with tori throughout regions of phase space), will seem to posses one-parameter degeneracies, but these are fragile and turn into two-parameter degeneracies when studied with exponential precision.

There are some special integrable systems whose levels are all proportional to integers, and these are not naturally embedded in continuous families. Their degeneracy structure can be very strange as I now indicate with two examples.

Consider first the equal-frequency harmonic oscillator with $N=2$, whose Hamiltonian is

$$H=\frac{p_1^2+p_2^2}{2\mu}+\tfrac{1}{2}\mu\omega^2(q_1^2+q_2^2). \tag{4.16}$$

The exact levels are

$$E_{m_1,m_2}=\hbar\omega(m_1+m_2+1) \qquad (0\le m_1, m_2<\infty). \tag{4.17}$$

By considering contours of constant E in the m_1, m_2 plane (fig. 41), we see that levels occur at energies

$$E=\hbar\omega p \qquad (p=1,2\ldots) \tag{4.18}$$

with degeneracies p. The mean density of states (cf. section 5.1) is

$$\bar{d}(E)=\overline{\sum_1^\infty \delta(E-\hbar\omega p)p}=\int_0^\infty \mathrm{d}p\; p\delta(E-\hbar\omega p)=\frac{E}{\hbar^2\omega^2}. \tag{4.19}$$

Therefore the mean spacing irrespective of degeneracies is of order $\hbar^2/E$, and in terms of this mean spacing the levels (4.18) arrive at ever-increasing intervals in ever-more-degenerate groups.

The second example is the 45° right triangle billiard with hypotenuse $2^{-1/2}$, i.e. the desymmetrized Sinai billiard (fig. 35) with $R=0$. This is a separable system whose exact levels are

$$E_{m_1,m_2}\equiv\frac{2\pi^2\hbar^2}{\mu}\mathscr{E}_{m_1,m_2}=\frac{2\pi^2\hbar^2}{\mu}(m_1^2+m_2^2) \qquad (1\le m_1<m_2<\infty). \tag{4.20}$$

Thus the levels $\mathscr{E}$ are all those integers which can be written as the sum of two squares. For small m_1, m_2, this square decomposition can be carried out in only one way. But when $\mathscr{E}=65$ $(=7^2+4^2=8^2+1^2)$ the first degeneracy occurs (it can be seen in fig. 39 for $R=0$). When $\mathscr{E}=325$ $(=15^2+10^2=18^2+1^2=17^2+6^2)$ the first triple degeneracy occurs. It seems that degeneracies are rare. But as $\mathscr{E}$ increases, they come to dominate

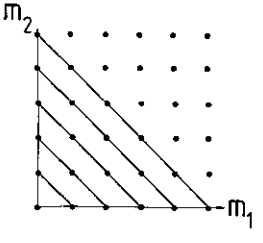


Fig. 41.

the spectrum. This curious conclusion, obtained in ref. [70], follows from two facts. Firstly, the average density of states (in $\mathscr{E}$) can be found from eq. (4.20) by counting lattice points, or from formulae to be derived in section 5.1, to be

$$\bar{d}(\mathscr{E}) = \pi/8, \tag{4.21}$$

so that the asymptotic mean level spacing irrespective of degeneracy is $8/\pi$. Secondly, it is a number-theoretic result that for large integers $\mathscr{E}$ the probability that $\mathscr{E}$ can be written as the sum of two squares decreases as $(\ln \mathscr{E})^{-1/2}$, so that the levels are separated by widening gaps of size $\sim (\ln \mathscr{E})^{1/2}$. These two facts are consistent only if the degeneracies increase as $(\ln \mathscr{E})^{1/2}$ to keep the net level density constant. In the original energy units, the resulting semiclassical spectrum has levels with degeneracies of order $[\ln(\text{constant} \times E/\hbar^2)]^{1/2}$ separated by gaps of order $\hbar^2[\ln(\text{constant} \times E/\hbar^2)]^{1/2}$. Pinsky [76] has found similar strange behaviour for the levels of the equilateral triangle billiard.

It is instructive to see how pseudointegrable systems fit into this general picture of degeneracy structure. Recall from section 2.5 that these are delicate "marginal" systems with $N=2$ and two constants of motion, whose nonintegrability consists in the fact that the two-dimensional phase-space surfaces that their orbits fill are not tori but multiply-handled spheres.

To study these we consider the square torus billiard of fig. 15, parameterized by the side length L of the reflecting square. Quantum-mechanically, this was desymmetrized so as to become effectively the billiard shown in fig. 42. Richens and Berry [21] computed some of the levels $E_j(L)$ "exactly", with results shown in fig. 43. There are no degeneracies except at $L=0$, suggesting that all the levels avoid each other which would mean that this class of systems behave like typical Hamiltonian systems and require two parameters to produce degeneracy. It can be

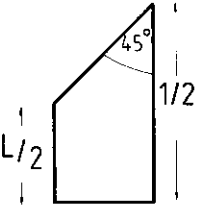


Fig. 42.

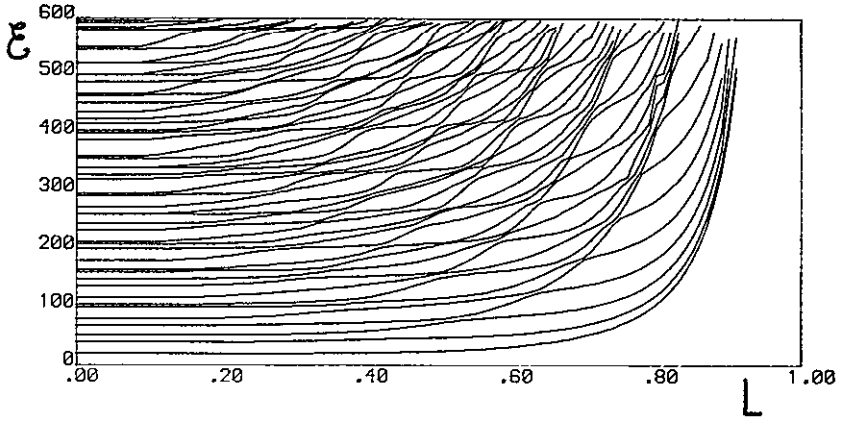


Fig. 43.

shown [21] that this is not quite correct: for every rational value of L there is a subset of all the levels (which gets smaller as the order of rationality increases) for which an exact torus-like quantization formula can be found, and which exhibit the increasing number-degeneracies discussed in connection with the 45° right triangle. None of these degeneracies (except for three at $L=0$) appears in fig. 43 because the energy range is not large enough.

Marcus [77] suggests that the presence of many near-degeneracies (overlapping avoided crossings), in curves of energy levels as functions of a single parameter, is an indication that the corresponding quantum states are irregular in the sense described in section 3.4, i.e. associated with chaotic classical orbits filling large fractions of the $2N-1$ dimensional energy surface. Figure 43 shows this implication to be unjustified: there are many near-degeneracies, but for the square torus billiard all orbits lie on $N-$(i.e. $2-$) dimensional surfaces and explore them without exponential chaos. Marcus's further suggestion that the implication holds in the reverse direction, i.e. that the energy levels of quantum states associated with classically chaotic motion will show many avoided crossings, is almost certainly correct. I conclude by drawing attention again to the extreme fragility of one-parameter crossings; these are analogous to the nodal-line intersections of wave functions (section 3.5), in that any departure from separability (which need not amount to classical chaos or even non-integrability) may destroy them. For a detailed study of diabolical points in a quantum billiard system, see ref. [69].

4.3. Level spacings

The next step in understanding spectra is to study their finescale texture as embodied by the distribution of spacings between neighbouring levels, that is, to study spectra on the scale of the mean level spacing, which as we shall see in section 5.1 is of order $\hbar^N$. The spacings distribution $P(S)$ is defined by

$$P(S)\mathrm{d}S = \text{probability that the spacings of a randomly chosen pair of neighbouring levels will between } S \text{ and } S+\mathrm{d}S, \tag{4.22}$$

where S is measured as a fraction of the mean level spacing $[\bar{d}(E)]^{-1}$ at the energy considered, so that the jth spacing S_j is defined as

$$S_j \equiv \bar{d}(E_j)(E_{j+1} - E_j). \tag{4.23}$$

The function $P(S)$ was devised by Wigner and Landau and Smorodinsky (see e.g. ref. [78] in nuclear physics, where the many-body levels of nuclei are modelled by the eigenvalues of matrices in an ensemble (e.g. all matrices whose elements are Gaussian-distributed). How can an ensemble be defined in the case of a single Hamiltonian, for which the levels are deterministic rather than random? The answer is: by taking the semiclassical limit $\hbar \to 0$, so that infinitely many levels lie near any given E and eq. (4.22) is the average over their spacings. For billiards this procedure is equivalent to averaging over all levels in the spectrum with $\hbar$ fixed, i.e.

$$P(S) = \lim_{J\to\infty} \frac{1}{J} \sum_{j}^{j=J} \delta(S - S_j). \tag{4.24}$$

We shall study the behaviour of $P(S)$ as $S \to 0$, because this tells us about the finest scales of level *clustering*. If $P(S) \to 0$ as $S \to 0$, neighbouring levels can be considered to "repel" each other, leading to a degree of regularity in the arrangement of levels, which can be quantified by the manner in which $P(S)$ vanishes. If on the other hand $P(S) \to$ constant as $S \to 0$, neighbouring levels cluster rather than repel.

The form of $P(S)$ as $S \to 0$ for a given Hamiltonian H depends on the degeneracy structure of families of similar Hamiltonians in which H can be embedded. In the generic case, we saw in the last section how H can be embedded in a family with two parameters A, B which has degeneracies at

diabolical points in E, A, B space as in fig. 37. Let the actual Hamiltonian under study have parameters A_0, B_0. Then the line with $A = A_0$, $B = B_0$ (fig. 44) will thread its way among cones in E, A, B space, and the cones will presumably be distributed thickly if $\hbar$ is small. In eq. (4.24) the sum over j corresponds to an "energy average"

$$P(S) = \overline{\delta[S - S_j(A_0, B_0)]} \tag{4.25}$$

over the spacings S_j "above" A_0, B_0. For small S the only contributions to the average come from spacings for which the line in fig. 44 passes near diabolical points (labelled by k) whose parameters A_k^*, B_k^* lie close to A_0, B_0. The corresponding spacing S_k takes the positive-definite "conical" form [cf. eq. (4.15) with the matrix elements expanded in $A - A^*$, $B - B^*$] as follows:

$$S_k(A_0, B_0) =$$
$$[a_k(A_0 - A_k^*)^2 + 2b_k(A_0 - A_k^*)(B_0 - B_k^*) + c_k(B_0 - B_k^*)^2]^{1/2}, \tag{4.26}$$

where a_k, b_k and c_k describe the geometry of the kth cone.

Now, by hypothesis, there is nothing special about the system with parameters A_0, B_0, and so the energy average in eq. (4.25) can be augmented by an ensemble average over a region of A, B near A_0, B_0. Let $\varrho(A, B, E)$ be the density of diabolical points, and let $\pi(a, b, c)$ be the probability distribution of cone geometry parameters; the forms of the functions ϱ and π are unknown. Then the average (4.25) becomes, on using eq. (4.26) because S is small,

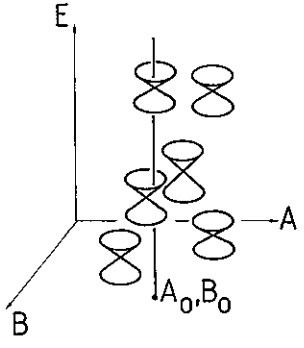


Fig. 44.

$$P(S) = \frac{\varrho}{\bar{d}(E)}(A_0, B_0, E) \int \mathrm{d}a \int \mathrm{d}b \int \mathrm{d}c \; \pi(a, b, c)$$
$$\int \mathrm{d}A \int \mathrm{d}B \; \delta[S - (aA^2 + 2bAB + cB^2)^{1/2}]. \tag{4.27}$$

The integral over A, B can be extended to infinity because the positive definiteness of the quadratic form ensures convergence. Now change variables to

$$\alpha \equiv A/S, \qquad \beta \equiv B/S, \tag{4.28}$$

and eq. (4.27) becomes

$$P(S) = \left(\frac{\varrho(A_0, B_0, E)}{\bar{d}(E)} \int \mathrm{d}a \int \mathrm{d}b \int \mathrm{d}c \; \pi(a, b, c) \right.$$
$$\left. \int \mathrm{d}\alpha \, \mathrm{d}\beta \; \delta[1 - (a\alpha^2 + 2b\alpha\beta + c\beta^2)^{1/2}] \right) \times S. \tag{4.29}$$

The factor in the two-line round brackets is a purely geometric average not involving S, so that

$$P(S) \sim S \qquad \text{as } S \to 0. \tag{4.30}$$

This argument therefore leads us to expect level repulsion for generic systems. The essential feature of the degeneracy structure which gives rise to this conclusion is the fact that the total length of intersection of a diabolo by two parallel sheets separated by S, where the diabolical point lies between the sheets, is proportional to S as $S \to 0$.

Figure 45 shows a test of this prediction of "linear" level repulsion, obtained by calculating eigenvalues of the desymmetrized Sinai billiard (fig. 42) and making a histogram of all spacings (several hundred in all) with $0.20 \leq R \leq 0.44$. It is clear that the levels to repel, and that the linear law gives a good fit. McDonald and Kaufman [57] and Casati, Valz-Griz and Guarneri [79], in computations of $P(S)$ for the desymmetrized stadium billiard, also obtain level repulsion, but their histograms do not show sufficient resolution to say whether the linear law is obeyed. Figure 46 shows another test, this time for the more delicate case of the (pseudo-integrable) desymmetrized square torus billiard (fig. 42). Once again the linear law of level repulsion gives a good fit, and it appears that pseudointegrable systems behave generically as far as $P(S)$ is concerned. (Because fig. 46 is

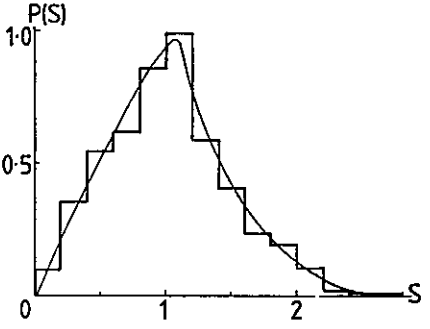


Fig. 45.

a compilation for different values of L, there might be a delta-peak in $P(S)$ at $S=0$ because of the increasing number-theoretic degeneracies for each rational L, discussed in section 4.2, but it is probable that this delta-peak has zero height because of the zero measure of rationals.)

Zaslavsky [80] predicts that for classically chaotic systems $P(S)$ will display not the limiting form (4.30) but a nonlinear repulsion S^γ where γ depends on the rate of exponential separation of trajectories. I have criticised Zaslavsky's argument elsewhere [70], but point out that my own argument leading to linear repulsion is not watertight: it could fail if the cone-

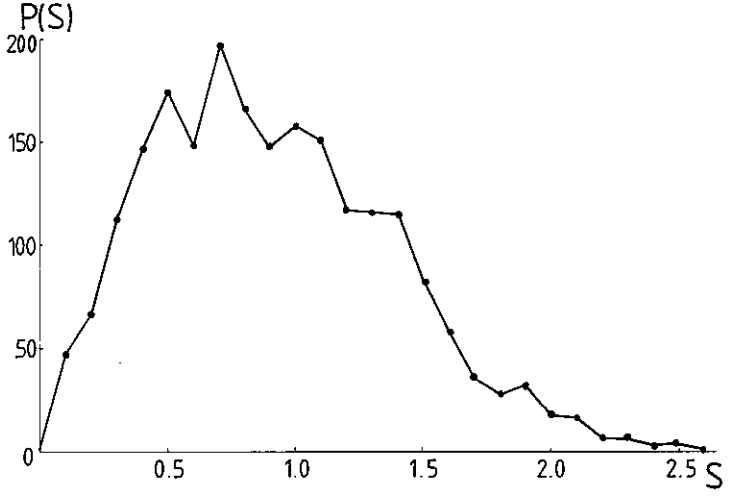


Fig. 46.

shape distribution $\pi(a, b, c)$ becomes singular as $\hbar \to 0$ in such a way that the previously mentioned two-line geometric factor in eq. (4.29) diverges.

Now let us consider $P(S)$ for some classes of non-generic system. Suppose we have a system so special that when embedded in an ensemble of similar systems it exhibits m-parameter degeneracies, i.e. degeneracies with codimension m, where m differs from the generic value of two. Then an argument precisely analogous to that based on cones when $m = 2$, employing the fact that S as a function of parameters is given not by eq. (4.26) but by the square root of a quadratic form in m variables, leads instead of eq. (4.30), to

$$P(S) \sim S^{m-1} \qquad \text{as } S \to 0. \tag{4.31}$$

For systems with torus quantization we have seen in eq. (4.2) that degeneracies are produced by varying only one parameter, so $m = 1$ and eq. (4.31) predicts that $P(S) \to$ constant as $S \to 0$. Therefore systems with tori should show level clustering rather than level repulsion. This is consistent with a more elaborate argument by Berry and Tabor [81], indicating that for integrable systems (with $N > 1$), $P(S)$ generically has the universal form

$$P(S) = \mathrm{e}^{-S}, \tag{4.32}$$

corresponding to levels arriving irregularly. Figure 47 shows tests of this prediction for (a) the spectrum of a rectangle with side ratio 2, and (b) a

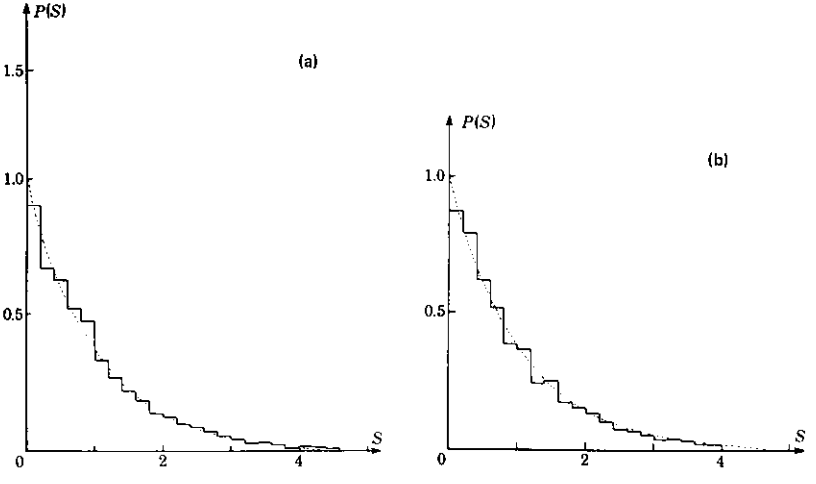


Fig. 47.

two-dimensional system whose potential is a square well in one direction and a harmonic oscillator in the other. Evidently the negative exponential is a very good fit to the computed histograms.

To summarize, we have found that for a generic system, e.g. one whose classical motion is irregular, the quantum levels exhibit linear repulsion and so are fairly regularly distributed. For a typical system with tori, whose classical motion is regular, the energy levels are irregularly distributed [and indeed from eq. (4.32) arrive at random as in a Poisson process such as radioactive decay].

I conclude with two curiosities. Consider first a class of systems in which degeneracies are strictly forbidden, so that even infinitely many parameters will not produce them. Then we must take $m \to \infty$, and eq. (4.31) suggests that $P(S)$ will vanish faster than any power of S as $S \to 0$. One such class of system is the eigenstates of the one-dimensional Schrödinger equation in a finite range with vanishing wave function at the ends. And indeed Pokrovski [82] showed that for electron states in a disordered linear chain of N delta-function potentials,

$$P(S) \sim \exp(-C/S^2)/S^2 \qquad \text{as } S \to 0, \tag{4.33}$$

where C depends on N but not S) in conformity with this prediction. (Recently, Molcanov [83], in a rigorous analysis, obtained the exponential distribution (4.32) for a one-dimensional potential – in the form of a non-differentiable ("fractal") curve derived from Brownian motion - in apparent contradiction with the prediction. But an essential feature of Molcanov's argument is the taking of a limit in which the potential exists over an infinite range, so that localized wave functions separated by great distances can have essentially independent eigenvalues and the basis of the prediction of level repulsion breaks down).

The second curiosity concerns two-dimensional harmonic oscillators with frequencies ω_1 and ω_2. Recall fig. 41 and eqs. (4.18) and (4.19), which show that if $\omega_1 = \omega_2$ degeneracies increase with energy, so that on the scale $\bar{d}^{-1}$ of the mean level spacing the gaps between neighbouring levels increase. Therefore the function $P(S)$ does not exist: the levels never settle down to a limiting distribution. A similar result holds for all rational values of ω_1/ω_2. What about irrational values? These are classically integrable systems and so might be expected to obey the negative-exponential rule (4.32). But for irrational harmonic oscillators the basis of the rule breaks down, because it can be shown quite easily that the levels, given by

$$E_{m_1 m_2} = \hbar[\omega_1(m_1 + 1/2) + \omega_2(m_2 + 1/2)], \tag{4.34}$$

are never degenerate. A lengthy number-theoretic analysis [81] of ways levels can come close together strongly suggests that $P(S)$ should show level repulsion, and indeed computation confirms this, as fig. 48 (for $\omega_1/\omega_2 = 2^{1/2}$) shows. The decay as $S \to 0$ is faster than for generic systems (cf. figs. 45 and 46), and suggest that $P(S)$ is more like eq. (4.33) – i.e. vanishing faster than any power of S – than the merely linear repulsion (4.30).

5. Eigenvalues: spectra on larger scales

5.1. Mean level density

Now we study spectra on scales large compared with the mean level spacing. It will be convenient to work with two functions: the level density $d(E)$ and the *mode number* $\mathcal{N}(E)$, defined in terms of the Hamiltonian operator $\hat{H}$ or the levels $E_1 \ldots E_j \ldots$ (in increasing E with degenerate states counted separately) by

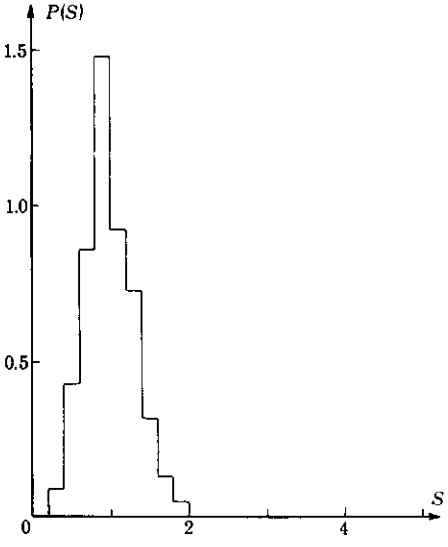


Fig. 48.

$$d(E) \equiv \sum_{j=1}^{\infty} \delta(E - E_j) = \mathrm{Tr}\,\delta(E - \hat{H}) \tag{5.1}$$

and

$$\mathcal{N}(E) \equiv \sum_{j=1}^{\infty} \Theta(E - E_j) = \mathrm{Tr}\,\Theta(E - \hat{H}). \tag{5.2}$$

These functions have of course been chosen so that their singularities (spikes for d and steps for $\mathcal{N}$) give the positions of the levels. $\mathcal{N}(E)$ is simply the number of states with energies below E. Obviously,

$$d(E) = \mathrm{d}\mathcal{N}(E)/\mathrm{d}E. \tag{5.3}$$

Sometimes $\mathcal{N}(E)$ is the more convenient function to work with, and sometimes $d(E)$ is more convenient.

There is a beautiful semiclassical theory for d and $\mathcal{N}$, whose principal architects were Gutzwiller [84–88] and Balian and Bloch [89–92]. This is based on representing $d(E)$ in the form

$$d(E) = \bar{d}(E) + d_{\mathrm{osc}}(E), \tag{5.4}$$

where $\bar{d}$ is the mean level density and d_{osc} is a series of oscillatory corrections. There is a similar representation for $\mathcal{N}(E)$. The terms on the right side of eq. (5.4) correspond to successive *smoothings* of the singular function $d(E)$. On the coarsest scale, that is after smoothing over energy ranges ΔE large enough to obliterate all traces of individual levels and all scales of level clustering, only $\bar{d}(E)$ survives. As ΔE is made smaller, more and more terms $d_{\mathrm{osc}}(E)$ contribute, with faster oscillations, until eventually they sum up to give a series of delta functions at the exact level positions. In simple terms, the semiclassical representation (5.4) is a generalization of the following expansion for a series of equally spaced unit delta functions:

$$\sum_{n=-\infty}^{\infty} \delta(E - n) = 1 + 2\sum_{m=1}^{\infty} \cos 2\pi m E. \tag{5.5}$$

We begin in this section by studying the mean level density $\bar{d}(E)$ and the mean mode number $\bar{\mathcal{N}}(E)$. These are given by the simple semiclassical rule (Landau and Lifshitz [93]) that each quantum state is associated with a phase-space volume h^n. Therefore $\bar{\mathcal{N}}(E)$ is h^{-N} times the volume of phase space for which $H(q,p)$ is less than E, i.e.

$$\bar{\mathcal{N}}(E) \approx \frac{1}{h^n} \int \mathrm{d}q \int \mathrm{d}p\, \Theta[E - H(q,p)], \tag{5.6}$$

and

$$\bar{d}(E) \approx \frac{1}{h^N} \int \mathrm{d}q \int \mathrm{d}p\, \delta[E - H(q,p)], \tag{5.7}$$

so that the density is proportional to the size of the energy surface.

It is instructive to obtain these formulae in several different ways. The simplest is via the Weyl association [39–42] of a phase-space function $a_w(q,p)$ with any quantal operator $\hat{a}$ which is a function of the $\hat{q}$ and $\hat{p}$ operators:

$$a_w(q,p) = \frac{1}{h^N} \mathrm{Tr}\left(\hat{a} \int \mathrm{d}Q \int \mathrm{d}\Pi \exp\left\{ \frac{\mathrm{i}}{\hbar} \left[(\hat{p} - p) \cdot Q + (\hat{q} - q) \cdot \Pi \right] \right\} \right), \tag{5.8}$$

and implies

$$\mathrm{Tr}\,\hat{a} = \frac{1}{h^N} \int \mathrm{d}q \int \mathrm{d}p\; a_w(q,p). \tag{5.9}$$

It is not hard to verify that if $\hat{a}$ is the sum of an operator depending only on $\hat{q}$ and an operator depending only on $\hat{p}$, $a_w(q,p)$ is the function obtained by replacing $\hat{q}$ q and $\hat{p}$ by p , but that in general a_w $(q, p$ does not equal the classical function obtained in this way. But, by the correspondence principle this classical replacement must hold *in the semiclassical limit*. We therefore choose for $\hat{a}$ the operator in eq. (5.1), namely

$$\hat{a} = \delta(E - \hat{H}), \tag{5.10}$$

so that semiclassically a_w is simply obtained by replacing $\hat{H}$ by the classical Hamiltonian function, and eq. (5.7) follows at once.

The second method, which we shall also use to study d_{osc} in section 5.2, is based on writing eq. (5.1) as

$$\begin{aligned} d(E) &= -\frac{1}{\pi} \lim_{\varepsilon \to 0} \mathrm{Im}\, \mathrm{Tr}\left(\frac{1}{E + \mathrm{i}\varepsilon - \hat{H}} \right) \\ &= -\frac{1}{\pi} \mathrm{Im} \int \mathrm{d}q\, G^+(q, q'; E)_{q=q'}, \end{aligned} \tag{5.11}$$

where G^+ is the outgoing energy-independent *Green function*, defined by

$$G^+(q, q'; E) \equiv \left\langle q' \left| \frac{1}{E + \mathrm{i}\varepsilon - \hat{H}} \right| q \right\rangle \tag{5.12}$$

and satisfying

$$[E+\mathrm{i}\varepsilon-H(q,-\mathrm{i}\hbar\nabla_q)]G^+(q,q')=\delta(q-q'). \tag{5.13}$$

For simplicity let us restrict ourselves to Hamiltonians of the "kinetic + potential" type (3.26). Then the Green function satisfies a Helmholtz equation

$$\left(\nabla_q^2+\frac{p^2(q)}{\hbar^2}\right)G^+(q,q')=\frac{2\mu}{\hbar^2}\delta(q-q') \tag{5.14}$$

with varying local momentum

$$p^2(q)=2[E-V(q)]. \tag{5.15}$$

Let

$$\varepsilon=|q-q'|. \tag{5.16}$$

Then from eq. (5.11), $d(E)$ depends on G as $\varepsilon\to 0$. The mean density $\bar{d}$ can be obtained by a local approximation in which eq. (5.14) is solved as though $p(q)$ had everywhere the value it has at q. The resulting outgoing "free-particle Green function with momentum $p(q)$" *is*

$$G^+(q,q';E)_{q\approx q'}\approx\frac{-2\mathrm{i}\mu}{(2\pi)^{N/2+1}}\left(\frac{p(q)}{\varepsilon}\right)^{N/2-1}$$
$$H^{(1)}_{N/2-1}\left(\frac{p(q)\varepsilon}{\hbar}\right), \tag{5.17}$$

where $H^{(1)}$ is a Hankel function [72]. This diverges as $\varepsilon\to 0$ but its imaginary part remains finite. The small-argument limiting form for the Hankel functions [72], together with eq. (5.11), now gives $\bar{d}$ as an integral over the classically accessible space, namely

$$\bar{d}(E)\approx\left(\frac{\mu}{2\pi\hbar^2}\right)^{N/2}\frac{1}{\Gamma(N/2)}\int \mathrm{d}q\,[E-V(q)]^{N/2-1}\Theta(q)]. \tag{5.18}$$

This is precisely what eq. (5.7) gives for Hamiltonian of type (3.26) (after integrating away the quadratic p-dependence).

For billiard Hamiltonians with $N=2$, whose boundary encloses area $\mathscr{A}$ eqs. (5.6) and (5.7) give

$$\bar{N}(E)\approx\frac{\mathscr{A}\mu E}{2\pi\hbar^2}=\frac{\mathscr{A}k^2}{4\pi} \tag{5.19}$$

and

$$\bar{d}(E) \approx \frac{\mathscr{A}\mu}{2\pi\hbar^2}. \tag{5.20}$$

These results are known as "Weyl formulae" (for a review, see Baltes and Hilf [94]).

Of course it is also possible to obtain $\bar{d}(E)$ from the quantum conditions discussed in section 4.1. For integrable systems, the torus quantization rule (4.1), together with eq. (5.1), gives

$$d(E) = \sum_m \delta[E - H(I_m)]. \tag{5.21}$$

To find $\bar{d}$ we simply replace the sum by an integral and change variables from $\boldsymbol{m}$ to $\boldsymbol{I}$ using eq. (3.20), giving

$$\bar{d}(E) \approx \frac{1}{\hbar^N} \int \mathrm{d}\boldsymbol{I}\, \delta[E - H(\boldsymbol{I})]. \tag{5.22}$$

Introducing the torus angles $\boldsymbol{\theta}$ as dummy integration variables, we obtain an integral over phase space $\boldsymbol{I}$, $\boldsymbol{\theta}$:

$$\bar{d}(E) \approx \frac{1}{h^N} \int \mathrm{d}\boldsymbol{I} \int \frac{\mathrm{d}\boldsymbol{\theta}}{(2\pi)^N} \delta[E - H(\boldsymbol{I})]. \tag{5.23}$$

After a canonical transformation from $\boldsymbol{I}$, $\boldsymbol{\theta}$ back to the original $\boldsymbol{q}$, $\boldsymbol{p}$ variables (a transformation whose Jacobian is unity), this gives precisely eq. (5.7).

It is more tricky to extract $\bar{d}(E)$ from the implicit quantization condition (4.6) for Sinai's billiard, and I simply outline the procedure (details are given in ref. [70]). A slight manipulation of eq. (4.6) gives

$$F(E) \equiv \det_{l,l'} \left(\frac{\mathrm{e}^{-\mathrm{i}\eta_l(E)}}{\sin \eta_l(E)} \delta_{l,l'} + S_{l-l'}(E) \right) = 0, \tag{5.24}$$

where $F(E)$ is a function which can be shown to be real for real E. At each zero E_j (energy level of Sinai's billiard), the phase of F jumps by π. F also has poles E_p (where the structure constants S_l diverge and where $\sin \eta_l$ vanishes), and at E_p the phase of F also jumps by π. This leads to the following representation for $\mathscr{N}(E)$:

$$\mathscr{N}(E) = \lim_{\varepsilon \to 0} -\frac{1}{\pi} \operatorname{Im} \ln[F(E + \mathrm{i}\varepsilon)] + \sum_p \Theta(E - E_p). \tag{5.25}$$

After some algebra this gives the exact formula

$$\mathcal{N}(E) = \sum_m \Theta(k^2 - 4\pi^2 m^2) + \frac{1}{\pi}\sum_{l=-\infty}^{\infty} \eta_l(E)$$

$$- \frac{1}{\pi}\mathrm{Im}\,\mathrm{Tr}\,ln[\delta_{ll'} + \sin\eta_l(E)e^{i\eta_l(E)}S_{l-l'}(E)]. \tag{5.26}$$

The first term is the "unperturbed" mode number corresponding to the integrable billiard with $R=0$, with steps at each level of the "empty" torus. The other terms embody scattering by the discs. Only the first two terms contribute to the mean mode number in lowest approximation, and a little analysis [70] leads to

$$\bar{\mathcal{N}}(E) = \frac{(1-\pi R^2)k^2}{4\pi}, \tag{5.27}$$

which is precisely the "Weyl" result (5.20) for this case.

In order to test numerically these theories for $\bar{\mathcal{N}}(E)$, it is necessary to compute the levels E_j, construct the exact stepped curve $\mathcal{N}(E)$ using eq. (5.2) and compare its trend with eq. (5.6). Such a test has been carried out for the desymmetrized Sinai billiard (fig. 35) and is very instructive. Figure 49 shows the comparison of $\mathcal{N}(E)$ (stepped curve) with the Weyl formula (5.27) (divided by 8 because of the reduced area after desymmetrization) (full curve), for values of R from zero (integrable) to $R=0.4$. Evidently the agreement is poor. The reason is that eq. (5.27) is an *asymptotic* formula which is here being tested on *lowlying* states. It is necessary to include *corrections* to eq. (5.27). These are not contributions to $\mathcal{N}_{\mathrm{osc}}$ (which will be discussed in the next section), but are smooth terms of lower order in k, depending on aspects of the billiard geometry other than its area $\mathcal{A}$. For the desymmetrized Sinai billiard, general formulae [94] give

$$\bar{\mathcal{N}}(E) \approx \underbrace{\frac{1}{32\pi}(1-\pi R^2)k^2}_{\text{area}} - \underbrace{\frac{1}{4\pi}\left[1 + 2^{-1/2} - R\left(2-\frac{\pi}{4}\right)\right]k}_{\text{perimeter}} + \underbrace{\frac{31}{96}}_{\substack{\text{curvature} \\ + \text{ corners}}}. \tag{5.28}$$

This corrected formula is shown in fig. 49 as dashed curves; evidently the agreement is dramatically improved. Computations of the spectrum of the (pseudo-integrable) desymmetrized square torus billiard [21] give essentially the same result.

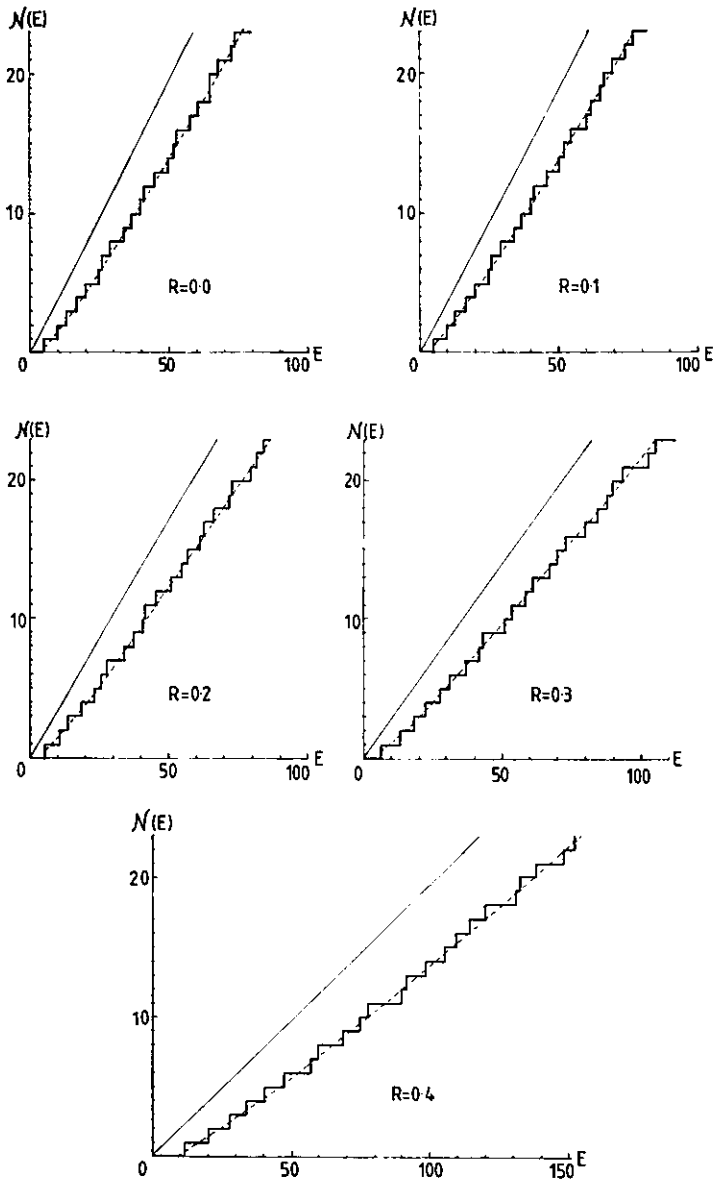


Fig. 49.

This study of the spectrum on the coarsest scales, embodied in $\bar{d}$ and $\mathcal{N}$, has revealed that with this degree of smoothing all evidence of regularity of the underlying classical motion is lost.

5.2. Oscillatory corrections and closed orbits

The results of the last section confirm what was asserted in section 4, that the mean level spacing $\bar{d}^{-1}$ is of order $\hbar^N$. This might lead us to expect that the mean level density $\bar{d}(E)$ would be obtained simply by smoothing the exact $d(E)$ over an energy range of order $\hbar^N$. But such an expectation is mistaken, because as we shall see the corrections $d_{\rm osc}$ oscillate with energy "wavelength" of order $\hbar$, which when $N>1$ (i.e. in nontrivial cases) is infinitely larger than the mean spacing as $\hbar \to 0$. In this important respect the spectrum in the general case contrasts with those special cases (which actually correspond to $N=1$ as analyzed for example in ref. [3]) represented by eq. (5.5), in which the slowest-oscillating corrections have energy wavelength equal to the mean spacing.

The clearest route to understanding the nature of the contributions to $d_{\rm osc}$ is the Green function method based on eqs. (5.12)–(5.15). Because full details can be found in the original papers by Gutzwiller [84–88] and Balian and Bloch [89–92], which have been partially reviewed before [3]. I confine myself here to using the very simplest arguments leading to the essential results.

As defined by eq. (5.12) the Green function $G^+(q,q';E)$ gives the amplitude at q of waves continuously emitted from q' with energy E. The corresponding classical paths form an N-parameter family (N components of momentum p' of the emitted particles, minus 1 because p' must lie on the energy surface $H(q',p')=E$ with q' and E fixed, plus 1 because particles are emitted at all times). Therefore the method described in section 3.1 can be employed to construct the semiclassical Green function as a series of contributions from *all classical paths leading from q' to q with energy E.*

For the jth such path, let the momentum at q be $p_j(q;q',E)$. Then the phase of this path's contribution to G^+ is $S_j(q,q';E)/\hbar$, where

$$S_j(q,q';E)=\int_{q'}^{q} p_j(q_1,q';E)\cdot \mathrm{d}q'. \tag{5.29}$$

Derived from S_j is an amplitude factor $a_j(q,q';E)$ whose form will not be specified here (see ref. [84]), and this contributes to G^+ after multiplica-

tion by $\hbar^{-(N+1)/2}$ (as can be seen for example from the "free-particle" formula (5.17) on replacing the Hankel function by its asymptotic approximation). Therefore

$$G^+(q,q';E) \approx \frac{1}{\hbar^{(N+1)/2}} \sum_j a_j(q,q';E) \exp\left(\frac{\mathrm{i}}{\hbar} S_j(q,q';E)\right). \tag{5.30}$$

To find $d(E)$ it is necessary to let $q' \to q$ and take the trace (5.11) by integrating over q. When $q' \to q$, eq. (5.32) includes all classical paths beginning and ending at q. These are of two kinds: the "paths of zero length" – i.e. the limit as $\varepsilon \to 0$ of the direct path from q to $q+\varepsilon$—and the paths looping back to q after a finite excursion. The paths of zero length were discussed in the last section, and shown to give rise to the smooth term $\bar{d}$ [cf. eqs. (5.17) and (5.18)]. The looping paths contribute to d_{osc} as I now show.

A looping path need not be a closed orbit, because it may (and usually does) return to q with a momentum p different from its initial momentum p'. But in the integration (5.11) over q such non-closed looping paths give negligible contributions. The reason is that $S_j(q,q;E)$ must be stationary under local variations of q if the contributions from nearby positions are not to cancel by destructive interference, and this implies

$$\begin{aligned} \nabla_q S(q,q;E) &= \lim_{q' \to q} [\nabla_q S(q,q';E) + \nabla_{q'} S(q,q';E)] \\ &= \lim_{q' \to q} [p(q) - p'(q')] = 0, \end{aligned} \tag{5.31}$$

so that the initial and final momenta must in fact be equal.

We have now reached the following central conclusion: *only closed orbits with energy E contribute to* d_{osc} (E); these include repetitions of primitive orbits, i.e. closed orbits traversed once. To study the form of these contributions, let j now label the primitive closed paths, and let the action round the jth path be

$$S_j(E) = \oint p_j(q;E) \cdot \mathrm{d}q. \tag{5.32}$$

Then it is clear that the phase of the contribution to d_{osc} from the jth primitive closed orbit, when traversed p times, is $pS_j(E)/\hbar$.

The amplitude is found by performing the integral over q in eq. (5.11), and the manner of doing this depends on two aspects of the primitive closed orbits: whether they are *isolated or nonisolated*, and whether they are *stable or unstable*.

Let the jth primitive closed orbit be embedded in an l_j-parameter family of closed orbits. l_j can vary from zero (for an isolated orbit) to $N-1$ (for the torus-filling orbits of an integrable system - section 2.2 - taking account of the fact that each closed orbit occupies one dimension). Then in eq. (5.11) l_j+1 dimensions of q integration (one along the orbit and l_j "across" the family) can be performed easily because the phase $S_j/\hbar$ is constant, and gives a factor corresponding to the measure of the family. The remaining $N-l_j-1$ dimensions, over the looped non-closed paths in the neighbourhood of the family, must be performed by stationary phase; each dimension gives a factor $\hbar^{1/2}$. Using eq. (5.30) and eq. (5.11), the resulting formula for the oscillatory level density corrections is

$$d_{\mathrm{osc}}(E)=\sum_j \sum_{p=1}^{\infty} \frac{A_{j,p}(E)}{\hbar^{1+l_j/2}} \sin\left(\frac{pS_j(E)}{\hbar}+p\alpha_j\right). \tag{5.33}$$

This is a sum over all primitive orbits j and repetitions p. The phases α_j are analogous to the Maslov indices discussed in section 3.2, and depend on the focusing of trajectories close to the closed orbit. The behaviour of the amplitudes $A_{j,p}$ as function of repetition number p depends on the stability of the primitive orbit j. For isolated orbits, $A_{j,p}$ oscillates with p if the orbit is stable, and decays exponentially if the orbit is unstable - types of behaviour to be expected in view of the repeated focusing or continued defocusing of beams of trajectories. For integrable systems, where all orbits are embedded in $N-1$ parameter families, Berry and Tabor [95,96] showed that $A_{j,p}$ decreases as $p^{-(N-1)/2}$. The Green function theory leading to eq. (5.33) is reviewed in detail by Rajaraman [97] and de-Witt-Morette et al. [98].

Just as for $\bar{d}(E)$ it is instructive to see how the general formula (5.33) for $d_{\mathrm{osc}}(E)$ is implicit in the quantization conditions obtained in section 4.1.

Consider first the integrable case, where torus quantization gave eq. (5.21) for $d(E)$. This is an N-dimensional sum over quantum numbers $\boldsymbol{m}$, which can be transformed exactly into a sum of integrals in action space $\boldsymbol{I}$ by using the Poisson summation formula and the relation (3.20) between $\boldsymbol{I}$ and $\boldsymbol{m}$. Recall first that the Poisson formula transforms sums over a unit lattice $\boldsymbol{m}$ into sums over another unit lattice $\boldsymbol{M}$: for any function $f(\boldsymbol{m})$ defined on the lattice,

$$\sum_{\boldsymbol{m}} f(\boldsymbol{m})=\sum_{\boldsymbol{M}} \int \mathrm{d}\boldsymbol{m} \exp(2\pi \mathrm{i}\boldsymbol{m}\cdot\boldsymbol{M}) f(\boldsymbol{m}), \tag{5.34}$$

where the integrals involve any interpolation of f for continuous $\boldsymbol{m}$. Applied to eq. (5.21) this procedure gives

$$d(E)=\frac{1}{\hbar^N}\sum_{\boldsymbol{M}}\exp(-\mathrm{i}\frac{\pi}{2}\alpha\cdot\boldsymbol{M})\int \mathrm{d}\boldsymbol{I}\,\delta[E-H(\boldsymbol{I})]\exp\left(\frac{2\pi\mathrm{i}}{\hbar}\boldsymbol{M}\cdot\boldsymbol{I}\right). \quad (5.35)$$

The term $\boldsymbol{M}=0$ corresponds to replacing the original $\boldsymbol{m}$-sum by an integral, which as we saw in eq. 5.1 gives the mean level density $\bar{d}$. The terms $\boldsymbol{M}\neq 0$ give the oscillatory corrections we want to study now.

Action space $\boldsymbol{I}$, that is the space whose points are tori, is stratified by the surfaces of constant E, which have $N-1$ dimensions. The delta function in eq. (5.35) restricts integration to the surface with energy E. Introduce coordinates $\xi=\{\xi_1\ldots\xi_{N-1}\}$ on this surface (fig. 50). Then for small $\hbar$ eq. (5.35) involves an integration with respect to ξ over a rapidly-oscillating exponential whose phase is $2\pi\boldsymbol{M}\cdot\boldsymbol{I}(\xi)/\hbar$. The principal contributions will come from those values $\xi^{\boldsymbol{M}}$ for which this phase is stationary, i.e. for which

$$\boldsymbol{M}\cdot\frac{\partial\boldsymbol{I}}{\partial\xi_i}=0 \qquad \text{when } \xi=\xi^{\boldsymbol{M}} \qquad (i=1,2\ldots N-1). \quad (5.36)$$

Because $\partial\boldsymbol{I}/\partial\xi_i$ are tangent vectors in the energy surface, this condition for $\xi^{\boldsymbol{M}}$ has the geometric meaning that the tori $\boldsymbol{I}^{\boldsymbol{M}}$ contributing to d_{osc} are those lying on the energy surface at places where it is perpendicular to $\boldsymbol{M}$. But the frequency vector ω[eq. (2.10)] is also perpendicular to the energy surface, so that ω and $\boldsymbol{M}$ are parallel in action space. Now, the components M_i are integers, so that the ω_i are mutually commensurate. Therefore the tori $\boldsymbol{I}^{\boldsymbol{M}}$ are just those on which orbits are *closed* [cf. the discussion follow-

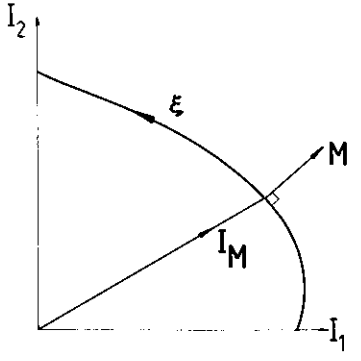


Fig. 50.

ing eq. (2.11)]. Moreover the lattice vector $\boldsymbol{M}$ specifies the topology of the orbit in terms of its winding numbers round the cycles of the torus. Therefore we have shown that the terms $\boldsymbol{M} \neq 0$ in eq. (5.35) correspond to all the topologically distinct closed orbits with energy E. The phase of each contribution is $2\pi \boldsymbol{M} \cdot \boldsymbol{I}^{\boldsymbol{M}}/\hbar$, which according to the definition (2.4) is precisely the action S round the orbit. This argument shows that for integrable systems torus quantization gives terms for d_{osc} precisely in accord with the general formula (5.33). For further details see refs. [95,96].

It is at first surprising to find that d_{osc} depends on the closed orbits in this way, because the representation (5.23) from which we started involved the actions $\boldsymbol{I}_m$ corresponding to the quantized tori, and not the actions $\boldsymbol{I}^{\boldsymbol{M}}$ of the "rational" tori supporting closed orbits. These actions are different, and in particular the actions $\boldsymbol{I}_m$ corresponding to individual quantum states will not generally have commensurable frequencies and so will not support closed orbits. I shall discuss this point further in the next section, but remark that the apparently paradoxical result that all details of $d(E)$, including positions of individual levels, can be obtained by summing closed orbits is in fact analogous to, and no more surprising than, representing an irrational number as a sum over infinitely many rationals.

In the case of Sinai's billiard, extraction of the oscillatory contributions from the implicit quantization condition (4.6) is a much more difficult business and I only indicate the results (a full derivation is given in ref. [70]). The oscillatory contributions come from the first and third terms in the exact representation (5.26). The first term (which corresponds to the spectrum of the integrable "empty torus" or "empty lattice") can be transformed by the Poisson formula as in the general integrable case just discussed. This gives terms $\mathcal{N}_{\mathrm{osc}}$ corresponding to each closed orbit belonging to the *nonisolated* set, which do not strike a disc (cf. fig. 12). Two examples are shown in fig. 51. But also included in this set of contributions are terms corresponding to nonisolated orbits that would pass through "ghost" discs, such as those in fig. 52. These are obviously un-

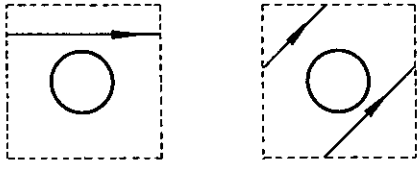


Fig. 51.

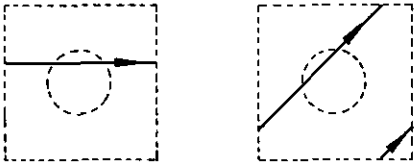


Fig. 52.

physical and in fact cancel as will soon be explained.

The contributions $\mathcal{N}_{\rm osc}$ from the third term in eq. (5.26) are extracted as follows. First the Tr ln is expanded using the formula

$$\mathrm{Tr}\, ln(\delta_{ll'} + M_{ll'}) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left(\sum_{l_1} \dots \sum_{l_n} M_{l_1 l_2} M_{l_2 l_3} \dots M_{l_n l_1} \right). \tag{5.37}$$

Next, the phase shift factors are written as

$$\sin \eta_l e^{i\eta_l} = \frac{1}{2i}(e^{2i\eta} - 1) \tag{5.38}$$

and the "-1" neglected for the moment. Then the sums over angular momenta l are replaced by integrals. Next, all Bessel functions are replaced by their asymptotic approximations. Finally the multiple l-integrals are evaluated by the method of stationary phase. The result of this lengthy procedure is that the nth term in eq. (5.37) gives contributions to $\mathcal{N}_{\rm osc}$ from *isolated* closed orbits with n bounces from discs (cf. fig. 13). Two examples with $n = 4$ are shown in fig. 53. But also included in this set of contributions are oscillatory terms corresponding to closed orbits that would pass through "ghost" discs, as shown in fig. 54. Just like their nonisolated counterparts, these are obviously unphysical.

Both sorts of impossible path are cancelled by contributions from the "-1" terms in eq. (5.38). This occurs through a subtle diagrammatic identity [70] whose action can be glimpsed from the equation represented sym-

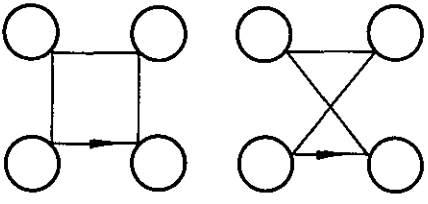


Fig. 53.

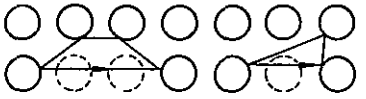


Fig. 54.

bolically in fig. 55, which refers to the impossible leg of the first diagram in fig. 54.

After all this, the final result is that $d_{\rm osc}$ contains contributions from all topologically distinct physically possible closed orbits. These are of two sorts. Firstly, terms from nonisolated orbits, which give contributions of order $\hbar^{-3/2}$ in precise accord with eq. (5.33) for this case where the orbits form one-parameter families. Secondly, terms from isolated orbits, which give contributions of order $\hbar^{-1}$, again in accord with eq. (5.33). The isolated orbits are all unstable and their contributions decay with increasing repetition as Δ_j^{-p}, where Δ_j (whose value exceeds unity) is a "determinant of bounces" corresponding to the jth primitive closed orbit; the order of Δ_j is the number n_j of bounces in the primitive orbit.

5.3. Comments on the closed-orbit sum

The most important feature of eq. (5.33) is its implication that each closed orbit (labelled by j and p) contributes an oscillation to the level density $d(E)$. The oscillation has energy "wavelength" ΔE given by

$$\frac{p}{\hbar}\frac{\mathrm{d}S_j}{\mathrm{d}E}\Delta E = 2\pi\,. \tag{5.39}$$

Now from eq. (5.34),

$$\frac{\mathrm{d}S_j}{\mathrm{d}E} = \oint \frac{\mathrm{d}p_j}{\mathrm{d}E}\cdot\mathrm{d}q = \int \mathrm{d}t\,\dot{q}\cdot\frac{\mathrm{d}p_j}{\mathrm{d}E} = \int \mathrm{d}t\,\nabla_p H\cdot\frac{\mathrm{d}p_j}{\mathrm{d}E} = T_j(E), \tag{5.40}$$

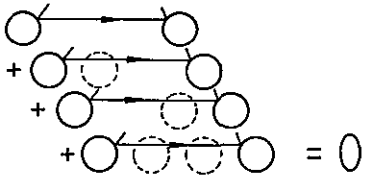


Fig. 55.

where $T_j(E)$ is the period of the jth closed orbit with energy E, so that

$$\Delta E = h/pT_j(E), \tag{5.41}$$

so that the closed orbit describes level clustering on scales of order $\hbar$. Now recall from eq. (5.7) that the mean level spacing of order $\hbar^N$. Therefore if $N>1$ the closed orbits do indeed describe level clustering on scales much larger than the mean level spacing, as asserted at the beginning of section 5.2.

For billiards,

$$\frac{S_j}{\hbar} = \oint \frac{\boldsymbol{p}_j \cdot \mathrm{d}\boldsymbol{q}}{\hbar} = kL_j, \tag{5.42}$$

where L_j is the spatial length of the jth orbit, implying that the corresponding k-wavelength Δk of the oscillations is

$$\Delta k = 2\pi/L_j. \tag{5.43}$$

Now I want to dispose of a fallacy based on a misinterpretation of eq. (5.33). Consider the terms with given j, i.e. those corresponding to all repetitions of a single primitive closed orbit. The terms will interfere constructively if

$$S_j(E_m) = (2\pi m - \alpha_j)\hbar, \tag{5.44}$$

defining a series of energies E_m corresponding to integers m. At these energies, the sum of the contributions to d_{osc} depends on the amplitude factors $A_{j,p}(E)$: for stable isolated orbits, the sum gives delta-functions at E_m [87], for unstable isolated orbits, the sum gives Lorentzian peaks at E_m [87], while for nonisolated orbits the sum gives other singularities at E_m [91,95], e.g. of logarithmic or inverse square root type. The fallacy is to suppose that E_m are eigenvalues of the Hamiltonian, i.e. that eq. (5.44) is a quantization condition associating individual quantum states with repetitions of individual closed orbits.

Why is this a fallacy? For a start, eq. (5.44) gives energies separated by distance of order $\hbar$, whereas the energy levels have separation $\hbar^N$. But why not superpose the level sequences obtained from eq. (5.44) with all topologically diffeent orbits j? Because this would give too many levels! An instructive demonstration of this is provided by free motion of a particle with mass μ on a coordinate 2-torus, represented by a rectangle with sides a, b and periodic boundary conditions. This is an integrable system,

whose exact levels E_m are labelled by $m=(m_x, m_y)$ and given by

$$E_m=\frac{2\pi^2\hbar^2}{\mu}\left(\frac{m_x^2}{a^2}+\frac{m_y^2}{b^2}\right). \tag{5.45}$$

Compare this with what eq. (5.44) would give: there are closed orbits labelled by $M=(M_x, M_y)$, corresponding to lattice translations $(M_x a, M_y b)$ (windings round the torus), with length

$$L_M=(M_x^2a^2+M_y^2b^2)^{1/2}. \tag{5.46}$$

Then eq. (5.44) would give, on using eq. (5.42),

$$E_{m,M}=\frac{\hbar^2}{2\mu}\frac{(2\pi)^2m^2}{L_M^2}=\frac{2\pi^2\hbar^2m^2}{\mu(M_x^2a^2+M_y^2b^2)}. \tag{5.47}$$

This is clearly nonsense: it describes "levels" labelled by three quantum numbers rather than two, with an "infrared catastrophe" of levels with arbitrary small E corresponding to slowly-traversed closed orbits which are nevertheless long enough for their action to exceed $\hbar$.

Nevertheless, there are two circumstances where eq. (5.44) does give semiclassical quantal levels correctly. The first (and trivial) case is potential wells with $N=1$, when there is only one topology of closed orbit and the levels do have separation $\hbar$ (see for example ref. [3] and references therein). The second case occurs when the closed orbits are isolated and stable. Then Miller [99] showed that by considering lowest-order fluctuations about the closed orbit it was possible to generalize eq. (5.44) into a condition with a full set of N quantum numbers. But Voros [44] pointed out that isolated stable orbits are always surrounded by tori, and explained how this quantum condition is really an approximate version of torus quantization, appropriate for the thin tori surrounding isolated stable orbits (when applied to a circularly symmetric potential with $N=2$, for example, the approximation is accurate for levels near the bottoms of the wells in the effective potentials for each angular momentum).

In general, though, a single closed orbit gives not individual levels but a collective property of the spectrum, namely an oscillatory clustering with scale ΔE given by eq. (5.41). Conversely the determination of individual levels from of eq. (5.33) involves the closed orbits collectively, and would require the summation over sufficiently many closed orbits for individual delta functions to emerge as a result of constructive interference at certain energies and destructive interference at all other energies. Is this a feasible

procedure for calculating the individual levels? I shall now argue that it is not.

To begin to see delta functions emerging from the path sum (5.33), it is necessary to include at least all orbits giving oscillations whose energy wavelength ΔE, given by eq. (5.41), exceeds the mean spacing $\delta E = \bar{d}^{-1}$ between neighbouring levels. Since the lower orbits give faster oscillations, it is necessary to sum over all closed orbits whose periods $pT_j(E)$ are less than $T_{\max}$, given by

$$T_{\max} = \frac{\int \mathrm{d}q \int \mathrm{d}p\, \delta[E - H(q,p)]}{h^{N-1}}. \tag{5.48}$$

For billiards with $N=2$, it is necessary to sum over all closed orbits with length less than $L_{\max}$, where [using eq. (5.19)]

$$L_{\max} = k\mathscr{A}, \tag{5.49}$$

where $\mathscr{A}$ is the area of the billiard.

Semiclassically, i.e. as $\hbar \to 0$ or $k \to \infty$, $T_{\max}$ and $L_{\max}$ increase, and it is necessary to include ever more closed orbits in the sum. How many? This depends on the behaviour as $T \to \infty$ of the function

$$\nu(T) = \text{number of closed orbits with periods less than } T, \tag{5.50}$$

or its billiard analogue $\nu(L)$. For integrable systems, closed orbits are classified by the N-dimensional winding number $\boldsymbol{M}$, eq. (2.11), and T is proportional to the components of $\boldsymbol{M}$, so

$$\nu(T) \sim T^N \qquad \text{as } T \to \infty \tag{5.51}$$

(this is just the number of unit lattice points within an N-dimensional sphere whose radius is proportional to T). Therefore the number ν_s of steps required to determine the levels of an integrable system by direct summation over closed orbits is

$$\nu_s \sim \nu(T_{\max}) \sim \hbar^{-N(N-1)}. \tag{5.52}$$

Of course this would be a foolish way to calculate the levels of an integrable system, because the torus quantization formula (4.1) gives the levels explicitly. Nevertheless it is instructive to see the delta functions emerging as more topologies of closed orbit are included, and I now illustrate this with a two-dimensional centralforce Hamiltonian (cf. fig. 2). Orbits classified by (M_1, M_2) close after M_1 librations and M_2 rotations;

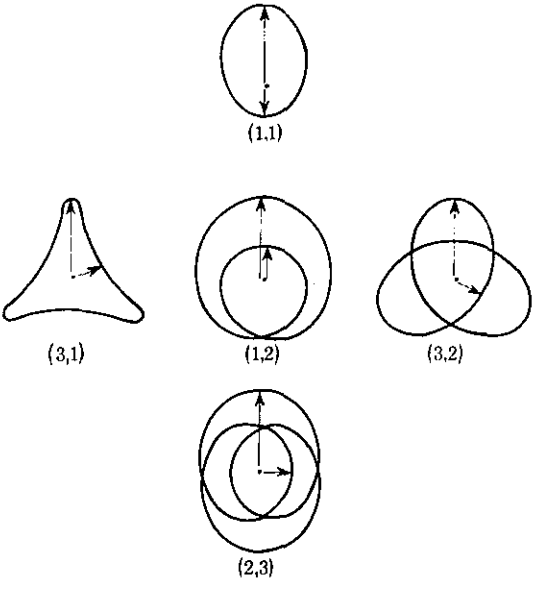


Fig. 56.

fig. 56 shows some of the simplest topologies. The Hamiltonian is a particle moving in a Morse potential, namely

$$H=\frac{p_x^2+p_y^2}{2\mu}+V_0(e^{-2\delta(r-r_0)}-2e^{-\delta(r-r_0)}) \tag{5.53}$$

($\mu=1$ proton mass, $V_0=0.2$ eV, $r_0=0.25$ nm, $\delta=10$ nm^{-1}) which has a total of 166 bound levels. Figure 57 shows the effect of including increasing numbers of topologies M. In the last frame, which includes several hundred closed orbits, the delta functions are beginning to emerge very clearly (arrows mark exact levels, and $\bar{d}$ is given by the chain curves). For more details, see ref. [95].

What about ergodic systems? These will be discussed with reference to Sinai's billiard. Simple arguments [70] indicate that there is only a finite number of nonisolated closed orbits, but that the isolated orbits proliferate exponentially, i.e.

$$\dot{\nu}(L)\sim\exp(\text{constant}\times L) \qquad \text{as } L\to\infty, \tag{5.54}$$

so that the isolated orbits dominate the sum (5.31) in spite of their individually weaker contributions. As mentioned at the end of section 5.2,

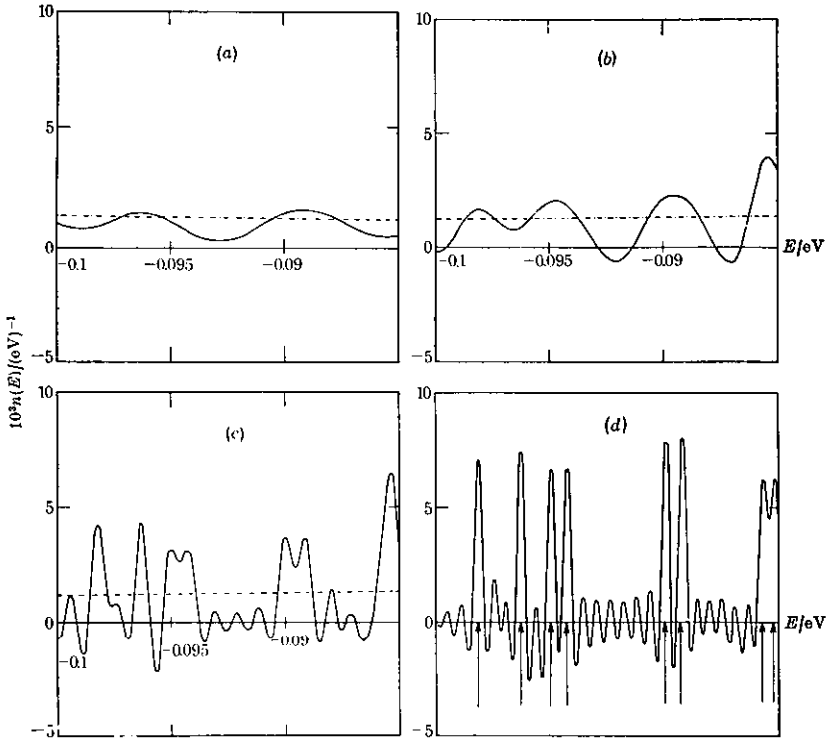


Fig. 57.

the contribution of each such orbit involves the $n_j \times n_j$ determinant of bounces. The computation of such a determinant (by Gaussian elimination) takes a number of steps of order n_j^3. As L increases, the majority of new orbits are traversed only once, and so the bounce number n_j is proportional to L. Therefore the number ν_s of steps required to evaluate the path sum for this ergodic system is

$$\nu_s \sim \nu(L_{\max}) \times L_{\max}^3 \sim (k\mathscr{A})^3 \exp(\text{constat} \times k\mathscr{A}),$$
$$\sim \hbar^{-3} \exp(\text{constant}/\hbar) \tag{5.55}$$

and greatly exceeds the corresponding number (5.51) for an integrable system.

However, there is in this case no torus quantization procedure to give the levels, and the closed orbit sum must compete instead with the determinant (4.6), whose zeros give the levels. How many steps ν_s are required to

evaluate this? The answer lies in eq. (4.9), which gives the effective size of the determinant and leads [70] to

$$\nu_s \sim \hbar^{-3}. \tag{5.56}$$

It is obvious, therefore, that in the semiclassical limit the "compact" representation in the form of a determinant is vastly (i.e. exponentially) more efficient than the "expanded" representation as a closed path sum, as a means for calculating individual levels.

In a tour de force, Gutzwiller [101] has recently evaluated the path sum for an ergodic system, namely the anisotropic Kepler problem with Hamiltonian (2.12). The exponentially large number of closed orbits below any given length are classified in terms of binary sequences, which enables the path sum (5.33) to be interpreted as the partition function for a system of interacting spins, and this can be approximated by a determinant in terms of whose eigenvalues the energy levels can be determined. In this way the first 18 levels are computed with an accuracy of a few per cent. But a 50×50 determinant must be diagonalized in order to achieve this, so that [as expected on the basis of eq. (5.55)] the closed path sum is hardly a practical method for calculating the levels.

The enormous labour involved in the evaluation of closed orbit sums for nonintegrable systems does not mean that they are useless in understanding the spectrum. Suppose for example that what is required is not a complete knowledge of all the levels, but the level density *smoothed* over energy ranges $\Delta E \gg \delta E$. Then the path sum need only be taken up to paths with period pT_j given by eq. (5.41), which will involve only a small number of paths if ΔE is large. In this way Richens and Berry [21] obtained with 225 closed orbits a good fit to the smoothed $d(E)$ in the case of the (pseudointegrable) square torus billiard. A physical problem for which the closed path sum is a promising approach is the acoustics of auditoriums, where absorption at the walls causes individual eigenfrequencies to be broadened into resonances much wider than their separation, thus effecting a natural smoothing.

Finally, although I have presented these closed orbit expansions as semiclassical approximations, and this is how they arise in the general theories of Gutzwiller [84–88] and Balian and Bloch [89–91], it is encouraging that there are two *exact* connections known between the spectrum and the closed orbits for nonintegrable systems. Firstly, for the wave equation on a smooth closed Riemannian manifold (which need not have constant

curvature), if the eigenwavenumbers are k_j $[=(2\mu E_j)^{1/2}/\hbar]$, then the function

$$Q(\lambda) \equiv \sum_{j=1}^{\infty} e^{ik_j\lambda} \tag{5.57}$$

was proved by Chazarain [102] to have singularities for each λ equal to the length of a closed geodesic [$Q(\lambda)$ is the Fourier transform of the k density of states]. Balian and Bloch [92] obtain a similar exact result for the Schrödinger equation: the transform with respect to $\hbar^{-1}$ is singular at the action values of classical closed orbits. And secondly. the "Selberg trace formula" (reviewed by McKean [103] and Hejhal [104]) shows that for the ergodic system consisting of geodesic motion on a compact surface of constant negative curvature a sum over eigenvalues of the Laplacian is exactly given by a sum over (isolated, unstable) closed orbits. Unfortunately, neither side in the (exact or asymptotic) equality

$$\sum \text{(eigenvalues)} \approx \sum \text{(closed orbits)} \tag{5.58}$$

is exactly known for any nonintegrable system.

6. Quantum maps

6.1. Evolving states

Recall the association described in sections 3.1 and 3.2, between quantal wave functions $\psi(q)$ and N-dimensional surfaces Σ in phase space. The association, embodied in eq. (3.7), was natural because under semiclassical conditions it persists as the state ψ and surface Σ evolve, one classically and the other quantally, under the action of the Hamiltonian H. Now I want to consider the largely unsolved problem of what happens to this association over long times t. The following argument indicates why we can expect this problem to be difficult. In almost all cases (i.e. when Σ is not an invariant surface and H is not linear or quadratic) Σ develops, as will be explained, an infinitely complicated morphology as $t \to \infty$, with foldings and convolutions down to arbitrarily fine scales. But it is hard to see how phase-space fine structure in volumes smaller than h^N can have quantal significance. This suggests that the asymptotic association between ψ and Σ holds only up to a time $t_{\max}(\hbar)$ (which increases with $\hbar$), after which ψ

and Σ evolve differently. Alternatively stated, we expect that the limits $\hbar \to 0$ and $t \to 0$ cannot be interchanged.

A wide-ranging and perceptive discussion of the problem is given by Chirikov et al. [106], especially in the context of classically chaotic systems (although even the semiclassical approximation as $t \to \infty$ of nonstationary states in classically *integrable* systems presents difficulties). Possible meanings that can be assigned to the term "quantum stochasticity" for evolving states have given rise to an extensive chemical literature which I shall not attempt to summarize; it is well discussed by Hutchinson and Wyatt [107].

A direct analytical attack on the problem appears to be extremely difficult, and so it is natural to seek simple models where it is possible to compute the evolution of both ψ and Σ and see if and when the association between them breaks down. This is where area-preserving maps M of the phase plane q, p (discussed in section 2.6) become useful (and indeed almost indispensible, given the limitations of present computers). There are four main reasons for using such maps as models. Firstly, they can easily be quantized as I show in a moment. Secondly they show all varieties of regular and chaotic behaviour. Thirdly, phase space is a plane, so that "surfaces" Σ are curves whose evolution is easy to depict. And fourthly the coordinate space q is one-dimensional, so that probability densities $|\psi(q)|^2$ can be depicted as graphs. The study of such "quantum maps" was pioneered by Casati et al. [108] in a particular case. Here I present some theory for a rather more general class of maps, as independently developed later by Berry et al. [109]. I shall emphasize, however, that the most interesting questions are still open.

The map to be quantized is eq. (2.24), derived from the T-periodic Hamiltonian (2.23), and some results will be presented for the special case, eq. (2.25). For quantization, we seek the unitary evolution operators $\hat{U}$ corresponding to the action of $\hat{H}$ over time T. This operator transforms states $|\psi_n>$ into states $|\psi_{n+1}>$ according to the quantum map

$$|\psi_{n+1}> = \hat{U}|\psi_n>. \tag{6.1}$$

To find $\hat{U}$ explicitly, we notice that in the time $0 \le t < T$, $\hat{H}$ can be written as the limit $\tau \to 0$ of

$$\begin{aligned} H(\hat{q}, \hat{p}, t) &= \hat{p}^2/2\mu \qquad 0 \le t < T - \tau \\ &= \frac{TV(\hat{q})}{\mu} \qquad T - \tau \le t < T \end{aligned} \tag{6.2}$$

In the first time interval, $0 \le t < T - \tau$, states evolve under

$$\hat{U}_1 = \exp\left(\frac{-\mathrm{i}}{2\mu\hbar}\hat{p}^2(T-\tau)\right), \tag{6.3}$$

and in the second interval, $T - \tau \le t < T$, the evolution operator is

$$\hat{U}_2 = \exp\left(\frac{-\mathrm{i}}{\hbar}TV(\hat{q})\right). \tag{6.4}$$

The full evolution operator $\hat{U}$ is therefore given by

$$\hat{U} = \hat{U}_2\hat{U}_1 = \exp\left(\frac{-\mathrm{i}}{\hbar}TV(\hat{q})\right)\exp\left(\frac{-\mathrm{i}}{2\hbar\mu}\hat{p}^2T\right), \tag{6.5}$$

after letting $\tau \to 0$.

In the position representation, this operator inserted into eq. (6.1) gives the following discrete-time Schrödinger equation giving the quantum map transforming wave functions $\psi_n(q)$ into wave functions $\psi_{n+1}(q)$:

$$\psi_{n+1}(q) = \left(\frac{\mu}{2\pi\hbar T}\right)^{1/2} \exp\left[\frac{\pi}{4} + \frac{TV(q)}{\hbar}\right)\Bigg] \times \int_{-\infty}^{\infty} \mathrm{d}q' \exp\left(\frac{\mathrm{i}\mu}{2\hbar T}(q-q')^2\right)\psi_n(q'). \tag{6.6}$$

This innocent-looking equation can, as we shall see, generate a wealth of interesting structure.

To explain how this was calculated, first note that in the limit $T \to 0$ the Hamiltonian (2.23) can be replaced by its time-average, which is the (integrable) stationary Hamiltonian $\bar{H}$ given by

$$\bar{H}(q,p) = \frac{p^2}{2\mu} + V(q), \tag{6.7}$$

corresponding to an oscillator. Therefore T is a parameter which turns on nonintegrability. For power-law anharmonic oscillators T can be scaled away as in the quartic case (2.25), and this simply means that a point in the original q, p plane [mapping under eq. (2.24)] is associated with a point in the scaled q, p plane which is closer to the origin the smaller T is.

For the initial curve Σ we [109] chose a (closed) contour of the average Hamiltonian (6.7). Of course Σ is not an invariant curve of M when $T > 0$. Nearby points on Σ will map apart under repeated application of M and so Σ will become infinitely long as $n \to \infty$. On the other hand, because of

eq. (2.17) the area inside Σ must remain the same. Therefore Σ will develop infinite complexity. This complexity is due to the development of two principal morphologies in Σ, which are called "whorls" and "tendrils" [109,46] and which are associated with phase plane regions where motion is, respectively, regular and irregular. As can be seen from fig. 26, both sorts of motion are generated by M given by eq. (2.25).

Whorls are associated with invariant curves surrounding a stable fixed point of M. They can arise, for example in the twist map (2.21) (see fig. 22), provided the rotation number α is not independent of radius. What happens is that points at different radii rotate around the central fixed point at different rates. Therefore radii map to spirials, and parts of Σ passing close to stable fixed points will wrap around them as indicated in fig. 58. For the particular M being studied now, we expect from fig. 26 that whorls will occur whenever Σ passes through the central area of the qp plane (covered with invariant curves) or through the surrounding island chain.

Fig. 58.

Tendrils are associated with the chaotic areas surrounding unstable fixed points. Iterates of curves Σ passing through such areas flail wildly back and forth as the individual points of Σ map exponentially apart, as indicated in fig. 59. From fig. 26 we expect tendrils whenever Σ passes through chaotic outer areas of the qp plane.

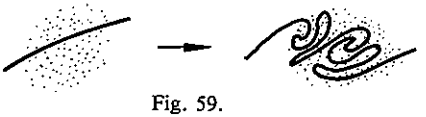


Fig. 59.

Figure 60a and 61 show how these morphologies develop for two initial curves Σ (contours of $\bar{H}$). The first curve ($n=0$ on fig. 60a) lies mainly in the central area, and so begins to wrap into a whorl around the origin. By $n=20$ the "spiral galaxy" structure is clearly visible and two subwhorls have started to appear, associated with other islands. The second curve ($n=0$ on fig. 61) lies largely in the outer, escaping region of the plane and rapidly shoots tendrils out towards infinity.

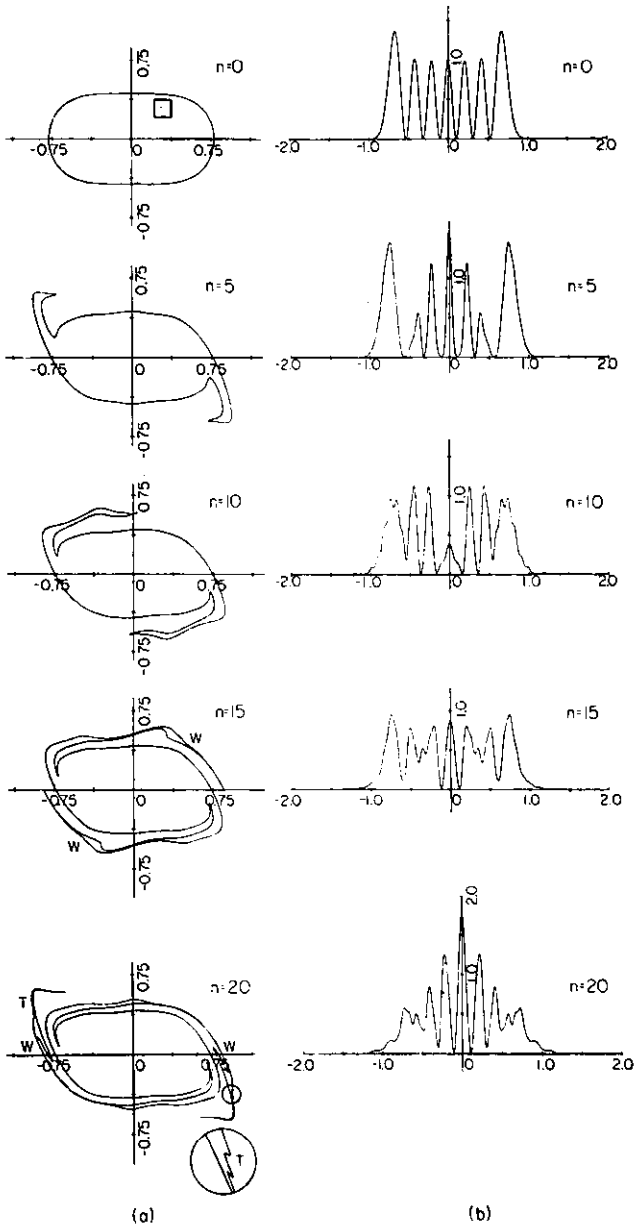


Fig. 60.

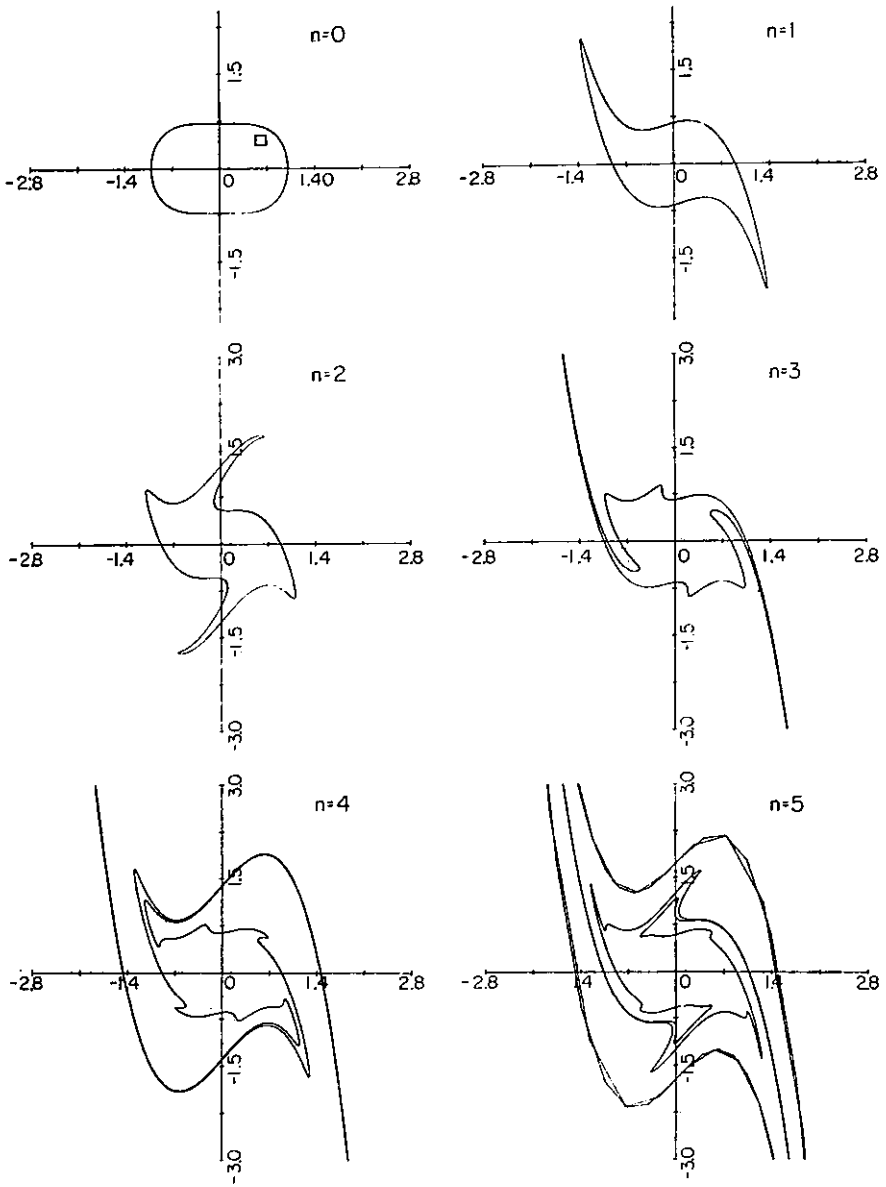


Fig. 61.

The development of a whorl can be seen very clearly in computations by Lewis (reported by Berry [110] of the evolution of a curve under the "nonlinear Harmonic oscillator" Hamiltonian

$$H=f(q^2+p^2), \tag{6.8}$$

whose points map continuously in circles with angular velocity depending on radius. Beautiful tendrils have been calculated by Shepelyansky [111] in the evolution of the curve $p=0$ under the map (2.24) with a sinusoidal potential (the "kicked quantum pendulum").

To study the corresponding quantum map (i.e. eq. (6.6) with a quartic potential $V(q)$, we must choose as initial state $\psi_0(q)$ a wave associated with Σ. Because Σ is an invariant curve of $\bar{H}$, ψ_0 must be an eigenstate of $\bar{H}$. For the curve mapped on fig. 60a, ψ_0 was chosen to be the sixth eigenstate of $\bar{H}$. This state, and its iterates under eq. (6.6), are shown in fig. 60b. The graph of $|\psi|^2$ for $n=0$ clearly shows the association with the initial curve Σ, with maxima at the caustics of projection of Σ, and, between these, quasisinusosidal oscillations caused by the interference of two waves associated with the intersections $p_i(q)$ with the fibre $q=$constant. The value of $\hbar$ corresponding to ψ_0 is the area of the square drawn inside Σ.

By the 20th iteration the classical curve has developed complexity (arms of the whorl) on scales smaller than $\hbar$, and so we expect ψ_n not to be associated with every detail. Instead, it is reasonable according to the "semiclassical hypothesis" stated in section 3.3 (in the context of eigenfunctions) and developed in section 3.4, to associate ψ with the area filled by the whorl, and on the basis of fig. 30 we expect the projection $|\psi(q)|^2$ to display anticaustics. Moreover, associated with the quasicontinuum of p-values intersected by fibres $q=$constant (each giving a wave with wavelength h/p contributing to the interference pattern making up the total ψ) is an obvious loss of spectral purity of ψ. Therefore the predictions of section 3 are confirmed in this case.

For the initial curve of fig. 61, ψ_0 was chosen as the eighteenth eigenstate of $\bar{H}$. The graph of $|\psi_0|^2$, and its iterates, is shown in fig. 62a. The waves ω_0 and ψ_1 are spectrally pure (2-wave interference), because for $n=1$ the classical curve (fig. 61) has not yet "curled over" to give multiple intersections. But it does curl over when $n=2$, with immediate and dramatic effect on $|\psi_2(q)|^2$.

For a more careful classical-quantum comparison, the classical curves of

fig. 61 are projected onto q, and the resulting "classical $|\psi|^2$" graphs are shown in fig. 62c. The most striking features are the caustic spikes, which proliferate as n increases and the classical curve throws off tendrils. The caustics are evidently related to features of the wave functions (fig. 62a) until $n=2$ but thereafter there is no obvious association. This corresponds precisely with the throwing-off for $n-3$ of a fine tendril, smaller than $\hbar$. Related to this is the fact that for large n the neighbouring caustics in fig. 62c are closer than the de Broglie wavelength and so cannot be directly associated with features of ψ. This suggests, in the spirit of the smoothing procedure described in section 3.4 and embodied in eq. (3.29), that a better comparison between classical and quantal calculations will be obtained if we smooth both the graphs of $|\psi|^2$ (fig. 62a) and the curve projections (fig. 62c), on the scale of the average de Broglie wavelength. The smoothed curves are shown in fig. 62b (quantal) and fig. 62d (classical). Clearly the agreement is much better, suggesting a continuing association between ψ and Σ in some average sense if not in detail.

The breakdown of the detailed association between classical and quantum mechanics for this map can be followed very nicely in terms of the Wigner function defined by eq. (3.12). By employing this definition in conjunction with the quantum map equation (6.6), it is not hard [109] to obtain the following quantum map transforming Wigner functions $W_n(q,p)$ into their iterates $W_{n+1}(q,p)$:

$$W_{n+1}(q,p)=\int_{-\infty}^{\infty} dq' \int_{-\infty}^{\infty} dp' K(q,p;q',p')\, W_n(q,p), \tag{6.9}$$

where for the potential $V=Aq^4/4$ the Wigner propagator K is

$$K(q,p;q',p')=\frac{2\delta(q'-q+p'T/\mu)}{\hbar^{2/3}(6AT|q|)^{1/3}}\,\mathrm{Ai}\left(\frac{-2\mathrm{sign}(q)(p'-p-TAq^3)}{\hbar^{2/3}(6AT|q|^{1/3}}\right), \tag{6.10}$$

Ai being the Airy function [72]. As $\hbar\to 0$, the Airy function turns into a delta function, and eq. (6.10) becomes the Liouville propagator transforming classical densities pointwise in accordance with the classical map M (2.25).

For the initial Wigner function, Korsch and Berry [112] chose a Gaussian centred on $q=p=0$, with contours in the form of concentric circles. The half-width of the Gaussian corresponds to a substantial fraction of W lying in the chaotic escaping region of fig. 26. Contours of the first three iterates

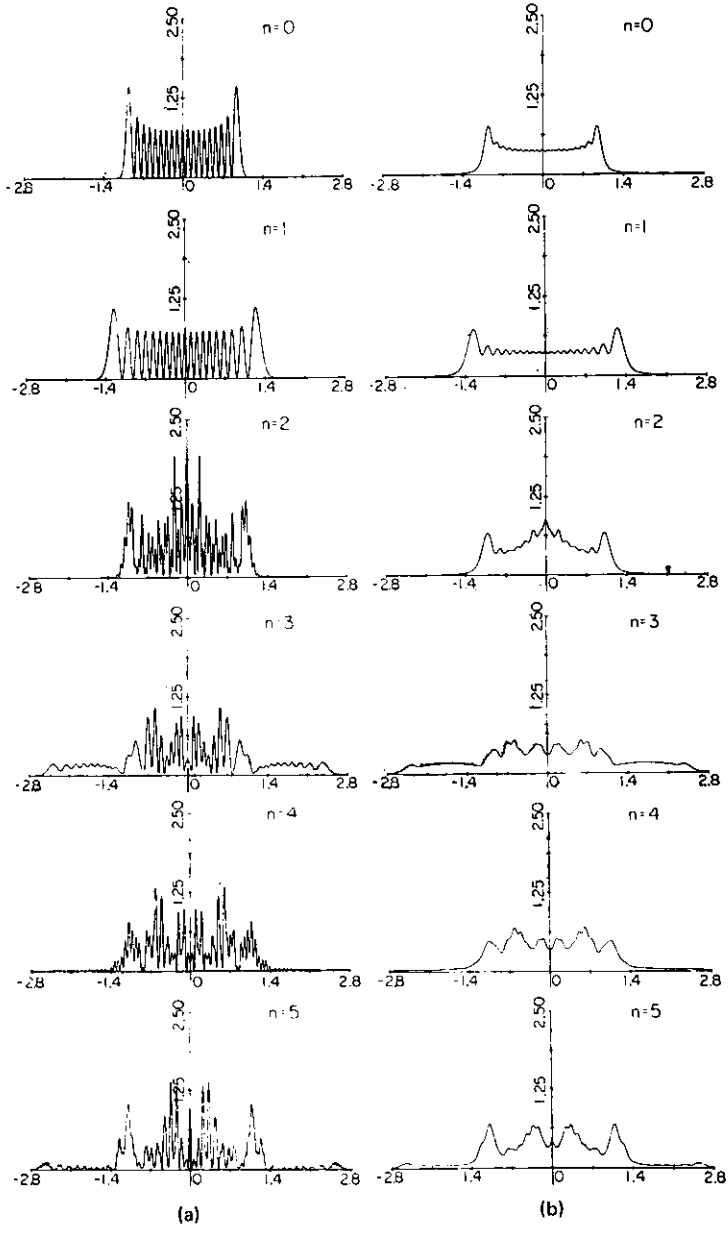


Fig. 62a,b.

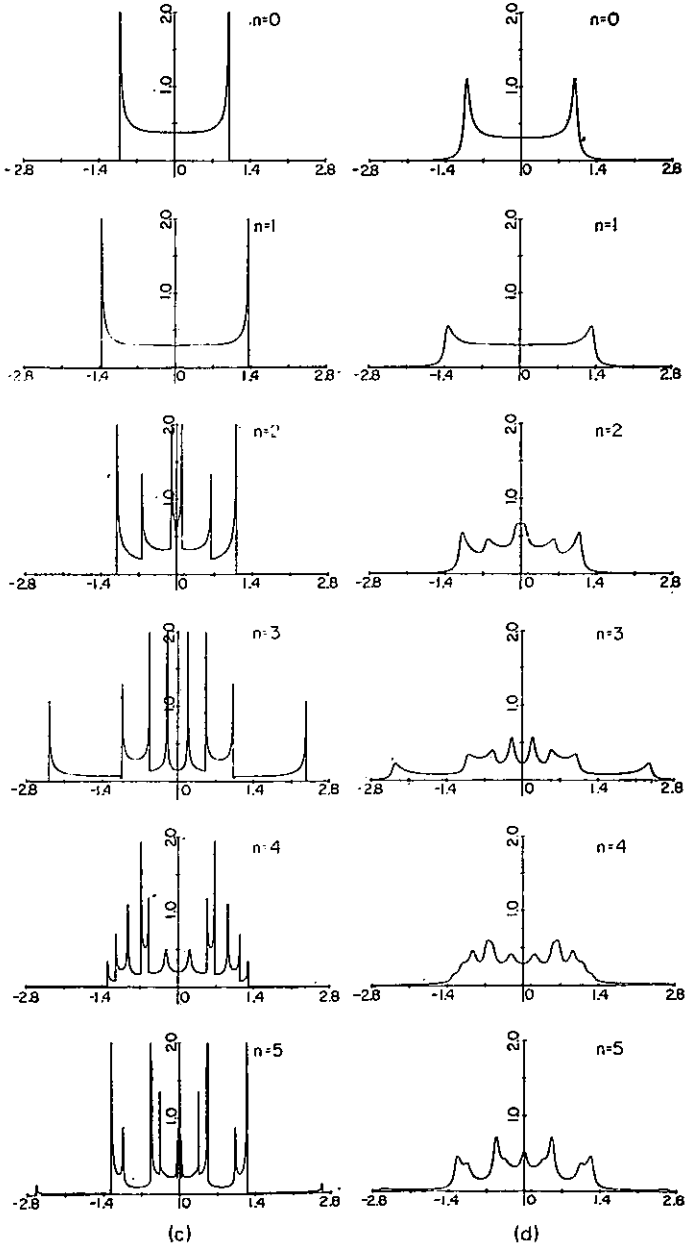


Fig. 62c,d.

W_1, W_2 and W_3 are shown in fig. 63 for values of $\hbar$ shown as the areas of squares on the right. The classical Wigner maps ($\hbar = 0$) are shown in fig. 63e. As $\hbar$ increases (figs. 63d,c,b,a), more and more of the classical complexity is smoothed away, and the extreme quantal functions (fig. 63a) show little resemblance to their classical counterparts (fig. 63e).

These classical curve maps and Wigner contour maps show strong resemblances to the morphologies developed when cream spreads in coffee, and indeed both have a common origin in generic measure-preserving maps. In quantum mechanics the development of complexity on infinitely fine scales is inhibited by the finite value of $\hbar$; in fluid mechanics the fine scale of turbulent mixing is determined by viscosity (see for example the vorticity contour maps computed by Zabusky [114]).

It is worth pointing out that these computations of the evolution of states under quantum maps show Planck's constant playing an unfamiliar role. Instead of adding quantum detail to a smooth underlying classical structure (e.g. the initial curve Σ), $\hbar$ acts in the opposite way when $t \to \infty$, to impose a quantum smoothing onto classical structures with infinitely fine detail (e.g. fully-developed whorls and tendrils).

To my knowledge there are no analytical results giving information about $\psi(q)$ for small $\hbar$ as $t \to \infty$, i.e. in the regime where the $\Sigma - \psi$ association (3.7) no longer holds (although it is plausible that ψ will resemble the irregular wave functions discussed in section 3.4). But by estimating the corrections to eq. (3.7) in the case of a classically chaotic system, Shepelyansky [111] (see also ref. [106]) concludes that the new regime begins at a transition time $t_{\max}$ of order $\hbar^{-1}$; for $t < t_{\max}$, therefore, ψ should be associated with Σ, even though this curve can be very complicated and give rise to many momenta $p_i(q)$ giving contributions to ψ. Associated with these many contributions are many caustics (cf. fig. 62d for $n = 5$), near which eq. (3.7) fails and must be modified as explained in section 3,2 and references therein. Shepelyansky [111] also estimates that for chaotic systems these caustics, although numerous, are greatly outnumbered by non-degenerate contributions P_i for which eq. (3.7) is valid, so that caustics give an insignificant net contribution to ψ.

6.2. Stationary states

So far we have used quantum maps to study evolving states. Can they be used to study stationary states? Unitary operators have eigenvalues on the

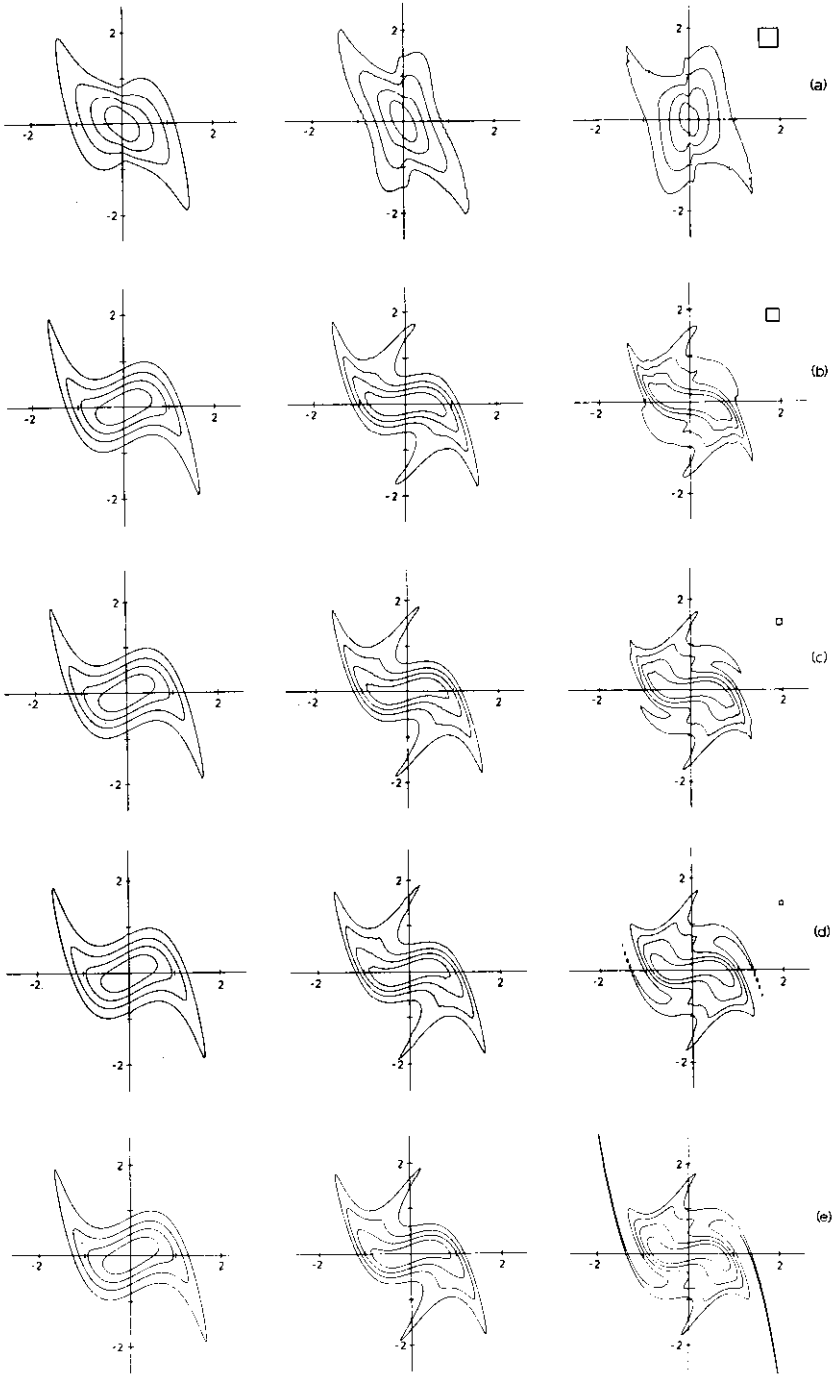


Fig. 63.

unit circle, and so the eigenstates $|\phi>$ must iterate to themselves under eq. (6.1), apart from a phase factor. Thus

$$\hat{U}|\phi> = e^{i\alpha}|\phi>. \tag{6.11}$$

For maps of the type (6.5), the "eigenangle" α may be written as

$$\alpha = \tilde{E}T/\hbar, \tag{6.12}$$

where $\tilde{E}$ is a "quasi-energy" eigenvalue (as described by Zeldovich [115]). The problem with using such maps to study the asymptotics of eigenstates is that little seems to be known about the spectra generated by integral equations of the type (6.6). Are they continuous or discrete? Does the nodal structure of the eigenfunctions (or their real and imaginary parts if they should be essentially complex) bear any relation to the spectrum? For what sort of potential $V(q)$ does eq. (6.6) exhibit bound states? We do not know the answers to these questions.

But there is another sort of quantum map whose spectra can be studied analytically, namely those obtained by quantizing maps M of the "Arnol'd cat" type [eq. (2.22) and fig. 24]. These are linear maps on a phase space consisting of the unit torus; their general form is

$$M: \begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} q_n \\ p_n \end{pmatrix} \text{mod. } 1 \tag{6.13}$$

For M to be continuous, the elements T_{ij} must be integers; for M to be area-preserving, $|\det T_{ij}|$ must equal unity; and if $|T_{11}+T_{22}|>2$ the transformation of area elements is hyperbolic and M is ergodic on the torus. This is a rather artificial class of system (although an optical analogue is in principle realizable) whose interest in the present context lies in the fact that it is the only ergodic system for which the exact quantum mechanics is at all well-understood. The following treatment is a summary of the paper by Hannay and Berry [116], but I also draw attention to a study by Izraelev and Shepelyansky [117] of a quantum map on the torus arising not from a linear transformation but from the kicked pendulum.

Quantization has two aspects. The first is purely kinematic. Because phase space is the unit torus, coordinate wave functions $\psi(q)$ must have unit period in q, implying that momentum wave functions $\bar{\psi}(p)$ must consist of a series of delta functions at $p = nh$, where n is an integer. But $\bar{\psi}(p)$ must also have unit period in p, so that $\psi(q)$ must consist of delta functions at $q = m$h, where m is an integer. These conditions are mutually consistent

only if the torus area is an integer multiple of Planck's constant, which in this case (unit torus) gives the strange condition

$$h = 1/N, \tag{6.14}$$

where N is an integer. Measurements of q and p can yield only values in the "quantum lattice"

$$q = Q/N,\ p = P/N \qquad (1 \leq Q, P \leq N), \tag{6.15}$$

with Q and P integers. The semiclassical limit is $N \to \infty$; as it is reached, h vanishes by discontinuous steps and q and p become continuous variables. For a technical reason [116], N will henceforth be restricted to be odd.

The second aspect of quantization is the introduction of dynamics, in the form of a unitary operator $\hat{U}$ corresponding to M. In the discrete representation (6.15), $\hat{U}$ propagates states $\psi(Q)$ according to the analogue of eq. (6.6), namely

$$\psi_{n+1}(Q) = \sum_{Q'=1}^{N} U_{QQ'} \psi_n(Q'). \tag{6.16}$$

$\hat{U}$ is constructed by "periodizing" the (simple) unitary operator corresponding to M acting as a map in the original "untorized" infinite q, p plane. It turns out that $\hat{U}$ preserves periodicity in both $\psi(Q)$ and $\bar{\psi}(P)$ only if the elements T_{ij} in eq. (6.13) have the form

$$M: \begin{pmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{pmatrix} \text{or} \begin{pmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{pmatrix}. \tag{6.17}$$

This excludes the familiar cat map (2.22) but includes for example, the simple ergodic map with matrix

$$M: \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}, \tag{6.18}$$

The explicit construction of $\hat{U}$ for an arbitrary matrix in the "permitted" class (6.17) is an intricate process [116] involving the Gauss sums of number theory, but for the particular case (6.18) the evolution operator takes the simple form

$$U_{QQ'} = \frac{1}{N^{1/2}} \exp\left[\frac{-i\pi}{4} + \frac{2\pi i}{N}\left(Q^2 + Q'^2 - QQ'\right)\right]. \tag{6.19}$$

According to eq. (6.11), eigenstates of $\hat{U}$ map into themselves under eq. (6.16) apart from phase factors $\exp(i\alpha)$, given by

$$\det_{QQ'}(U_{QQ'} - e^{i\alpha}\delta_{QQ'}) = 0. \tag{6.20}$$

Because $U_{QQ'}$ is an $N \times N$ matrix, there are N eigenvalues $\alpha_j (1 \leq j \leq N)$, and it is interesting to ask how they are distributed round the unit circle, especially in the semiclassical limit $N \to \infty$.

To answer this question, consider the effect of the *classical* map M on points in the *quantum* lattice (6.15). Being rational, these points map around closed orbits ("cycles") in contrast to generic points on the torus, which have irrational coordinates and so never return to their starting points. For each N, there will be a number $n(N)$ of iterations after which every rational point with denominator N will have completed at least one cycle. $n(N)$ is the *period* of the map, defined as the smallest number satisfying

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}^{n(N)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}_{\text{mod. } N}, \tag{6.21}$$

and given by the lowest common multiple of the lengths of cycles of points in the quantum lattice.

Now, for these linear maps the corresponding Hamiltonians are quadratic. For example, eq. (6.18) corresponds to

$$H = \frac{\operatorname{arsinh} 3^{1/2}}{2}\left(\frac{p^2}{3^{1/2}} - q^2 \cdot 3^{1/2}\right), \tag{6.22}$$

acting for unit time. It follows from this that certain quantum-mechanical quantities (e.g. Wigner's function) evolve classically. Therefore after $n(N)$ iterations of the quantum map, wave functions will have returned to their original values, apart from a possible phase factor, i.e.

$$\hat{U}^{n(N)} = \hat{1} \exp[i\sigma(N)], \tag{6.23}$$

where σ is in general unknown. Thus the eigenvalues α must be multiples of $2\pi/n(N)$, apart from a shift, i.e.

$$\alpha_j = \frac{2\pi m_j}{n(N)} + \frac{\sigma(N)}{n(N)} \qquad \begin{pmatrix} 1 \leq j \leq N \\ 1 \leq m_j \leq n(N) \end{pmatrix}. \tag{6.24}$$

The spectrum of $\hat{U}$ therefore consists of N eigenangles distributed among $n(N)$ possible eigenlevels. $n(N)$ is an extremely erratic function, defined

number-theoretically by eq. (6.21). Sometimes $n<N$, in which case some levels must be multiply occupied, and sometimes $n>N$, in which case some levels must be empty. In the map (6.18), for instance,

$$n(1)=1,\ n(3)=6,\ n(5)=3,\ n(7)=8,\ n(9)=18,\ n(11)=10. \qquad (6.25)$$

Number-theoretic arguments, together with numerical experiments, suggest that in a suitably-defined asymptotic sense

$$n(N)\sim CN \qquad \text{as } N\to\infty. \qquad (6.26)$$

where C is constant, but the limit is approached very slowly and erratically. Whatever the value of $n(N)$, the angles seem to be fairly uniformly distributed over their possible sites, behaviour reminiscent of the level repulsion discussed in section 4.3 for generic systems of more conventional type. Computations also show that the Wigner functions for individual eigenstates spread all over the quantum lattice in accordance with the semiclassical eigenfunction hypothesis (section 3) applied to these ergodic systems, and do not concentrate about individual cycles (closed orbits).

The erratic behaviour of $n(N)$ as $N\to\infty$ shows that the semiclassical mechanics of this chaotic system is very different from that of an integrable system whose eigenvalues are given by torus quantization (4.1) as smooth functions of $\hbar$. In fact decreasing $\hbar$, and hence increasing N, causes the spectrum to depend on the iteration of points in an ever-finer quantum lattice, i.e. on an increasingly intricate cycle structure. Here we again see $\hbar$ playing the same role as in the evolution of nonstationary states over long times, namely obscuring an underlying classical structure which has infinite complexity.

Acknowledgments

I thank the Institute for Theoretical Physics in the University of Utrecht for hospitality whilst these lectures were written. The work was not supported by any military agency.

References

[1] I.C. Percival, Adv. Chem. Phys. 36 (1977) 1-61.
[2] J.J. Duistermaat, Commun. Pure. Appl. Math. 27 (1974) 207-281.

[3] M.V. Berry and K.E. Mount, Rep. Prog. Phys. 35 (1972) 315–397.
[4] J. Ford, in: Fundamental Problems in Statistical Mechanics, ed., E.G.D. Cohen, vol. III (North-Holland, Amsterdam, 1975) pp. 215–255.
[5] V.I. Arnol'd, Mathematical Methods of Classical Dynamics (Springer, New York, 1978).
[6] M.V. Berry, in: Topics in Nonlinear Dynamics, ed., S. Jorna, Am. Inst. Conf. Proc. 46 (1978) 16–120.
[7] R. Helleman, in: Fundamental Problems in Statistical Mechanics, ed., E.G.D. Cohen, vol. V (North-Holland, Amsterdam, 1980) pp. 165–233.
[8] J.L. Lebowitz and O. Penrose, Physics Today 26 (February 1973) 23–29.
[9] M.V. Berry, Eur. J. Phys. 2 (1981) 91–102.
[10] Ya. G. Sinai, Russ. Math. Surv. 25 (1970) 137–189.
[11] L.A. Bunimovich, Funct. Anal. Appl. 8 (1974) 254–255.
[12] L.A. Bunimovich, Commun. Math. Phys. 65 (1979) 295–312.
[13] M.C. Gutzwiller, J. Math. Phys. 14 (1973) 139–152; 18 (1977) 806–823.
[14] A.N. Kolmogorov, Dokl. Akad. Nauk 98 (1954) 527–530.
[15] V.I. Arnol'd, Russ. Math. Surv. 18, No. 5 (1963) 13–39; No. 6, 61–196.
[16] J. Moser, Nachr. Akad. Wiss. Göttingen 1 (1962) 1–20.
[17] A.M. Ozorio de Almeida and J.H. Hannay, Ann. Phys. (N.Y.) 138 (1982) 115–154.
[18] M.V. Berry, Singularities in Waves and Rays, Lectures at Les Houches 1980, Session XXXV, Physics of Defects, eds., R. Balian, M. Kléman, J.P. Poirier (North-Holland, Amsterdam, 1981) p. 453.
[19] V.F. Lazutkin, Math. Izv. USSR. 37 (1973) 186–216.
[20] G. Benettin and J.M. Strelcyn, Phys. Rev. A17 (1978) 773–785.
[21] P.J. Richens and M.V. Berry, Physica 2D (1981) 495–512.
[22] A.N. Zemlyakov and A.B. Katok, Math. Notes (USSR) 18 (1975) Nos. 1–2, 760–764.
[23] D.W. Noid and R.A. Marcus, J. Chem. Phys. 67 (1977) 559–567.
[24] G. Casati and J. Ford, J. Comp. Phys. 20 (1976) 97–109.
[25] B.B. Mandelbrot, in: Nonlinear Dynamics, ed., R.H.G. Helleman, Ann. N.Y. Acad. Sci. 357 (1980) 249–259.
[26] V.P. Maslov and M.V. Fedoriuk, Semiclassical Approximation in Quantum Mechanics (Reidel, Dordrecht, 1981). (original Russian edition 1965).
[27] J.H. van Vleck, Proc. Natl. Acad. Sci. USA 14 (1928) 178–188.
[28] J.B. Keller, Ann. Phys. N.Y. 4 (1958) 180–188.
[29] Yu. A. Kravtsov, Sov. Phys. Acoust. 14 (1968) 1–17.
[30] V. Guillemin and S. Sternberg, Geometric Asymptotics, Am. Math. Soc. Surv. No. 14 (Providence, USA, 1977).
[31] J.P. Eckmann and R. Sénéor, Arch. Rat. Mech. Anal. 61 (1976) 153–173.
[32] P.A.M. Dirac, The Principles of Quantum Mechanics (Clarendon Press, Oxford, 1947).
[33] M.V. Berry and C. Upstill, Prog. Opt. 18 (1980) 257–346.
[34] J. Hayden and E.C. Zeeman, Math. Proc. Camb. Phil. Soc. 89 (1981) 193–200.
[35] V.I. Arnol'd, Funct. Anal. Appl. 1 (part 1 1967) 1–13.
[36] A. Einstein, Ver. Deut. Phys. Ges. 19 (1917) 82–92.
[37] I.C. Percival, J. Phys. B6 (1973) L229–232.
[38] E.P. Wigner, Phys. Rev. 40 (1932) 749–759.

[39] H.J. Groenewold, Physica 12 (1946) 405–460.
[40] J.E. Moyal, Proc. Camb. Phil. Soc. 45 (1949) 99–124.
[41] T. Takabayasi, Proc. Theor. Phys. Jap. 11 (1954) 341–373.
[42] G.A. Baker, Jr. Phys. Rev. 109 (1958) 2198–2206.
[43] K.S.J. Nordholm and S.A. Rice, J. Chem. Phys. 61 (1974) 203–223, 768–779.
[44] A. Voros, Ann. Inst. H. Poincaré XXIV (1976) 31–90.
[45] M.V. Berry, Phil. Trans. Roy. Soc. Lond. 287 (1977) 237–271.
[46] M.V. Berry and N.L. Balázs, J. Phys. A:12 (1979) 625–642.
[47] E.J. Heller, J. Chem. Phys. 65 (1976) 1289–1298; 67 (1977) 3339–3351.
[48] N.L. Balázs and G.G. Zipfel, Jr. Ann. Phys. N.Y. 77 (1973) 139–156.
[49] H.J. Korsch, J. Phys. A:12 (1979) 811–823.
[50] A.M. Ozorio de Almeida, Physica 110 A (1982) 501–517.
[51] M.V. Berry and F.J. Wright, J. Phys. A:13 (1980) 149–160.
[52] M.V. Berry, J. Phys. A:10 (1977) 2083–2091.
[53] A. Voros, in: Stochastic Behaviour in Classical and Quantum Hamiltonian Systems, eds., G. Casati, J. Ford, Lecture notes in Physics 93 (Springer, Berlin, 1979) pp. 326–333.
[54] J.S. Hutchinson and R.E. Wyatt, Chem. Phys. Lett. 72 (1980) 378–384.
[55] A.I. Shnirelman, Usp. Mat. Nauk 29, No. 6 (1974) 181–182.
[56] M.V. Berry, J. Phys. A:10 (1977) 2061–2081.
[57] S.W. McDonald and A.N. Kaufman, Phys. Rev. Lett. 42 (1979) 1189–1191.
[58] R.M. Stratt, N.C. and W.H. Miller, J. Chem. Phys. 71 (1979 3311–3322.
[59] K. Uhlenbeck, Am. J. Math. 98 (1976) 1059–1078.
[60] P. Pechukas, J. Chem. Phys. 57 (1972) 5577–5594.
[61] R. Courant and D. Hilbert, Methods of Mathematical Physics, vol. I (Interscience, New York, 1953) pp. 451–456.
[62] N. Fröman and P.O. Fröman, JWKB approximation (North-Holland, Amsterdam, 1965).
[63] W. Eastes and R.A. Marcus, J. Chem. Phys. 61 (1974) 4301–4306.
[64] D.W. Noid and R.A. Marcus, J. Chem. Phys. 62 (1975) 2119–2124.
[65] D.W. Noid, M.L. Koszykowski and R.A. Marcus 71 (1979) 2864–2873.
[66] I.C. Percival and N. Pomphrey, Mol. Phys. 31 (1976) 97–114.
[67] C. Jaffé and W.P. Reinhardt, J. Chem. Phys. 71 (1979) 1862–1869.
[68] S. Chapman, B. Garrett and W.H. Miller, J. Chem. Phys. 64 (1976) 502–509.
[69] M.V. Berry and M. Wilkinson, submitted to Proc. Phys. Soc. Lond. (1983).
[70] M.V. Berry, Ann. Phys. N.Y. 131 (1981) 163–216.
[71] J.M. Ziman, Solid State Phys. 26 (1971) 1–101.
[72] M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions (Washington, U.S. Nat. Bur. Standards, 1964).
[73] J. von Neumann and E.P. Wigner, Phys. Z. 30 (1929) 467–470.
[74] E. Teller, J. Phys. Chem. 41 (1937) 109–116.
[75] H.C. Longuet-Higgins, Proc. Roy. Soc. A344 (1975) 147–156.
[76] M. Pinsky, SIAM J. Math. Anal. 11 (1980) 819–827.
[77] R.A. Marcus, in: Nonlinear Dynamics, ed., R.H.G. Helleman, Ann. N.Y. Acad. Sci. 357 (1980) pp. 169–182.
[78] C.F. Porter, ed., Statistical Theory of Spectra: Fluctuations (Academic Press, New York, 1965).

[79] G. Casati, F. Valz-Gris and I. Guarneri, Nuovo Cimento Lett. 28 (1980) 279–182.
[80] G.M. Zaslavskii, Sov. Phys. JETP 46 (1977) 1094–1098.
[81] M.V. Berry and M. Tabor, Proc. Roy. Soc. A356 (1977) 375–394.
[82] V.L. Pokrovskii, JETP Lett. 4 (1966) 96–99.
[83] S.A. Molcanov, Commn. Math. Phys. 78 (1981) 429–446.
[84] M.C. Gutzwiller, J. Math. Phys. 8 (1967) 1979–2000.
[85] M.C. Gutzwiller, J. Math. Phys. 10 (1969) 1004–1020.
[86] M.C. Gutzwiller, J. Math. Phys. 11 (1970) 1791–1806.
[87] M.C. Gutzwiller, J. Math. Phys. 12 (1971) 343–358.
[88] M.C. Gutzwiller, in: Path Integrals and Their Application in Quantum Statistical and Solid State Physics, eds., G.J. Papadopoulos, J.T. Devreese (Plenum, New York, 1978) pp. 163–200.
[89] R. Balian and C. Bloch, Ann. Phys. N.Y. 60 (1970) 401–447.
[90] R. Balian and C. Bloch, Ann. Phys. N.Y. 64 (1971) 271–307.
[91] R. Balian and C. Bloch, Ann. Phys. N.Y. 69 (1972) 76–160.
[92] R. Balian and C. Bloch, Ann. Phys. N.Y. 85 (1974) 514–545.
[93] L.D. Landau and E.M. Lifshitz, Quantum Mechanics (Nonrelativistic Theory) Pergamon, Oxford, 1965).
[94] H.P. Baltes and E.R. Hilf, Spectra of Finite Systems (B – I Wissenschaftsverlag, Mannheim, 1978).
[95] M.V. Berry and M. Tabor, Proc. Roy. Soc. A349 (1976) 101–123.
[96] M.V. Berry and M. Tabor, J. Phys. A:10 (1977) 371–379.
[97] R. Rajaraman, Phys. Rep. 21 (1975) 227–313.
[98] C. DeWitt-Morette, A. Maheshwari and B. Nelson, Phys. Rep. 50 (1979) 255–372.
[99] W.H. Miller, J. Chem. Phys. 63 (1975) 996–999.
[100] V.I. Arnol'd and A. Avez, Ergodic Problems of Classical Mechanics (Benjamin, Reading, MA, 1968).
[101] M.C. Gutzwiller, Phys. Rev. Lett. 45 (1980) 150–153.
[102] J. Chazarin, Inv. Math. 24 (1974) 65–82.
[103] H.P. McKean, Comm. Pure Appl. Math, 25 (1972) 225–246.
[104] D. Hejhal, The Selberg Trace Formula for PSL(2,R), Lecture Notes in Mathematics, Vol. 548 (Springer, New York, 1976).
[105] D.S. Ornstein, in Proc. 6th IUPAP conf on Stat. Mech., eds., S.A. Rice, K.F. Freed, J.C. Light (University of Chicago Press, 1972).
[106] B.V. Chirikov, F.M. Izraelev and D.L. Shepelyansky, Preprint from Institute of Nuclear Physics, Novosibirsk.
[107] J.S. Hutchinson and R.E. Wyatt, Phys. Rev. A23 (1981) 1567–1584.
[108] G. Casati, B.V. Chirikov, J. Ford and F.M. Izraelev, in: Stochastic Behaviour in Classical and Quantum Hamiltonian Systems, eds., G. Casati, J. Ford, Lecture Notes in Physics 93 (Springer, Berlin, 1979) pp. 334–352.
[109] M.V. Berry, N.L. Balázs, M. Tabor and A. Voros, Ann. Phys. N.Y. 122 (1979) 26–63.
[110] M.V. Berry, in: Nonlinear Dynamics, ed., R.H.G. Helleman, Ann. N.Y. Acad. Sci. 357 (1980) 183–203.
[111] D.L. Shepelyansky, Sov. Phys. Dokl. 26 (1981) 80–82.
[112] H.J. Korsch and M.V. Berry, Physica 3D (1981) 627–636.
[113] J.B. Keller and S.I. Rubinow, Ann. Phys. N.Y. 9 (1960) 24–75.

[114] N.J. Zabusky, in: Proc. Orbis Scientiae (Miami) on the Significance of Nonlinearity in the Natural Sciences, eds., A. Perlmutter, L.F. Scott (Plenum, New York, 1977) pp. 145–205.
[115] Ya.B. Zel'dovich, Sov. Phys. JETP 24 (1967) 1006–1008.
[116] J.H. Hannay and M.V. Berry, Physica 1D (1980) 267–290.
[117] F.M. Izraelev and D.L. Shepelyansky, Sov. Phys. Dokl. 24 (1979) 996–8.
[118] G.M. Zaslavsky, Phys. Rep. 80 (1981) 157–250.
[119] A.M. Ozorio de Almeida, Ann. Phys. (N.Y.) (1983) in press.
[120] M.V. Berry, J.H. Hannay and A.M. Ozorio de Almeida, Physica D (1983) in press.