Reprinted from ANNALS OF PHYSICS
All Rights Reserved by Academic Press, New York and London

Vol. 131, No. 1, January 1981
Printed in Belgium

Quantizing a Classically Ergodic System: Sinai's Billiard and the KKR Method

M. V. BERRY

H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

Received April 29, 1980

Sinai's "billiards on a torus," i.e., free motion of a particle in a plane amongst reflecting discs of radius R centred on points of the unit square lattice, is a classically ergodic system with two freedoms, parametrized by R. Quantal energy levels E_n are given by the vanishing of the Korringa–Kohn–Rostoker (KKR) determinant of solid state theory. This gives a rapid computational scheme for computing E_n as functions of R. Except for the integrable case $R = 0$, no degeneracies are found, illustrating the theorem that two parameters, not one, are required to make levels cross in a generic system. The same theorem leads to the prediction that the probability distribution of the spacings S of neighbouring levels is $\mathcal{O}(S)$ as $S \to 0$, in good agreement with computation. The KKR determinant is transformed analytically to give the level density $d(E)$ semiclassically (i.e., as $\hbar \to 0$) as the sum of a steady contribution $\bar{d}(E)$ and an oscillatory contribution $d_{\mathrm{osc}}(E)$. $\bar{d}$ is $\mathcal{O}(\hbar^{-2})$ and is given by the Weyl "area" formula plus "edge," "corner" and "curvature" corrections, in excellent agreement with computation. d_{osc} is given by a sum over classical closed orbits (all unstable). Nonisolated closed orbits (not hitting discs) contribute terms with $\mathcal{O}(\hbar^{-3/2})$ to d_{osc}, while isolated closed orbits (bouncing between discs) contribute terms with $\mathcal{O}(\hbar^{-1})$ to d_{osc}. The isolated orbits are vastly more numerous than the nonisolated orbits and their contributions cannot be neglected. As a means of calculating the individual E_n (rather than the smoothed spectrum), the KKR method is much more efficient than the classical path sum.

In the heaven of Indra is said to be a huge net, which bears at each intersection of its cords a reflecting pearl. Each pearl by imaging those immediately adjacent to it images the infinity of pearls in the outer spaces of the whole net, for each pearl is the bearer of its neighbour's image.

From the Avatamsaka Sutra

1. INTRODUCTION

A major unsolved problem in semiclassical quantum mechanics is the asymptotic approximation (as $\hbar \to 0$) of the energy levels of bound systems whose classical motion is nonintegrable. For a classical system with N freedoms, integrability means that there exist N constants of motion (including the Hamiltonian) expressible as functions on phase space, so that orbits are confined to N-dimensional manifolds in the $2N$-dimensional phase space; by a theorem of Arnol'd [1], these manifolds are N-tori. Whenever tori exist, the action integrals around their irreducible cycles can be quantized according to the Bohr–Sommerfeld rules as generalized by Einstein [2],

0003-4916/81/010163-54$05.00/0

Keller [3] and Maslov [4]. In a nonintegrable system (for which N must always exceed unity) there exist regions of phase space where tori do not exist; in these regions, orbits explore more than N dimensions over infinite times, and moreover do so in an unstable and irregular ("stochastic") manner. *Ergodic* systems represent the extreme case of nonintegrability: there are no tori at all, and almost all orbits explore the neighbourhood of every point on the $(2N - 1)$-dimensional surface of constant energy in phase space. The following question naturally arises: what is the nature of the semiclassical energy level spectrum for a classically ergodic system ?

My purpose is to discuss this question in the context of a particular system chosen to satisfy two conditions: first, the classical motion is ergodic, and second, the energy levels are given by an exact quantal formalism which can be explored analytically in considerable depth. The system is the so-called "Sinai torus billiard table" [5] for $N = 2$; it is described in Section 2. The quantal formalism is based on the "Korringa–Kohn–Rostoker" (KKR) method [6, 7] of solid state physics; it is described in Section 3. Application of the KKR method gives the energy levels as the zeros of a determinant which is in principle of infinite order but in practice rapidly convergent and ideally suited to numerical computations. Previously, the following ergodic systems have been studied in connection with quantization: (i) the anisotropic Kepler problem, for which Gutzwiller [12] calculated a single closed orbit and discussed its effect on the spectrum, (ii) the stadium, for which McDonald and Kaufman [32] carried out a pioneering numerical study of the distribution of eigenvalues and the nodal structure of eigenfunctions, and (iii) unstable linear maps (of "Arnold's cat" type) on a torus phase space, a system with one freedom for which Hannay and Berry [36] have carred out a detailed analytical and numerical study of the eigenvalues and eigenfunctions.

Sinai's billiard is actually a family of ergodic systems, labelled by a parameter R. Therefore the opportunity arises to study the energy levels as functions of R. In the special case $R = 0$, the system is integrable, and the spectrum exhibits degeneracies of a number-theoretic nature which leads to a very strange classical limit. But the most dramatic effect is the complete absence of degeneracies when $R \neq 0$; as discussed in Section 4, this is expected on the basis of a theorem of von Neumann and Wigner [8], and Teller [17] and contrasts with the behaviour of integrable systems. There are, however, very many near-degeneracies, whose interpretation in Section 5 makes it possible to obtain an analytical description of the so-called "level repulsion."

In Section 6 the KKR determinant is transformed analytically and approximated asymptotically to give the Weyl formula [9] for the mean level density. The level density itself is obtained from this by the addition of an infinite series of oscillatory correction terms, derived in Section 7 and discussed in Section 8. Each oscillatory term corresponds to a closed classical orbit. All closed orbits are unstable, but they may be isolated or nonisolated; as previously recognized by Balian and Bloch [10], this distinction reflects important analytical differences in the oscillatory contributions. Nonisolated orbits give rise to contributions closely similar to those calculated by Berry and Tabor [11] for closed orbits lying on the tori of integrable systems. Isolated orbits give contributions of the type previously calculated by Gutzwiller [12]. Although

the closed path sum is very helpful in understanding the clustering properties of the spectrum, it is far less efficient than the KKR method for calculating individual levels; this is shown in Section 8 by estimating the number of terms involved in the rival methods.

To make the paper as readable as possible I have relegated the more technical arguments to a series of appendices. Two other appendices (D and I) concern topics of more general interest arising out of the work described here, namely a proof of the sign change of wave functions taken round a circuit in parameter space enclosing a degeneracy, and an estimate of the strength of clustering of resonant frequencies of auditoriums with fractal walls.

2. Billiards on a Torus

The system under study is a classical particle moving freely on a two-dimensional torus on which there is a circular specularly reflecting obstacle (Fig. 1a). It is easier to understand the motion by representing the torus as a unit square with opposite edges identified (Fig. 1b) in the Euclidean xy plane, with the disc at its centre; the disc has radius R, so that $0 \leqslant R \leqslant \frac{1}{2}$. Particle paths are sequences of straight line

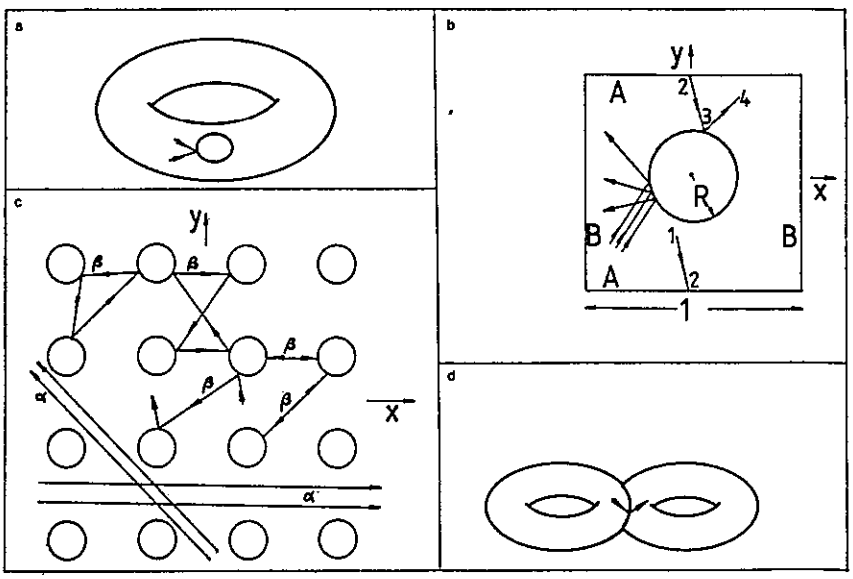

Fig. 1. Representations of Sinai's billiard: (a) as motion on a torus with reflecting disc; (b) as straight line motion on unit square with periodic boundary conditions and with a reflecting disc, and showing defocusing of beams of particles; (c) as free motion amongst reflecting discs centred on points of the unit square lattice, and showing closed orbits of nonisolated (α) and isolated (β) types; (d) as geodesic motion on a double torus creased along a line of infinitely negative Gaussian curvature.

segments on the square, such as 1234 on Fig. 1b. Alternatively, the torus can be represented by the infinite periodic xy plane, in which the particle moves in straight lines apart from specular reflections at discs centred on the points of the unit square lattice (Fig. 1c).

In this system, conservation of energy just means that the particle's speed is constant, so that position on the constant-energy surface in phase space is defined by x, y and the direction of motion θ. Energy is thus an unimportant parameter in the classical motion, affecting only the time taken for the motion to be carried out. The important features of paths are determined purely geometrically.

The limiting case $R = 0$ is integrable: the particle moves freely on the torus and both velocity components are constants of motion, so that θ never changes. But when $R > 0$ the motion is ergodic: for almost all initial x_0, y_0, θ_0, the particle will eventually come arbitrarily close to all other points x, y, θ; this was proved by Sinai [5]. Ergodicity derives from the continual defocusing of particle beams (Fig. 1b) reflected from the disc. Such defocusing prevents an orbit developing caustics (focal points and lines), which (by virtue of their interpretation as singularities of the projections of phase-space tori onto xy) would be an indication of nonergodicity. Another argument that makes it easy to understand why the motion is ergodic is based on representing the trajectories as geodesics on two tori ("sheets") joined at the central disc (Fig. 1d); reflection at the disc corresponds to moving from one sheet to the other. This double torus can be defined to have zero Gaussian curvature everywhere except along the join, where its curvature is infinitely negative. Now it is known that motion on a surface of negative curvature is ergodic (and moreover exponentially unstable) [13], and so it is reasonable that motion on the double torus is ergodic too.

Ergodicity means that in Sinai's billiard almost all orbits when traversed for infinite time will trace out three-dimensional regions in phase space. The absence of phase space tori rules out orbits tracing out two-dimensional regions in phase space. Moreover, when the energy is not zero there are no equilibrium points that would correspond to zero-dimensional orbits. But there do exist *closed orbits*, tracing out one-dimensional regions in phase space in a periodic manner. Although of measure zero in phase space, these closed orbits are infinitely numerous and indeed dense. All are unstable. They will play an important role in the quantum-mechanical energy spectrum.

It is important to distinguish two types of closed orbit. The first type never strike a disc, and derive their periodicity from the torus-periodicity of the plane. Two such orbits are labelled α on Fig. 1c. These orbits move in directions for which $\tan\theta$ is rational; they are parallel to lattice vectors $\rho = (m, n)$ in the periodized plane, where m, n are relatively prime integers. Only for ρ-values corresponding to directions not obscured by discs do such orbits exist. Because lattice lines perpendicular to ρ have spacing $1/(m^2 + n^2)^{1/2}$, the condition for orbits to exist is

$$2R\rho = 2R(m^2 + n^2)^{1/2} < 1. \tag{2.1}$$

Thus the (1, 0) orbit exists for all R, the (1, 1) orbit exists only if $R < 1/2(2)^{1/2} = 0.354$, and the (1, 2) orbit exists only if $R < 1/2(5)^{1/2} = 0.224$. For each R there are

only a finite number of distinct orbits. But these orbits are not isolated. For each ρ satisfying (2.1) the orbits form a continuous family related by parallel translation perpendicular to ρ. The natural measure M_ρ for a ρ-family is the fractional area of the xy plane covered by the orbits in the family. This is easily calculated to be

$$M_\rho = 1 - 2R\rho = 1 - 2R(m^2 + n^2)^{1/2}. \tag{2.2}$$

The second type of closed orbit do strike discs. Five such orbits are labelled β on Fig. 1c. All these orbits are isolated and unstable. As will be shown, their number and variety increases very rapidly with their length and they are not easily classified.

3. KKR Determinant

Let position in the xy plane be denoted by $\mathbf{r}$, and choose polar coordinates r, ϕ using the centre of a disc as origin. For particles with mass m and energy $\mathscr{E}$, quantum states $\psi(\mathbf{r})$ on the torus satisfy

$$\nabla^2\psi + \frac{2m\mathscr{E}}{\hbar^2}\psi = 0, \tag{3.1}$$

with boundary conditions

$$\psi(R, \phi) = 0; \qquad \psi(\mathbf{r} + \rho) = \psi(\mathbf{r}), \tag{3.2}$$

where ρ is a lattice vector. It will prove convenient to define the wave number k and scaled energy E by

$$\frac{2m\mathscr{E}}{\hbar^2} \equiv k^2 \equiv 4\pi^2 E. \tag{3.3}$$

Therefore the classical limit $\hbar \to 0$ corresponds to $k \to \infty$ or $E \to \infty$. In terms of solid state physics, the problem is that of finding the high-energy bands at the centre of the Brillouin zone.

To determine the allowed energies E by the KKR method [6, 7], the first step is to define the Green function $G(\mathbf{r}, \mathbf{r}')$ by

$$\nabla_r^2 G + k^2 G = \delta(\mathbf{r} - \mathbf{r}'). \tag{3.4}$$

Here and throughout the theoretical development we shall use outgoing boundary conditions, which corresponds to giving k or E a small positive imaginary part. Then

$$G(\mathbf{r}, \mathbf{r}') = \frac{-i}{4} H_0^{(1)}(k \mid \mathbf{r} - \mathbf{r}' \mid), \tag{3.5}$$

where $H_0^{(1)}$ denotes the Hankel function of the first kind of order zero [14].

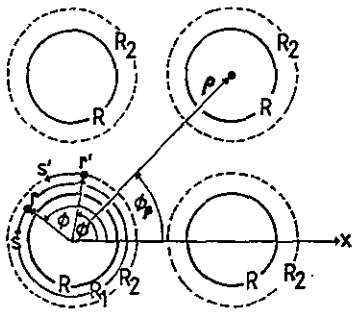

FIG. 2. Notations and geometric constructions associated with the KKR method.

Next, apply Green's identity to the area A consisting of the whole $\mathbf{r}$-plane minus discs of radius R_2 surrounding each lattice point, where $R_2 > R$ (Fig. 2). Thus

$$\iint_A d^2r' \{\psi(\mathbf{r}') \nabla_{\mathbf{r}'}^2 G(\mathbf{r}, \mathbf{r}') - G(\mathbf{r}, \mathbf{r}') \nabla_{\mathbf{r}'}^2 \psi(\mathbf{r}')\}$$

$$= \int_B ds' \, \hat{\mathbf{n}} \cdot \{\psi(\mathbf{r}') \nabla_{\mathbf{r}'} G(\mathbf{r}, \mathbf{r}') - G(\mathbf{r}, \mathbf{r}') \nabla_{\mathbf{r}'} \psi(\mathbf{r}')\}, \tag{3.6}$$

where B is the union of boundaries of all the R_2-dics, s' is arc-length round B and $\hat{\mathbf{n}}$ is the unit normal to B. Now choose $\mathbf{r}$ to lie on a circle centred on $\mathbf{r} = 0$ with radius R_1 satisfying $R < R_1 < R_2$, so that $\mathbf{r}$ does not lie in A. Together with Eqs. (3.1) and (3.4) satisfied by ψ and G, this implies that the left-hand side of (3.6) vanishes. Now let $R_2 \to R_1 \to R$. Then the first boundary condition in (3.2) (ψ vanishes on the R-discs) implies that the first term on the right-hand side of (3.6) vanishes. Thus

$$\int_B ds' \, \hat{\mathbf{n}} \cdot \nabla_{\mathbf{r}'} \psi(\mathbf{r}') \, G(\mathbf{r}, \mathbf{r}') = 0. \tag{3.7}$$

Using the periodicity condition (3.2) on ψ, the integral over B can be reduced to an integral round the R_2-disc centred on the origin, namely,

$$\int_0^{2\pi} d\phi' \, \frac{\partial \psi}{\partial r'} (R, \phi') \sum_{\rho} G(\mathbf{r}, \mathbf{r}' + \rho) = 0. \tag{3.8}$$

Next, $\partial\psi/\partial r'$ and $\sum G$ are expanded in terms of angular eigenfunctions:

$$\frac{\partial \psi}{\partial r'} (R, \phi) = \sum_{l=-\infty}^{\infty} a_l e^{il\phi'} \tag{3.9}$$

and

$$\sum_{\rho} G(\mathbf{r}, \mathbf{r}' + \rho) = \sum_{l=-\infty}^{\infty} \sum_{l'=-\infty}^{\infty} M_{ll'} e^{i(l\phi - l'\phi')}. \tag{3.10}$$

The a_l are unknown coefficients. In Appendix A it is shown that

$$M_{ll'} = \frac{J_l(kR)\,J_{l'}(kR)}{4}\left\{\frac{-iH_l^{(1)}(kR)}{J_l(kR)}\,\delta_{ll'} + \sum_{\rho}{}' (-i)\,H_{l-l'}^{(1)}(k\rho)\,e^{i(l'-l)\phi_\rho}\right\}, \quad (3.11)$$

where the prime denotes summation over all lattice points except $\rho = 0$ and where ϕ_ρ is the polar angle of ρ. Equation (3.8) must hold for all positions of $\mathbf{r}$ round the R_1-circle, that is for all ϕ; with the expansions (3.9) and (3.10), this condition yields

$$\sum_{l'=-\infty}^{\infty} a_{l'} M_{ll'} = 0 \quad (3.12)$$

as the equation determining the coefficients a_l. Therefore the determinant of $M_{ll'}$ must vanish.

Now define the *scattering phase shifts* η_l by

$$\tan \eta_l(E) \equiv J_l(kR)/Y_l(kR), \quad (3.13)$$

and the *structure constants* $S_l(E)$ by

$$S_l(E) \equiv -i \sum_{\rho}{}' H_l^{(1)}(k\rho)\,e^{il\phi_\rho}. \quad (3.14)$$

Then (3.12) and (3.11) imply

$$\det_{ll'}\{(\cot \eta_l(E) - i)\,\delta_{ll'} + S_{l-l'}(E)\} = 0 \qquad (-\infty < l, l' < +\infty) \quad (3.15)$$

as the quantization condition for the billiard eigenvalues E. This is the final result of the KKR method. My derivation of the KKR equation is simpler than the usual ones [6, 7], which are for three-dimensional lattices and general interaction potentials (rather than hard dises). The KKR method in two dimensions has been studied by Ozorio de Almeida [15] in the context of electron microscopy. In physical terms, Eq. (3.15) means that energy levels E are determined by the condition that a wave scattered by the disc at $\rho = 0$ is coherent with waves scattered from and amongst all the other discs, as anticipated in the quotation with which this paper begins.

For later theoretical analysis (Sections 6 and 7), the complex equation (3.15) is the form that will be used. But for numerical calculations a real determinant is preferable, and in Appendix A it is shown that (3.15) is equivalent to

$$\det_{ll'}\{\delta_{ll'} + \tan \eta_l(E)\,S_{l-l'}^{(r)}(E)\} = 0 \qquad (-\infty < l, l' < +\infty), \quad (3.16)$$

where $S_l^{(r)}$ are the real structure constants

$$S_l^{(r)}(E) \equiv \sum_{\rho}{}' Y_l(k\rho)\cos[l\phi_\rho]. \quad (3.17)$$

Because of the fourfold symmetry of the ρ-lattice, the structure constants (real and complex) satisfy

$$S_l(E) = 0 \qquad \text{unless } l/4 \text{ is an integer.} \tag{3.18}$$

Moreover

$$S_{-l}(E) = S_l(E). \tag{3.19}$$

The lattice sums in (3.14) and (3.17) converge conditionally, in an oscillatory manner unsuited to computation. In Appendix B a nontrivial application of the Ewald summation technique is employed to derive the following exponentially convergent representation for $S_l^{(r)}$:

$$\begin{aligned} S_l^{(r)}(E) &\approx \frac{1}{\pi^2}\sum_{\rho}{}' \left(\frac{\rho^2}{E}\right)^{l/2} \frac{\exp[(l/2)(1-\rho^2/E)]}{E-\rho^2}\cos[l\phi_\rho] \qquad (l \neq 0), \\ &\approx \frac{1}{\pi^2}\sum_{\rho} \frac{\exp[Q(1-\rho^2/E)]}{E-\rho^2} - \frac{E_l(Q)}{\pi} \qquad (l = 0), \end{aligned} \tag{3.20}$$

where again ρ denotes a vector in the unit square lattice, Ei is the exponential integral [14], and Q is any positive constant (in numerical calculations I took $Q = 3$). This is not exact, but the error is exponentially negligible for all relevant l and E.

An important part of this work will be a study of degeneracies and near-degeneracies. There are trivial degeneracies associated with (3.16), resulting from the symmetry of the lattice; for example, two states related by reflection in a diagonal of the square in Fig. 1b are degenerate. It proves very convenient to eliminate these degeneracies, by considering only states that are antisymmetric about the x axis and about the square diagonal. After this "desymmetrization," the quantum torus billiard problem becomes equivalent to the determination of the modes of vibration in the enclosure (drum, membrane) of Fig. 3. Waves ψ in such an enclosure must have angular dependence of the form $\sin 4n\phi$ ($n = 1, 2, 3,\ldots$,) [15], so that in expansion (3.9) the coefficients must satisfy

$$a_l = 0 \qquad \text{unless} \quad l = \pm 4n,\ a_{-l} = -a_l\,. \tag{3.21}$$

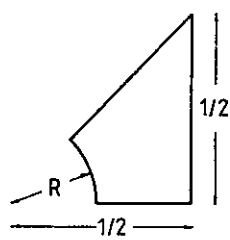

FIG. 3. Enclosure corresponding to the "desymmetrized" Sinai billiard. Eigenstates are determined by the condition that the wave function vanish on the boundary.

Equations (3.12) can now be written as

$$\sum_{n'=1}^{\infty} a_{4n'}(M_{l,4n'} - M_{l,-4n'}) = 0. \tag{3.22}$$

In this set the equations for which $l \neq 4n$, and for which $l = 0$, are vacuous by virtue of (3.18) and (3.19), and the equations for positive and negative l are equivalent. This leads to a determinant considerably simpler than (3.16), namely

$$\det_{nn'}\{\delta_{nn'} + \tan \eta_{4n}(E)[S^{(r)}_{4|n-n'|} - S^{(r)}_{4(n+n')}(E)]\} = 0 \qquad (1 \leqslant n, n' < \infty). \tag{3.23}$$

In the "unperturbed" billiard $R = 0$, it follows from elementary arguments that the eigenstates satisfying the Schrödinger equation (3.1) and boundary conditions (3.2) can be labelled by m, n and have energies

$$E_{mn} = m^2 + n^2 \qquad (-\infty < m, n < +\infty, m = n = 0 \text{ excluded}). \tag{3.24}$$

This correctly reproduces the degeneracy structure of the torus plane waves vanishing at $\mathbf{r} = 0$, namely

$$\begin{aligned} \psi &= \sin 2n\pi x \exp[\pm 2\pi i m y], \\ &= \sin 2n\pi y \exp[\pm 2\pi i m x], \\ &= \sin 2m\pi x \exp[\pm 2\pi i n y], \\ &= \sin 2m\pi y \exp[\pm 2\pi i n x]. \end{aligned} \tag{3.25}$$

The ground state energy is $E = 1$. For the desymmetrized unperturbed billiard (Fig. 3 with $R = 0$), the levels m, n are

$$E_{mn} = m^2 + n^2 \qquad (1 \leqslant n < m < \infty), \tag{3.26}$$

corresponding to the states

$$\psi = \sin 2n\pi x \sin 2m\pi y - \sin 2m\pi x \sin 2n\pi y. \tag{3.27}$$

The ground state energy is $E = 5$. Although not symmetry-degenerate, the states (3.26) have a marvellous number-theoretic degeneracy structure, beginning at $E = 65 = 7^2 + 4^2 = 8^2 + 1^2$; this will be further discussed in Section 4.

Of course these unperturbed results (3.24) and (3.26) must follow from the determinants (3.16) and (3.23) respectively as $R \to 0$. To see that this is so, note first that from (3.13) the phase shifts η_l vanish as $R \to 0$. Therefore both determinants are unity (rather than zero) for all energies except those for which the structure constants diverge. But from the representation (3.20) (see also Appendix B) the singularities of S_l occur precisely at the energies (3.24), so that only these energies can be unperturbed levels. Moreover, in the desymmetrized case the combination of structure constants in (3.23) leads to the following combination of phase factors multiplying the singularity at E_{mn} :

$$\exp[4in_1\phi_{mn}] - \exp[4in_2\phi_{mn}] \qquad (n_1, n_2 \text{ integers}). \tag{3.28}$$

This vanishes when ϕ_{mn} is a multiple of $\pi/4$ (i.e., when $|m| = |n|$ or when m or n is zero), thus cancellating the singularity and leaving only the values (3.26) as possible energies.

When R is small, the corrections to (3.26) can be found by perturbation theory as discussed in Appendix C; it is shown there that nondegenerate levels deviate from (3.26) as R^8, while in the case of number-theoretic degeneracies with multiplicity two one level deviates as R^8 and the other as R^{16}.

For general R the levels are found by computation. The KKR equation (3.23) is a particularly convenient representation for this purpose, for three reasons. Firstly, each element n, n' in the matrix is labelled by two single numbers, in contrast to the secular determinants obtained in conventional basis-set calculations where each matrix element is labelled by two pairs of numbers. Secondly, the KKR determinant converges very rapidly as n or n' increases, by virtue of the exponential decay of $\tan \eta_l(E)$ once the order of the Bessel functions in (3.13) exceeds their argument. This convergence occurs in spite of the divergence of S_l as $l \to \infty$, and depends on the fact that the discs do not overlap (i.e., $R \leqslant 1/2$), as explained in Appendix B. The effective order $n_{\max}$ of the determinant (3.23), as estimated from (3.13) and (3.3), is

$$n_{\max} = \pi R E^{1/2}/2. \tag{3.29}$$

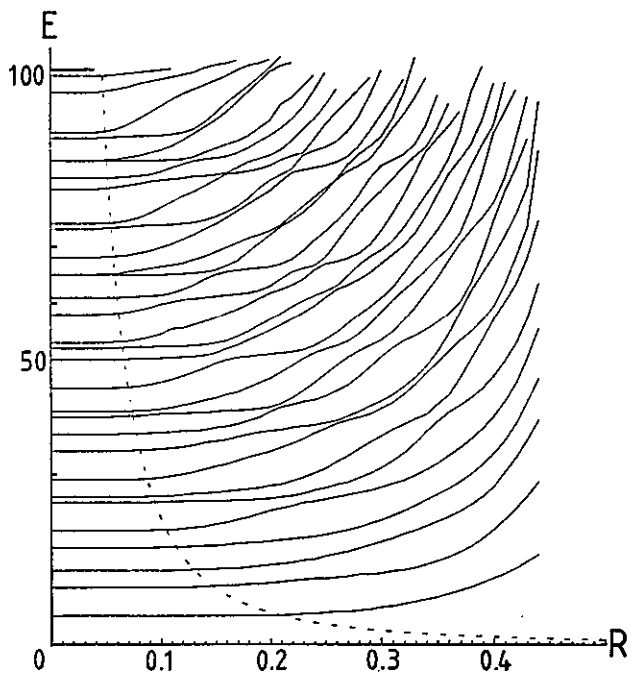

FIG. 4. Eigenvalues E of Sinai's quantum billiard, as functions of the disc radius R. The dashed line indicates the limit of applicability of the perturbation theory of Appendix C.

TABLE I
Sinai Billiard Eigenvalues E_n for Different Values of R

n \ R	0.0	0.1	0.2	0.3	0.4
1	5.00	5.00	5.17	6.53	11.78
2	10.00	10.01	10.82	12.67	20.27
3	13.00	13.03	13.92	18.24	28.05
4	17.00	17.07	18.21	22.07	33.87
5	20.00	20.16	23.52	27.78	40.38
6	25.00	25.08	25.71	31.14	47.34
7	26.00	26.17	27.93	36.87	57.40
8	29.00	29.46	34.08	41.57	59.60
9	34.00	34.41	37.66	42.91	68.51
10	37.00	37.26	40.95	50.68	74.74
11	40.00	40.60	41.79	53.58	77.83
12	41.00	41.37	46.86	58.53	89.91
13	45.00	46.08	51.02	60.92	91.43
14	50.00	50.27	55.22	66.98	98.87
15	52.00	52.31	57.12	69.36	105.52
16	53.00	54.50	61.76	72.85	109.63
17	58.00	59.76	63.07	79.86	114.72
18	61.00	61.38	66.44	84.21	121.05
19	65.00	65.02	72.07	87.93	131.23
20	65.00	66.40	72.72	89.29	132.24

And thirdly, once the structure constants (3.20) have been computed for a given E, they can be stored and used in the evaluation of the determinant for a range of values of R.

In this study the determinant was evaluated over a narrow grid of energies and disc radii in the ranges $0 < E < 100$, for $0 \leqslant R \leqslant 0.45$, and the trajectories of zeros in the E, R plane determined. The spectrum obtained is shown in Fig. 4, and the levels for $R = 0$, 0.1, 0.2, 0.3 and 0.4 are shown in Table I.

4. Degeneracies

The most remarkable feature of the desymmetrized billiard eigenvalues for $R > 0$ as shown on Fig. 4 is the complete absence of degeneracies. One might have expected that as the parameter R varied, pairs of energy level curves $E(R)$ would cross, thus producing degeneracies at particular values of R. The fact that these do not occur is a fine illustration of a theorem of von Neumann and Wigner [8], Teller [17] and Arnol'd [18]. This states that in order to produce level degeneracies in a generic real Hamiltonian system without symmetry, not one but two parameters must be varied. If the parameters are A and B, the two degenerating levels correspond to the two sheets of an elliptic double cone in the (A, B, E) space near the degeneracy, which occurs at a

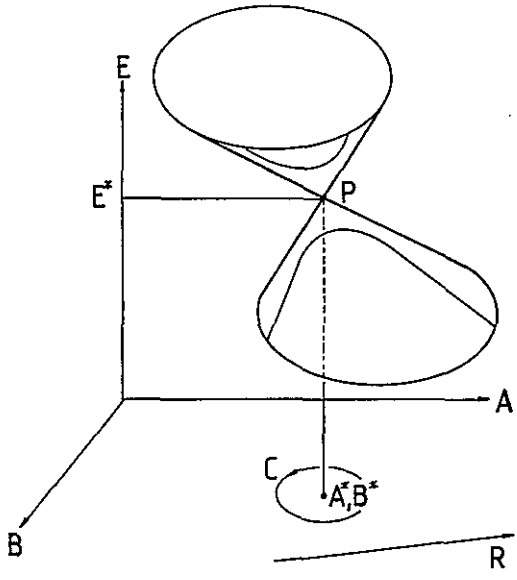

FIG. 5. Two sheets representing energy levels as functions of parameters A, B, with a degeneracy P at A^*, B^* and energy E^*. R is a path missing the degeneracy, and C a path enclosing it.

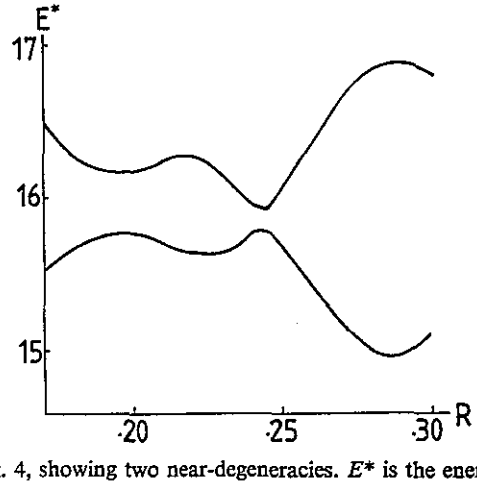

FIG. 6. Detail of Fig. 4, showing two near-degeneracies. E^* is the energy scaled so as to make the mean level spacing unity.

point $P = (A^*, B^*, E^*)$ as shown in Fig. 5. If only one parameter is varied (for example R) this corresponds to traversing a path in A, B space that misses (A^*, B^*), so that the energy level curves are hyperbolae cutting the cone. Many such near-degeneracies can be seen in Fig. 4 and two are shown in detail in Fig. 6; they will be further discussed in Section 5.

The following question naturally arises: how do we know when the point A^*, B^* on Fig. 5 is a true degeneracy, rather than a very close approach of two separate eigenvalue surfaces? The answer is that during a small circuit C in A, B space, surrounding (A^*, B^*), each of the two wave functions changes sign if (A^*, B^*) is a degeneracy of two levels, and the wave functions do not change sign if (A^*, B^*) is not a degeneracy. This result is proved in Appendix D; it has also been stated by Arnol'd [18], and the converse result (that sign change of eigenfunctions round a curve implies at least one degeneracy within the curve) was proved by Longuet-Higgins [19]. It is surprising to learn that the single-valuedness of wave functions in coordinate space does not imply single-valuedness in parameter space.

These two theorems about degeneracies would appear to be quite elementary, yet I have not seen them in any quantum mechanics textbook. The generic behaviour just discussed should not be confused into the Jahn–Teller effect, or with the hybridization of energy bands in crystals, which both concern degeneracies arising through symmetry.

The phenomenon of "repulsion of levels" exhibited by generic systems is very different from what happens in separable and integrable systems. In these cases, degeneracy typically occurs when just one parameter A is varied. To show this, it is sufficient to consider systems with two degrees of freedom. In the separable case, the wave function can be written as a product, and the levels labelled with two quantum numbers m, n:

$$E = E(m, n; A). \tag{4.1}$$

In the space m, n, each lattice point corresponds to a quantum state (Fig. 7). For a given state (m^*, n^*), the energy corresponds to the contour of the function E in (4.1) that passes through (m^*, n^*); usually no other lattice points lie on this contour, so that (m^*, n^*) is nondegenerate. But as A varies the contour passing through (m^*, n^*) typically changes not only its value but its slope, and can cross other lattice points, thus producing "one-parameter" degeneracies, as asserted.

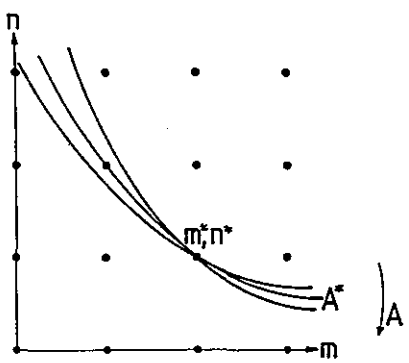

FIG. 7. Space of quantum numbers m, n for a separable system with two freedoms, showing the energy contours passing through a state m^*n^* for different values of a parameter A. There is a degeneracy when $A = A^*$.

An example of this behaviour is the family of rectangular boxes with sides 1, A. With definition (3.3) the eigenvalues are

$$E = m^2 + A^2n^2. \tag{4.2}$$

Degeneracies occur whenever A^2 passes through rational numbers; for example, the states (2, 1) and (1, 3) degenerate when $A^2 = \frac{3}{8}$. Another example is the "critical voltage effect" in transmission electron microscopy of thin metal foils [20]. The parameter A is the voltage of the microscope, the eigenvalues that degenerate are the Bloch wave vector components perpendicular to the foil surface, and the wave functions are the (real) periodic solutions of Schrödinger's equation in the one-dimensional symmetric potential due to planes of atoms.

If the system is integrable, but not separable in the coordinates, a representation in the form (4.1) is still possible at the level of semiclassical approximation; m and n are $\hbar$-multiples of classical action variables. When A varies, degeneracies are again predicted, but now these might be only approximate. My guess is that any splitting would be the result of something like barrier penetration, resulting in minimum level spacing of order $\exp(-1/\hbar)$ instead of zero; these "pseudo-degeneracies" would be negligible as $\hbar \to 0$ if the spectrum is viewed on the scale $\hbar^2$ corresponding to the mean level spacing.

Returning to the desymmetrized Sinai billiard, consider now the case $R = 0$. This is integrable, and the energy levels are given exactly by (3.26). These levels are all integers, making the system very special and not to be considered as representative of the class of integrable or even separable systems. Only those integers that can be represented as the sum of two squares correspond to eigenvalues. The lowest such level is $E = 5 = 2^2 + 1^2$. Some integers can be represented as a sum of squares in more than one way, thus giving rise to degeneracies. The lowest levels with multiplicities two and three are

$$\begin{aligned} E &= 65 = 7^2 + 4^2 = 8^2 + 1^2, \\ E &= 325 = 15^2 + 10^2 = 18^2 + 1^2 = 17^2 + 6^2. \end{aligned} \tag{4.3}$$

Degeneracies are rare for low-lying levels, but as E increases they get more numerous, and come to dominate the spectrum in the classical limit $E \to \infty$.

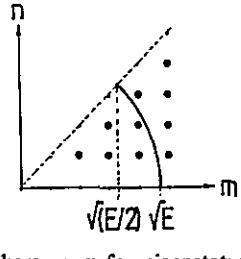

FIG. 8. Space of quantum numbers m, n for eigenstates of the desymmetrized Sinai billiard with $R = 0$.

In showing this, it is helpful to picture the states (3.26) as lattice points in a 45° sector of the m, n plane (Fig. 8). Then the average number of states with energies less than E is the area of a 45° sector of a circle with radius E, namely $\pi E/8$. The mean level density $\bar{d}(E)$ is therefore

$$\bar{d}(E) = \pi/8, \tag{4.4}$$

and the mean spacing between states is $8/\pi$. The degeneracy $D(E)$ of the level at the integer E is the number of ways in which E can be expressed as the sum of two squares, namely

$$D(E) = \sum_{\substack{m^2+n^2=E \\ 1<m<n<\infty}} = \sum_{m=[(E/2-1/4)^{1/2}+1/2]}^{[(E-1)^{1/2}]} \delta_{((E-m^2)^{1/2}-[(E-m^2)^{1/2}]),0}\,, \tag{4.5}$$

where $[x]$ denotes the integer part of x. The maximum possible degeneracy occurs when all terms in this sum are unity, so that

$$D(E) < [E]^{1/2} - [(E/2)^{1/2}] \to 0.293(E)^{1/2} \qquad \text{as} \quad E \to \infty. \tag{4.6}$$

Arbitrarily large degeneracies are permitted by this inequality and in fact must occur as $E \to \infty$. This follows from a theorem, attributed to E. Landau [21], which states that the asymptotic probability $\mathscr{P}(E)$ that E can be written as the sum of two squares is

$$\mathscr{P}(E) \to \frac{C}{(\ln E)^{1/2}} \qquad \text{as } E \to \infty. \tag{4.7}$$

A simple proof of this theorem is given in Appendix E. It implies that as $E \to \infty$ almost all integers are in fact not occupied by levels. But since the mean number of levels in any range ΔE (satisfying $1 \ll \Delta E \ll E$) is $\pi \Delta E/8$ (from 4.4), the mean degeneracy $\bar{D}_{\text{occ}}(E)$ of the energies that are occupied diverges as

$$\bar{D}_{\text{occ}}(E) \to \frac{\pi(\ln E)^{1/2}}{8C}\,; \tag{4.8}$$

the mean spacing between occupied energies diverges in the same way.

Reverting to natural units $\mathscr{E}$ (Eq. (3.3)) the following strange picture emerges of the semiclassical limiting behaviour of the $R = 0$ spectrum. The mean spacing between the energies of states irrespective of degeneracy is $\mathcal{O}(\hbar^2)$ as in any two-dimensional bound system. But the states are increasingly degenerate and separated by increasingly large gaps with lengths

$$\Delta\mathscr{E} \approx \frac{\hbar^2}{2m}(\ln(2m\mathscr{E}/\hbar^2))^{1/2}. \tag{4.9}$$

Such behaviour represents extreme level clustering and contrast sharply with the repulsion of levels when $R > 0$. (Pinsky [39] has discovered that the modes of an equilateral triangle also tend to have infinite degeneracy as $E \to \infty$.)

The divergence $(\ln E)^{1/2}$ is extremely slow, and direct computation of $\mathscr{P}(E)$ showed that the limiting behaviour is clearly apparent only when $E \gtrsim 3 \times 10^4$, that is for levels higher than the 10,000th. This discovery provides a salutary counterexample to semiclassical folklore, which on the basis of experience with WKB eigenvalues for one-dimensional smooth potential wells has led to the belief that semiclassical limiting forms are good approximations even for low-lying levels. The fact that even in this very simple system unexpected level structure appears for very high-lying states provides a warning against making general conclusions in the ergodic case ($R > 0$) solely on the basis of calculations of a small number of levels (e.g., Fig. 4).

5. Level Spacing Distribution

On fine scales, the texture of the energy spectrum may be described by the probability distribution $P(S)$ for spacings S between neighbouring levels. S is measured as a fraction of the mean level spacing, which will be discussed in Section 6. In the statistical theory of nuclear energy levels [22], $P(S)$ is defined with reference to an ensemble of random matrices. In semiclassical mechanics there is no ensemble, but $P(S)$ can be defined as an average over the infinitely many levels in any fixed range $\Delta\mathscr{E}$ as $\hbar \to 0$. For Sinai's billiard this procedure is equivalent to averaging over the spacings between all levels E.

Clustering of neighbouring levels is embodied in the behaviour of $P(S)$ as $S \to 0$, which is determined by near-degeneracies of the spectrum. To derive this limiting behaviour consider first generic systems, such as Sinai's billiard for $R > 0$. Let the system under study correspond to parameters A_0, B_0 when embedded in a family of systems labelled by two parameters A, B. Each level is a sheet in E, A, B space, and as discussed in Section 4 (see Fig. 5) and Appendix D the sheets typically meet at conical points. If these sheets are labelled in order of increasing energy, and the separation between the ith and $i+1$st sheets denoted by $S_i(A, B)$, then the level spacing distribution is

$$P(S) = \overline{\delta(S - S_i(A_0, B_0))}, \tag{5.1}$$

where the bar denotes an "energy average" over all levels i.

For small S the only contributions to $P(S)$ come from sheets i with conical intersections at points A_i^*, B_i^* lying close to A_0, B_0. Near such an intersection,

$$S_i(A, B) = \{a_i(A - A_i^*)^2 + 2b_i(A - A_i^*)(B - B_i^*) + c_i(B - B_i^*)^2\}^{1/2}, \tag{5.2}$$

where a_i, b_i and c_i are constants consistent with S_i^2 being a positive-definite quadratic form (cf. Appendix D). Now, there is nothing special about the points A_0, B_0, and as i increases the deviations $A_0 - A_i^*$, $B_0 - B_i^*$, by which the system's parameters differ

from those of the degeneracies, vary randomly. Therefore the energy average can be replaced by an ensemble average over A, B, and

$$P(S) = \frac{1}{\mathscr{A}} \iint_{\mathscr{A}} dA\, dB\, \overline{\delta(S - (a_i^2 A^2 + 2b_i AB + c_i B^2)^{1/2})} \qquad \text{for small } S, \tag{5.3}$$

where the average is now over the cone parameters a_i, b_i, c_i and where $\mathscr{A}$ is an area large enough to include the elliptic contours of S_i for the S values considered. On writing $A \equiv S\alpha$, $B \equiv S\beta$, $P(S)$ becomes

$$P(S) = S\left\{\frac{1}{\mathscr{A}} \iint d\alpha\, d\beta\, \overline{\delta(1 - (a_i\alpha^2 + 2b_i\alpha\beta + c_i\beta^2)^{1/2})}\right\}$$
$$\propto S \qquad \text{as } S \to 0, \tag{5.5}$$

since the quantity in curly brackets is independent of S.

This quantitative description of level repulsion should be compared with Fig. 9, which shows a histogram of $P(S)$ compiled from 492 spacings between billiard eigenvalues with $0.20 \leqslant R \leqslant 0.44$. While there are not enough levels to make a precise test possible, it appears that linear behaviour as $S \to 0$, as predicted by Eq. (5.5), is a good fit to the data. This conclusion is supported by computations by Casati (private communication) for the spectrum of a stadium, and is also compatible with results for the stadium published by McDonald and Kaufman [32].

The argument based on conical intersections applies also to ensembles of random matrices, and indeed analytical calculations [22] lead to the linear dependence (5.5) in that case also.

Now consider integrable systems, for which degeneracies occur when a single parameter is varied. The energy level sheets in E, A, B space now intersect in curves rather than points, and the calculation of $P(S)$ is somewhat different, and leads to

$$P(S) \propto \text{constant} \qquad \text{as } S \to 0. \tag{5.6}$$

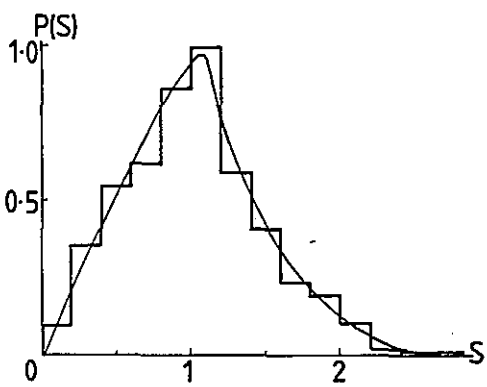

FIG. 9. Distribution of 492 level spacings S between Sinai billiard eigenvalues for $0.20 < R < 0.44$. The energies were scaled to make $\bar{S} = 1$.

This result skows that integrable systems do not exhibit level repulsion, and is in agreement with the analytical arguments and computations of Berry and Tabor [23], according to which levels are Poisson-distributed in typical integrable systems, so that

$$P(S) = e^{-S}. \tag{5.7}$$

Of course this does not apply to the exceptional case of Sinai's billiard when $R = 0$, where the degeneracies and spacings increase with E (as was shown in Section 4) so that no distribution $P(S)$ exists.

Equations (5.5) and (5.6) do not exhaust the possibilities for the limiting behaviour of $P(S)$. Suppose that a system is a member of a special class for which the production of a degeneracy requires n parameters to be varied. Then the argument based on the geometry of intersections leads to

$$P(S) \propto S^{n-1} \qquad \text{as } S \to 0, \tag{5.8}$$

a result which includes (5.5) and (5.6). In particular, if degeneracies are strictly forbidden, even an infinite number of parameters would not suffice to produce them, and $P(S)$ should vanish faster than any power of S as $S \to 0$. An example is provided by Pokrovskii [24] who shows that for electron states in a disordered one-dimensional linear chain of δ-function potentials,

$$P(S) \propto e^{-\text{const}/S^2}/S^2 \qquad \text{as } S \to 0. \tag{5.9}$$

What forbids degeneracy in this case is the boundary condition, which requires the wave function ψ to vanish at both ends of the chain: if ψ is fixed at one point, then the one-dimensional Schrödinger equation determines ψ uniquely (up to a constant factor) for given E.

Zaslavsky [25] argues that for a classically ergodic system $P(S)$ will be given not by (5.5) but instead by

$$P(S) \propto S^{\nu} \qquad \text{as } S \to 0, \tag{5.10}$$

where ν is a measure of the instability of classical trajectories. This conclusion follows from the hypothesis that the spectrum is generated by trajectories that are almost closed. I believe this to be false, and indeed shall show in Section 7 that the spectrum is generated by orbits that are exactly closed. (One way in which (5.5) might break down, however, would be for its domain of validity to shrink to zero in the classical limit $E \to \infty$, as a result of the cones becoming infinitely sharp or infinitely blunt, but I see no reason why this should happen.)

Repulsion of levels, as given by (5.5), is associated with a prevalence of near-degeneracies in the spectrum. The suggestion that many near-degeneracies in the quantum spectrum indicate stochasticity in the classical motion has been made by Marcus [26]. He also points out that individual near-degeneracies need not indicate stochasticity because they can occur in integrable systems. If the near-degeneracies

in Fig. 4 occurred only in well-defined pairs, it would be possible to regard Sinai's billiard as a perturbation of a "nearby" integrable system whose levels would actually cross; such a hope is frustrated by the prevalence of near-degeneracies of three or more levels, for which no unique underlying crossing patterns can be assigned.

6. Mean Level Density

A complete description of the spectrum consists of a list of energy levels E_i, where i labels states in order of increasing energy (degenerate states will have different, neighbouring, values of i, arbitrarily assigned). The same information is contained in the *level density* (or "density of states") $d(E)$, defined as

$$d(E) \equiv \sum_{i=1}^{\infty} \delta(E - E_i), \tag{6.1}$$

and the *mode number* (or integrated density of states) $\mathcal{N}(E)$, defined as

$$\mathcal{N}(E) \equiv \sum_{i=1}^{\infty} \Theta(E - E_i), \tag{6.2}$$

where Θ denotes the unit step function. Obviously

$$d(E) = \frac{d\mathcal{N}(E)}{dE}. \tag{6.3}$$

In this section and the next an analytical theory of the functions $d(E)$ and $\mathcal{N}(E)$ will be developed, based on the KKR determinant in its non-desymmetrized complex form (3.15). The mode number will be expressed as a semiclassical approximation in the form

$$\mathcal{N}(E) \approx \bar{\mathcal{N}}(E) + \mathcal{N}_{\text{osc}}(E), \tag{6.4}$$

where $\bar{\mathcal{N}}(E)$, the mean mode number, is a smooth function of E, and $\mathcal{N}_{\text{osc}}(E)$ is a series of oscillatory corrections.

On the coarsest energy scales the spectrum is described by $\bar{\mathcal{N}}(E)$. It is easy to get a formula for this function using the general semiclassical rule that each quantum state occupies a volume h^2 in the phase space $\mathbf{r}, \mathbf{p}$, so that, reverting to the natural energy $\mathscr{E}$,

$$\bar{\mathcal{N}}(\mathscr{E}) = \frac{1}{h^2} \iint_{\substack{\text{accessible} \\ \text{area of} \\ \text{torus}}} d\mathbf{r} \iint_{p^2/2m<\mathscr{E}} d\mathbf{p}$$

$$= \frac{(1 - \pi R^2)\, m\mathscr{E}}{2\pi\hbar^2} \tag{6.5}$$

so that

$$\bar{\mathcal{N}}(E) = \pi(1 - \pi R^2)E. \tag{6.6}$$

Results of this type, in which the mode number is proportional to the accessible area, were first obtained by Weyl [9].

Now formula (6.6) will be derived from definition (6.2) and the KKR determinant (3.15), as a necessary preparation for the more elaborate calculation of $\mathcal{N}_{\text{osc}}(E)$ that will be carried out in Section 7. The derivation begins with a transformation of (3.15) based on ideas developed by Lloyd [27] in studying the physics of electrons in disordered systems. Consider a function $F(E)$, analytic in a strip $E + i\epsilon$ of the upper half-plane, which is real when E is real and which has zeros E_i and poles E_p on the real axis. It is easy to show that

$$\lim_{\epsilon \to 0} \operatorname{Im} \ln\{F(E + i\epsilon)\} = -\pi \sum_i \Theta(E - E_i) + \pi \sum_p \Theta(E - E_p). \tag{6.7}$$

Application of this formula to (3.15) gives

$$\mathcal{N}(E) = -\frac{1}{\pi} \operatorname{Im} \ln \det \left\{ \frac{e^{-i\eta_l(E)}}{\sin \eta_l(E)} \delta_{ll'} + S_{l-l'}(E) \right\} + \sum_p \Theta(E - E_p), \tag{6.8}$$

where the limiting process involving $E + i\epsilon$ is no longer written explicitly.

Rearranging, and using the result that for any matrix M,

$$\ln \det M = \operatorname{tr} \ln M, \tag{6.9}$$

gives

$$\mathcal{N}(E) = \sum_p \Theta(E - E_p) + \frac{1}{\pi} \sum_{l=-\infty}^{\infty} \eta_l(E) + \frac{1}{\pi} \operatorname{Im} \sum_{l=-\infty}^{\infty} \ln\{\sin(\eta_l(E + i\epsilon))\}$$

$$- \frac{1}{\pi} \operatorname{Im} \operatorname{tr} \ln\{\delta_{ll'} + \sin \eta_l(E)\, e^{i\eta_l(E)} S_{l-l'}(E)\}. \tag{6.10}$$

The poles E_p of the KKR determinant are of two types: singularities of the structure constants S_l, which by (3.20) occur at the unperturbed billiard energies $E = m^2 + n^2$, and zeros of $\sin \eta_l$. The contribution of the latter poles to the first term in (6.10) are exactly cancelled by the third term in (6.10), as follows from (6.7). The result is an exact formula for $\mathcal{N}(E)$, namely

$$\mathcal{N}(E) = \sum_{-\infty < m,n < \infty} \Theta(E - m^2 - n^2) + \frac{1}{\pi} \sum_{l=-\infty}^{\infty} \eta_l(E)$$

$$- \frac{1}{\pi} \operatorname{Im} \operatorname{tr} \ln\{\delta_{ll'} + \sin \eta_l(E)\, e^{i\eta_l(E)} S_{l-l'}(E)\}. \tag{6.11}$$

This is the basis for calculating $\bar{\mathcal{N}}(E)$ and $\mathcal{N}_{\text{osc}}(E)$.

The third term of (6.11) contributes only to $\mathcal{N}_{\text{osc}}$, because the expansion of the logarithm involves products of Hankel functions (3.14) of nonzero argument. The steady part of the first term is simply the number of lattice points in a circle with radius $E^{1/2}$ in the m, n plane, so that

$$\overline{\sum \Theta(E - m^2 - n^2)} = \pi E. \tag{6.12}$$

In Appendix F the asymptotic value of the steady part of the second term isshown to be

$$\frac{1}{\pi} \overline{\sum_{l=-\infty}^{\infty} \eta_l(E)} = -\pi^2 R^2 E. \tag{6.13}$$

Combining these last two equations leads directly to the Weyl result (6.6) for $\bar{\mathcal{N}}(E)$.

When applied to the desymmetrized billiard (Fig. 3), which has ouly one-eighth of the area, the same arguments yield

$$\bar{\mathcal{N}}(E) = \frac{\pi}{8}(1 - \pi R^2) E. \tag{6.14}$$

It is most instructive to compare this result with the exact function $\mathcal{N}(E)$, as given by (6.2) together with the roots E_i of the desymmetrized KKR determinant (3.23). The comparison is shown in Fig. 10 for $R = 0.0$, 0.1, 0.2, 0.3 and 0.4. The stepped curves are the exact $\mathcal{N}(E)$ and the straight lines are graphs of (6.14). Clearly the agreement is not very good.

The origin of the discrepancy lies in the fact that (6.14) is an asymptotic formula, valid as $E \to \infty$. In order to get agreement with calculations for relatively low-lying levels, it is necessary to include correction terms in $\bar{\mathcal{N}}(E)$. Baltes and Hilf [9], reviewing the literature on this problem, quote the following result for modes in an enclosure with area $\mathcal{A}$, perimeter $\mathcal{L}$, interior corner angles θ_i and curvature $\mathcal{K}(s)$ (measured positive when convex outwards, and as a function of arc length s):

$$\bar{\mathcal{N}}(\mathcal{E}) = \frac{m\mathcal{A}\mathcal{E}}{2\pi\hbar^2} - \frac{\mathcal{L}((2m\mathcal{E})^{1/2})}{4\pi\hbar} + \sum_{\text{corners}} \frac{\pi^2 - \theta_i^2}{24\pi\theta_i} + \frac{1}{12\pi} \oint \mathcal{K}(s)\, ds. \tag{6.15}$$

This is a consistent semiclassical expansion, including all terms that do not vanish as $\hbar \to 0$. When applied to the enclosure of Fig. 3 this gives

$$\bar{\mathcal{N}}(E) = \frac{\pi}{8}(1 - \pi R^2) E - \frac{1}{2}\left\{1 + \frac{1}{2^{1/2}} - R\left(2 - \frac{\pi}{4}\right)\right\} E^{1/2} + \frac{31}{96}. \tag{6.16}$$

This formula holds when $R > 0$. When $R = 0$, 5/96 should be added. There is no inconsistency in having a discontinuity of $\bar{\mathcal{N}}(E)$ at $R = 0$, because (6.16) is strictly valid as $E \to \infty$, when modes have vanishing wavelengths that can discriminate any detail, however small, of the shape of the enclosure. In any case, 5/96 is undetectable at the level of accuracy with which $\bar{\mathcal{N}}(E)$ can be inferred from my computations.

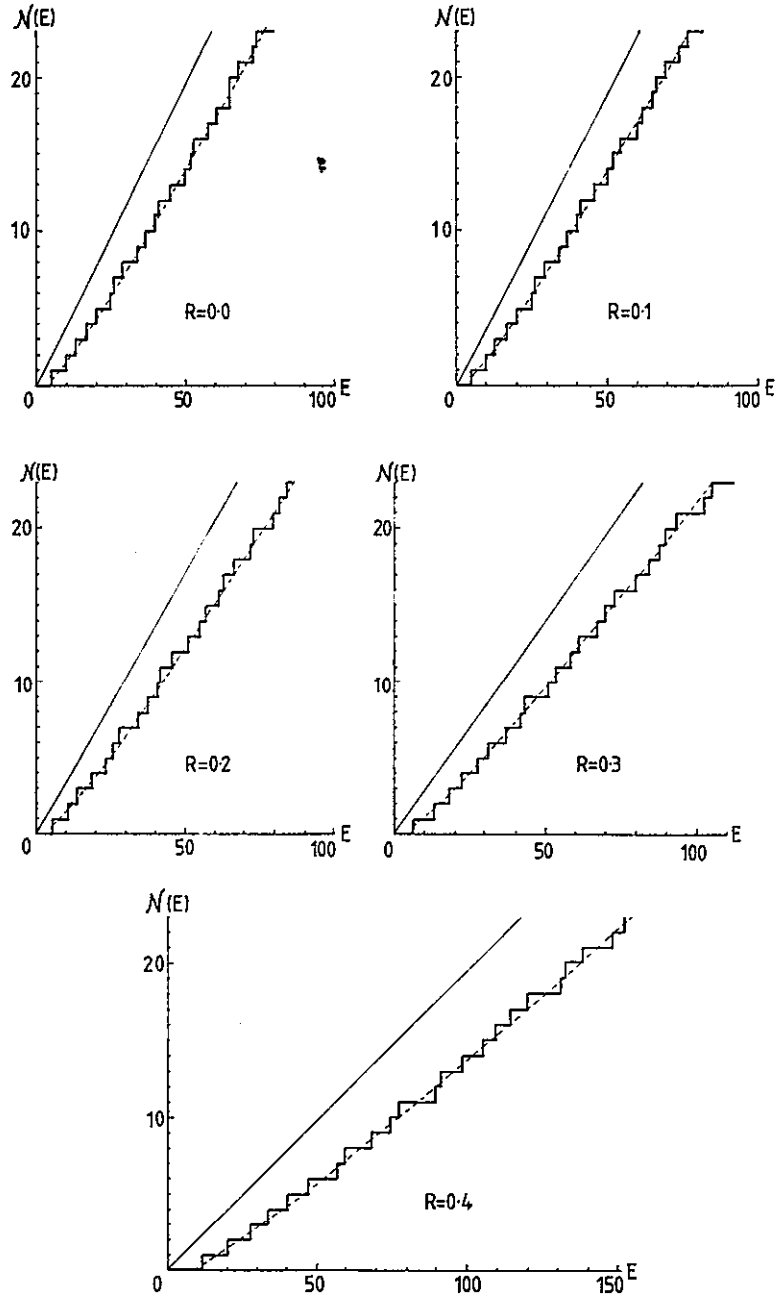

FIG. 10. Mode number $\mathcal{N}(E)$ for the indicated values of R. The stepped curves give the exact $\mathcal{N}(E)$, the straight lines give the Weyl "area" approximation and the dashed lines include the perimeter, corner and curvature corrections.

Graphs of (6.16) are shown as the dashed curves on Fig. 10. The improvement is dramatic, and shows the importance of the perimeter, corner and curvature corrections to the leading term in $\bar{\mathcal{N}}(E)$.

7. Derivation of Oscillatory Contributions to the Level Density, in Terms of Closed Classical Orbits

In the semiclassical limit, deviations from the mean mode number $\bar{\mathcal{N}}(E)$ (Eq. (6.6)) take the form of oscillatory corrections $\mathcal{N}_{\rm osc}(E)$, describing the spectrum on intermediate scales of energy. The derivation of $\mathcal{N}_{\rm osc}$ from (6.11) requires an unorthodox version of multiple scattering theory which gives neither a perturbation series nor an expansion in powers of the density of discs. Readers interested in the resulting formulae rather than their derivation should proceed directly to Section 8.

Expansion of the logarithm in (6.11) gives

$$\mathcal{N}_{\rm osc} = \mathcal{N}^{(1)}_{\rm osc} + \mathcal{N}^{(2)}_{\rm osc} + \mathcal{N}^{(3)}_{\rm osc}, \tag{7.1}$$

where

$$\mathcal{N}^{(1)}_{\rm osc} = \sum_{\rho} \Theta(E - \rho^2) - \pi E, \tag{7.2}$$

$$\mathcal{N}^{(2)}_{\rm osc} = \frac{1}{\pi} \sum_{l=-\infty}^{\infty} \eta_l(E) + \pi^2 R^2 E, \tag{7.3}$$

$$\mathcal{N}^{(3)}_{\rm osc} = \frac{\rm Im}{\pi} \sum_{n=1}^{\infty} \frac{1}{2^n n} \sum_{l_i} \cdots \sum_{l_n} \sum_{\rho_i}{}' \cdots \sum_{\rho_n}{}' (e^{2i\eta_1} - 1) \cdots (e^{2i\eta_n} - 1)$$

$$\times H^{(1)}_{l_1 - l_2}(k\rho_1) \cdots H^{(1)}_{l_n - l_1}(k\rho_n) \exp[-i\{(l_1 - l_2)\,\phi_1 + \cdots + (l_n - l_1)\,\phi_n\}], \tag{7.4}$$

and where η_i denotes η_{l_i} and ϕ_i denotes ϕ_{ρ_i}.

Consider first $\mathcal{N}^{(1)}_{\rm osc}$. The two-dimensional Poisson formula, namely

$$\sum_{\rho} F(\rho) = \sum_{\rho} \iint d\rho'\, F(\rho')\, e^{2\pi i \rho \cdot \rho'} \tag{7.5}$$

for any function F, transforms the lattice sum, to give

$$\mathcal{N}^{(1)}_{\rm osc} = E^{1/2} \sum_{\rho} \frac{J_1(2\pi\rho(E)^{1/2})}{\rho} - \pi E$$

$$= \frac{k}{2\pi}\, {\rm Im}\, i \sum_{\rho} \frac{H^{(1)}_1(k\rho)}{\rho}. \tag{7.6}$$

Semiclassically, $k\rho$ is large for all lattice vectors ρ, so that $H_1^{(1)}$ can be approximated by its asymptotic approximation. Thus

$$\mathcal{N}_{\rm osc}^{(1)} = \frac{1}{\pi}\left(\frac{k}{2\pi}\right)^{1/2} \sum_{\rho}{}' \frac{\exp[i(k\rho - \pi/4)]}{\rho^{3/2}}. \tag{7.7}$$

This contribution to $\mathcal{N}_{\rm osc}$ corresponds to the unperturbed system ($R = 0$). Each lattice vector ρ represents a closed orbit on the torus, which contributes to $\mathcal{N}_{\rm osc}^{(1)}$ in precisely the manner described by Berry and Tabor [11] for a general integrable system. When $R > 0$, however, not all ρ correspond to possible closed orbits; as explained in Section 2, the only surviving unperturbed orbits are those which do not strike a disc (e.g., α-type orbits on Fig. 1c). It will be shown that what happens is that the impossible orbits in (7.7) are cancelled by contributions from $\mathcal{N}_{\rm osc}^{(3)}$.

Next, consider $\mathcal{N}_{\rm osc}^{(2)}$. The one-dimension Poisson formula, namely,

$$\sum_{l=-\infty}^{\infty} f(l) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dl\, f(l)\, e^{2\pi i m l} \tag{7.8}$$

for any function f, can be used to transform the summation in (7.3). The term $m = 0$ gives the nonoscillatory contribution (6.13), leaving

$$\mathcal{N}_{\rm osc}^{(2)} = \frac{4}{\pi} \sum_{m=1}^{\infty} \int_0^{\infty} dl\, \eta_l \cos 2\pi m l. \tag{7.9}$$

In Appendix F this is shown to be negligible in the semiclassical limit, so that

$$\mathcal{N}_{\rm osc}^{(2)} \approx 0. \tag{7.10}$$

Now consider $\mathcal{N}_{\rm osc}^{(3)}$. As defined by (7.4), this is a quantum-mechanical multiple scattering series. Each term is a scattering path with n steps, starting with the disc at $\rho = 0$ with angular momentum l_1, propagating to hit the scatterer at ρ_1 with angular momentum l_2, etc., the last leg being from the scatterer at $\rho_1 + \cdots + \rho_{n+1}$ with l_n to the scatterer at $\rho_1 + \cdots + \rho_n$ with l_1. Semiclassically, the l-summations will be performed via the Poisson formula (7.8), with l replaced by the impact parameter b, defined by

$$l \equiv kb. \tag{7.11}$$

The Hankel functions can be replaced by their Debye large-argument asymptotic forms, namely

$$H_{kb}^{(1)}(k\rho) \approx \left(\frac{2}{\pi k}\right)^{1/2} \frac{\exp[ik\{(\rho^2 - b^2)^{1/2} - b \arccos(b/\rho)\} - i\pi/4]}{(\rho^2 - b^2)^{1/4}}. \tag{7.12}$$

The phase shifts will be approximated by their WKB asymptotic forms [28] (Appendix F), and written as

$$\eta_l \approx k\tilde{\eta}(b) - \pi/4. \tag{7.13}$$

If $|b| > R$, $\tilde{\eta} = 0$, reflecting the fact that classical particles with such impact parameters do not hit a disc. Summations over l will therefore be replaced by

$$\sum_{l=-\infty}^{\infty} \to \sum_{m=-\infty}^{\infty} k \int_{-R}^{R} db\, e^{2\pi i k b m}. \tag{7.14}$$

Now follow two crucial steps. The first is to multiply out the product of factors $(\exp i\eta_l - 1)$, to give a series of terms whose factors are $\exp i\eta_l$'s and -1's; all these 2^n terms will be considered as separate paths. The second step is to evaluate the integrals over b asymptotically for large k, by employing the method of stationary phase. As a result of these steps, only classical paths will be shown to contribute to $\mathcal{N}_{\text{osc}}^{(3)}$.

Consider first the terms involving only phase shift factors (i.e., no -1's). Define

$$P(b_1, b_2; \rho_1) \equiv \tfrac{1}{2} \exp[2ik\tilde{\eta}(b_1)]\, H^{(1)}_{k(b_1-b_2)}(k\rho_1) \exp[-ik\phi_1(b_1 - b_2)]. \tag{7.15}$$

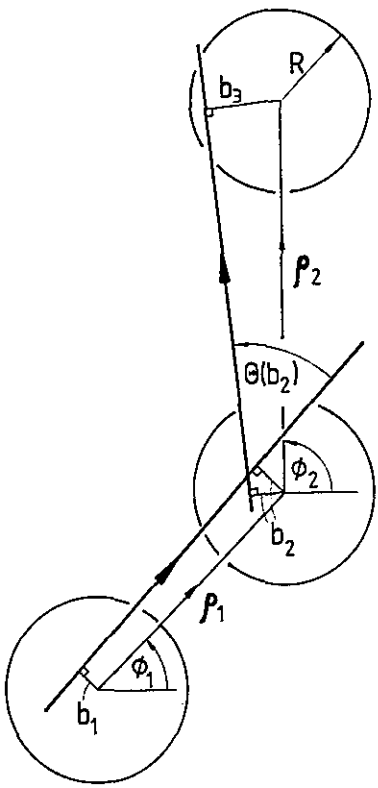

FIG. 11. Notation for typical intermediate scattering in one term of the multiple scattering expansion (7.4). The process illustrated is not a classical scattering because the trajectory is not specularly reflected from the disc.

A typical intermediate scattering in (7.4) then depends on

$$T \equiv \sum_{l_2} P(b_1, b_2; \rho_1)\, P(b_2, b_3; \rho_2). \tag{7.16}$$

Figure 11 illustrates this term. With (7.12)–(7.14), it can be written

$$T = \frac{e^{-i\pi/2}}{2\pi} \sum_{m_2=-\infty}^{\infty} \int_{-R}^{R} db_2 \frac{e^{ikx}}{\{\rho_1^2 - (b_1 - b_2)^2\}^{1/4}\{\rho_2^2 - (b_2 - b_3)^2\}^{1/4}}, \tag{7.17}$$

where

$$\begin{aligned}\chi = {} & 2\pi m_2 b_2 + 2\tilde{\eta}(b_1) + 2\tilde{\eta}(b_2) + \{\rho_1^2 - (b_1 - b_2)^2\}^{1/2} + \{\rho_2^2 - (b_1 - b_3)^2\}^{1/2} \\ & - (b_1 - b_2) \arccos\left(\frac{b_1 - b_2}{\rho_1}\right) - (b_2 - b_3) \arccos\left(\frac{b_2 - b_3}{\rho_2}\right) \\ & - (b_1 - b_2)\,\phi_1 - (b_2 - b_3)\,\phi_2 .\end{aligned} \tag{7.18}$$

Stationary phase occurs when

$$\frac{\partial \chi}{\partial b_2} = 0. \tag{7.19}$$

To see what this means, first realise that

$$2\frac{d\tilde{\eta}(b)}{db} = \Theta(b) = 2 \arccos\frac{b}{R}, \tag{7.20}$$

where $\Theta(b)$ is the classical deflection function [28], reckoned anticlockwise from the direction of incidence. Now (7.19) becomes

$$\Theta(b_2) + 2\pi m_2 = \phi_2 + \arcsin\left(\frac{b_3 - b_2}{\rho_2}\right) - \phi_1 - \arcsin\left(\frac{b_2 - b_1}{\rho_1}\right). \tag{7.21}$$

As can be seen from Fig. 11, this is precisely the condition that *the intermediate scattering is classical*, i.e., that the intermediate disc reflects specularly. In similar fashion, the scatterings at $b_3 \cdots b_n$ can be shown to be classical. This leaves only the summation over l_1 in (7.4), which involves

$$T = \sum_{l_1} P(b_1, b_2; \rho_1)\, P(b_n, b_1; \rho_n). \tag{7.22}$$

Just before the final scattering, by the disc at $\rho_1 + \cdots + \rho_n$, the particle has the same impact parameter b_1 with which it set out from the disc at the origin. By applying to (7.22) the same stationary phase argument that led to (7.21), it follows easily that just after the final scattering the particle is moving in its initial direction. Therefore the paths contributing to $\mathcal{N}_{\text{osc}}$ are not merely classical but *closed on the torus*. These are

the unstable closed orbits (all isolated) described in Section 2 (e.g., β-type orbits on Fig. 1c).

For the hard-disc scatterers considered here, the stationary phase conditions select at most one value of the integer m introduced at each scattering by the Poisson formula (7.8). The integers will depend on the total angle turned through along the whole orbit.

In finding the contribution $\mathcal{N}$ from a complete closed orbit with n steps, the simplest procedure is to evaluate all the integrals over $b_1 \cdots b_n$ together by the n-dimensional method of stationary phase. From (7.4), (7.12) and (7.13), and omitting the Poisson integers m,

$$\mathcal{N} = \frac{\mathrm{Im}}{\pi n}\left(\frac{k}{2\pi}\right)^{n/2} e^{-3\pi i n/4}\int_{-R}^{R} db_1 \cdots \int_{-R}^{R} db_n \prod_{\nu=1}^{n}\{\rho_\nu^2 - (b_\nu - b_{\nu+1})^2\}^{-1/4}$$

$$\times \exp\left[ik\sum_{\nu=1}^{n}\left\{2\tilde{\eta}_\nu + (\rho_\nu^2 - (b_\nu - b_{\nu+1})^2)^{1/2}\right.\right. \tag{7.23}$$

$$\left.\left. - (b_\nu - b_{\nu+1})\left(\arccos\left(\frac{b_\nu - b_{\nu+1}}{\rho_\nu}\right) + \phi_\nu\right)\right\}\right],$$

where $b_{n+1} \equiv b_1$. It is shown in Appendix G that the stationary value of the phase in the exponent is simply kL, where L is the length of the orbit. The amplitude of the stationary-phase contribution depends on the $n \times n$ determinant of second derivatives of the exponent, which will be written in the form

$$e^{i\pi n}k^n\left[\prod_{\nu=1}^{n} d_\nu^{-1}\right] D, \tag{7.24}$$

where

$$d_\nu \equiv \{\rho_\nu^2 - (b_\nu - b_{\nu-1})^2\}^{1/2} \tag{7.25}$$

and

$$D \equiv \det\{A_\mu\,\delta_{\mu\nu} + B_{\nu-1}\,\delta_{\mu,\nu-1} + B_{\nu-1}^{-1}\,\delta_{\mu,\nu+1}\}, \tag{7.26}$$

where

$$A_\mu \equiv 2\left(\frac{d_\nu\, d_{\nu-1}}{R - b_\nu^2}\right)^{1/2} - \left(\frac{d_{\nu-1}}{d_\nu}\right)^{1/2} - \left(\frac{d_\nu}{d_{\nu-1}}\right)^{1/2}, \qquad B_\nu \equiv \left(\frac{d_\nu}{d_{\nu-1}}\right)^{1/2}. \tag{7.27}$$

All repetitions of a given closed orbit will give stationary phase contributions to $\mathcal{N}_{\mathrm{osc}}$. Let the path under consideration consist of p repetitions of an s-step basic path (i.e., a path which cannot be decomposed into repeated sections) labelled by β (cf. Fig. 1c). Thus $n = ps$. The matrix elements in (7.26) now have period s, and by a slight extension of a method employed by Balian and Bloch [10] it can be shown that

the periodic determinant D can be expressed in terms of two $s \times s$ determinants corresponding to a single traversal of β, as follows:

$$D = \{\Delta_\beta^p - (-1)^{sp}\,\Delta_\beta^{-p}\}^2,$$

where

$$\Delta_\beta \equiv \tfrac{1}{2}(D_+^{1/2} + D_-^{1/2}) \tag{7.29}$$

and

$$D_\pm \equiv \det \begin{pmatrix} A_1 & B_s & 0 & 0 & \cdot & \cdot & \pm B_s \\ B_1 & A_2 & B_1^{-1} & 0 & \cdot & \cdot & 0 \\ 0 & B_2 & A_3 & B_2^{-1} & \cdot & \cdot & \\ 0 & 0 & 0 & 0 & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & B_{s-2} & A_{s-1} & B_{s-2}^{-1} \\ \pm B_s^{-1} & 0 & \cdot & \cdot & 0 & B_{s-1} & A_s \end{pmatrix}. \tag{7.30}$$

The length of this path is pL_β, where L_β is the length of the basic path. Each of the s discs may be taken as the initial one, and the path can be traversed in both directions; therefore its contribution must be multiplied by $2s$ (for paths where two reflections occur at normal incidence on a disc and which therefore retrace themselves, the factor 2 is absent). In addition, all eigenvalues of the matrix in (7.26) are positive. The stationary phase approximation to (7.23) can now be written down:

$$\mathcal{N} = \frac{2(-1)^{sp}\sin(kpL_\beta)}{\pi p(\Delta_\beta^p - (-1)^{sp}\,\Delta_\beta^{-p})}. \tag{7.31}$$

For the simplest case of one traversal of a 'bounce' orbit between two discs separated by lattice vector $\boldsymbol{\rho}$, this formula can be written explicitly as

$$\mathcal{N} = \frac{\sin[2k(\rho - 2R)]}{\pi p\{[(\rho/2R)^{1/2} + (\rho/2R - 1)^{1/2}]^2 - [(\rho/2R)^{1/2} + (\rho/2R - 1)^{1/2}]^{-2}\}}. \tag{7.32}$$

The terms (7.31) must be summed for all repetitions p and all basic paths β. It is very important to realize now that the *great majority of these paths are impossible* because, although their reflections at discs are specular and they propagate between discs in straight lines, they would have to pass through intermediate discs. I shall call these intermediate discs "ghosts"; Fig. 12a shows a 4-step path with two ghosts. It is obvious that these paths cannot appear in the final semiclassical approximation to $\mathcal{N}_{\text{osc}}$. Balian and Bloch [10] prove a general "cancellation theorem" concerning impossible scatterings involving boundaries of any shape, but it is instructive to see explicitly in the case of Sinai's billiard exactly how the cancellation operates in the multiple scattering expansion. What happens is that the contributions from impossible paths are cancelled by the "-1" terms in (7.4), which until now have been ignored. The subtle manner in which this cancellation occurs will now be explained.

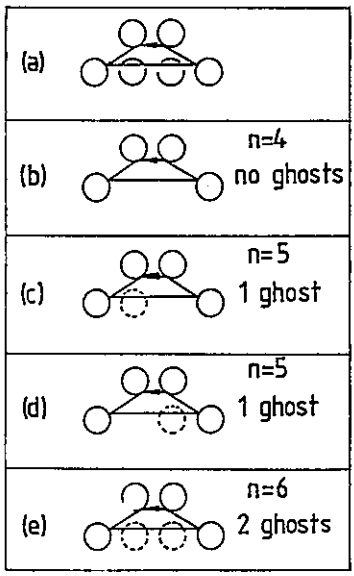

FIG. 12. (a) Four-step isolated impossible orbit with two "ghost" scatterings; (b–d) different paths by which the orbit in (a) gives contributions to $\mathcal{N}^{(3)}_{\text{osc}}$.

The general term in (7.4), once the factors $(\exp[i\eta_l] - 1)$ have been multiplied out, contains a string of factors $\exp[i\eta_l]$ and -1's. Each factor -1 corresponds to a "ghost scattering" from a disc with phase shift zero; this scattering, and the subsequent propagation, is described by a function $P_0(b_1, b_2; \rho_1)$, given by (7.15) with $\tilde{\eta}(b_1) = 0$. By analogy with (7.16), a typical intermediate ghost scattering (illustrated by Fig. 11 with a ghost as the middle disc) depends on

$$T_0 \equiv \sum_{l_2} P(b_1, b_2; \rho_1)\, P_0(b_2, b_3; \rho_2). \tag{7.33}$$

When applied to this function, the same asymptotics used on (7.16) lead to the simple result

$$T_0 \approx P(b_1, b_3; \rho_1 + \rho_2). \tag{7.34}$$

Therefore the only effect of each ghost scattering is to contribute a factor -1 to the path in which it occurs.

Each impossible path (e.g., Fig. 12a), can be traversed by several combinations of ghost scatterings (e.g., Figs. 12b–e), with different total numbers of steps n. Consider first an impossible path traversed once; each traversal can begin at any point, and this cancels the factors n in (7.4). Let the impossible path contain m ghosts. It may be traversed with j ghost scatterings, where $0 \leqslant j \leqslant m$, in $m!/(m-j)!\,j!$ ways, so that the total contribution to $\mathcal{N}_{\text{osc}}$ from an impossible path is

$$\sum_{j=0}^{m} \frac{(-1)^j\, m!}{(m-j)!\, j!} = \delta_{m0}. \tag{7.35}$$

Thus the contribution is zero unless there are no ghosts. For multiple traversals of an impossible path, cancellation can be established by applying (7.35) separately to paths where the same ghost scatterings occur on each traversal, and paths where different ghost scatterings occur on each traversal.

The final cancellation involves the term in (7.4) consisting entirely of -1 factors (i.e., no phase shift factors). These will cancel the impossible paths consisting entirely of ghost scatterings, which gave rise to unphysical contributions to $\mathcal{N}_{\rm osc}^{(1)}$ in Eq. (7.7). It is no longer possible to evaluate all b-integrations (in the analogue of (7.23) with $\tilde{\eta} = 0$) by stationary phase, because the paths now being considered are not isolated and the determinant D in (7.26) would vanish. But stationary phase can be applied to all but one integration, and in view of the ghost scattering law (7.34) this leads to a particularly simple result. For the closed path consisting of the primitive lattice vector $\rho_0 = (m, n)$ (where m and n are mutually prime) the contribution $\mathcal{N}$ can be written as

$$\mathcal{N} = \frac{\mathrm{Im}}{\pi}\left(\frac{k}{2\pi}\right)^{1/2}\frac{\exp[i(k\rho_0 - \pi/4)]}{\rho_0^{1/2}}\int_{-R}^{R} db_1\,\sigma(b_1\,,\,\rho_0). \tag{7.36}$$

The factor σ is defined with respect to Fig. 13. Let $n(b_1\,,\,\rho_0)$ be the total number of discs intersected by ρ_0 between the initial and final scatterings with impact parameter b_1 (in Fig. 13a, $n = 4$). The path ρ_0 can be realised in many ways (Figs. 13b–i); a typical such realisation has j intermediate ghost scatterings, and $j + 1$ steps where $0 \leqslant j \leqslant n - 1$. Summing over all possibilities gives σ as

$$\begin{aligned}\sigma(b_1\,,\,\rho_0) &= \sum_{j=0}^{n-1}\frac{(-1)^{j+1}(n-1)!}{(j+1)(n-1-j)!\,j!}\\ &= -\frac{1}{n(b_1\,,\,\rho_0)}\,.\end{aligned} \tag{7.37}$$

(The factor $1/(j + 1)$ has been retained in the summand because for these free paths, where each disc is intersected with a different b, it is not legitimate to consider each of the $j + 1$ discs as an alternative starting point for the path, as was done for the bouncing paths previously considered.)

If there are no intermediate ghosts, then $n = 1$ for all b_1. This occurs when there is a clear view between discs at 0 and ρ_0 and requires condition (2.1), i.e., $2R\rho_0 < 1$, to be satisfied. Equation (7.36) gives $\sigma = -1$ and the contribution (7.36) to $\mathcal{N}_{\rm osc}^{(3)}$ is

$$\mathcal{N} = -\frac{\mathrm{Im}}{\pi}\left(\frac{k}{2\pi}\right)^{1/2}\frac{\exp[i(k\rho_0 - \pi/4)]}{\rho_0^{3/2}}\,2R\rho_0\,. \tag{7.38}$$

When combined with the corresponding contribution to $\mathcal{N}_{\rm osc}^{(1)}$ (Eq. (7.7)), this yields the factor $1 - 2R\rho_0$, which according to (2.2) gives precisely the measure of the allowed paths of type α (non-disc-hitting). If there are intermediate ghosts for some

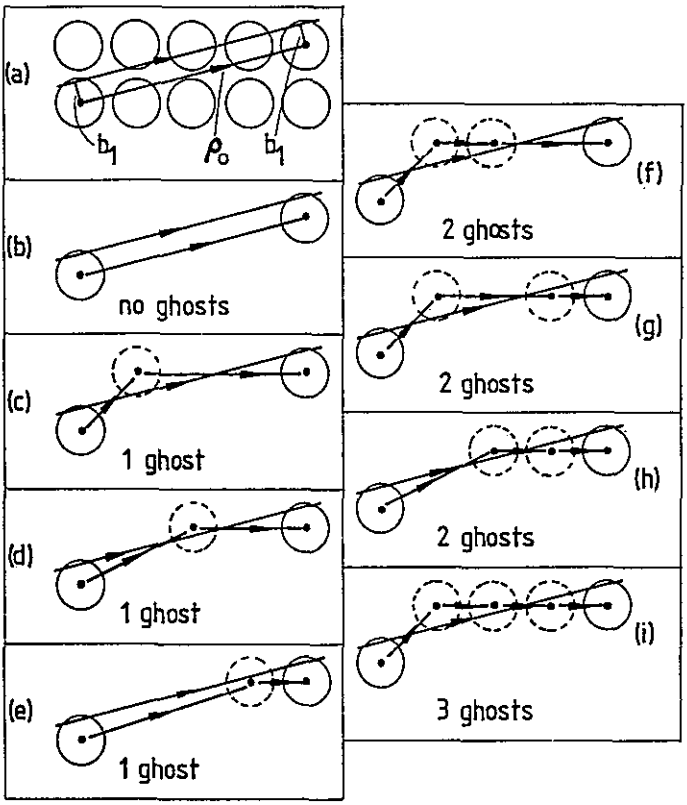

FIG. 13. (a) Impossible nonisolated path with impact parameter b_1 and lattice vector $\rho_0 = (4, 1)$; (b) realisation of this path with no intermediate ghosts; (c–e) realisation with one intermediate ghost; (f–h) realization with two intermediate ghosts; (i) realization with three intermediate ghosts.

or all values of b_1, i.e., if $2R\rho_0 > 1$, then detailed calculations for particular cases, as well as a statistical argument, strongly suggest that

$$\int_{-R}^{R} db_1\, \sigma(b_1, \rho_0) = -2\int_0^R \frac{db_1}{n(b_1, \rho_0)} = -\frac{1}{\rho_0} \tag{7.39}$$

but I have not found a general proof. This gives a contribution (7.36) to $\mathcal{N}_{\text{osc}}^{(3)}$ which exactly cancels the corresponding contribution (7.7) to $\mathcal{N}_{\text{osc}}^{(1)}$, as it should because there are no unscattered paths with $2R\rho_0 > 1$.

For completely ghostly paths consisting of p traversals of a primitive path ρ_0, corresponding to the lattice vector $\rho = p\rho_0$, a simple extension of the combinatorial arguments just presented shows that the contribution $\mathcal{N}$ to $\mathcal{N}_{\text{osc}}^{(3)}$ is given by (7.36) divided by $p^{3/2}$, with ρ_0 in the phase replaced by $p\rho_0$. Therefore the cancellation of impossible unscattered orbits from $\mathcal{N}_{\text{osc}}^{(1)}$ applies also for nonprimitive orbits.

The result of the lengthy arguments of this section is that $\mathcal{N}_{\rm osc}$ is a sum of contributions from possible isolated and nonisolated classical paths, i.e., those paths not intersecting intermediate discs. In the next section the resulting formulae will be summarised and discussed.

8. Discussion of Classical Path Sum

The outcome of the asymptotic analysis in the last two sections of the KKR determinant (3.15) is a series of semiclassical expressions for $\bar{\mathcal{N}}$ and $\mathcal{N}_{\rm osc}$. The division of $\mathcal{N}_{\rm osc}$ into $\mathcal{N}^{(1)}_{\rm osc}$, $\mathcal{N}^{(2)}_{\rm osc}$ and $\mathcal{N}^{(3)}_{\rm osc}$ (Eqs. (7.2)–(7.4)), whilst apparently mathematically natural, turned out in fact to be rather artificial, in that $\mathcal{N}^{(2)}_{\rm osc}$ was negligible and some contributions to $\mathcal{N}^{(3)}_{\rm osc}$ cancelled either one another or contributions to $\mathcal{N}^{(1)}_{\rm osc}$. It is much more natural instead to write the mode number as

$$\mathcal{N}(E) \approx \bar{\mathcal{N}}(E) + \mathcal{N}^{(\alpha)}_{\rm osc}(E) + \mathcal{N}^{(\beta)}_{\rm osc}(E), \tag{8.1}$$

where $\mathcal{N}^{(\alpha)}_{\rm osc}$ and $\mathcal{N}^{(\beta)}_{\rm osc}$ give, respectively, the contributions from the nonisolated and isolated closed orbits discussed in Section 2 (cf. Fig. (1c). A similar notation will be employed for the level density $d(E)$ defined by Eqs. (6.1) and (6.3). Explicit formulae will now be presented for the three contributions to $d(E)$.

The mean level density $\bar{d}$ is given by (6.6) as

$$\bar{d}(E) = \pi(1 - \pi R^2). \tag{8.2}$$

The contribution $d^{(\alpha)}_{\rm osc}$ from the nonisolated paths is given by (7.7), (7.35), (7.37) and (7.38) as a sum over primitive paths ρ_0 and traversals p as

$$d^{(\alpha)}_{\rm osc}(E) = E^{-1/4} \sum_{\rho_0}{}' \frac{(1 - 2R\rho_0)\,\Theta(1 - 2R\rho_0)}{\rho_0^{1/2}} \sum_{p=1}^{\infty} \frac{\cos(kp\rho_0 - \pi/4)}{p^{1/2}}, \tag{8.3}$$

where as before Θ denotes the unit step function.

The contribution $d^{(\beta)}_{\rm osc}$ from the isolated paths is given by (7.31), (7.29) and (7.30) as a sum over primitive paths β (of s_β steps) and traversals p as

$$d^{(\beta)}_{\rm osc}(E) = 2E^{-1/2} \sum_{\beta} L_\beta \sum_{p=1}^{\infty} \frac{(-1)^{ps_\beta} \cos(kpL_\beta)}{\Delta_\beta{}^p - (-1)^{ps_\beta} \Delta_\beta{}^{-p}}. \tag{8.4}$$

In understanding these contributions to $d(E)$ and $\mathcal{N}(E)$ it is essential to appreciate their different orders of magnitude in the variables $\hbar$, E or k, namely

$$\left.\begin{aligned} \bar{\mathcal{N}} &= \mathcal{O}(\hbar^{-2}) = \mathcal{O}(E^1) = \mathcal{O}(k^2) \\ \mathcal{N}^{(\alpha)}_{\rm osc} &= \mathcal{O}(\hbar^{-1/2}) = \mathcal{O}(E^{1/4}) = \mathcal{O}(k^{1/2}) \\ \mathcal{N}^{(\beta)}_{\rm osc} &= \mathcal{O}(\hbar^{0}) = \mathcal{O}(E^{0}) = \mathcal{O}(k^{0}) \end{aligned}\right\} \tag{8.5}$$

and

$$\left.\begin{aligned} \bar{d} &= \mathcal{O}(\hbar^2) = \mathcal{O}(E^0) = \mathcal{O}(k^0) \\ d_{\mathrm{osc}}^{(\alpha)} &= \mathcal{O}(\hbar^{-3/2}) = \mathcal{O}(E^{-1/4}) = \mathcal{O}(k^{-1/2}) \\ d_{\mathrm{osc}}^{(\beta)} &\quad \mathcal{O}(\hbar^{-1}) = \mathcal{O}(E^{-1/2}) = \mathcal{O}(k^{-1}). \end{aligned}\right\} \tag{8.6}$$

All the terms have previously been derived by general semiclassical methods. The steady terms, and higher-order corrections to them, have a long history, well reviewed by Baltes and Hilf [9]. Oscillatory corrections for isolated unstable orbits (β-type) were first derived, and their importance appreciated by Gutzwiller [12] in 1971. Shortly afterwards, Balian and Bloch [10] also derived the β-type corrections, in the context of a general analysis in which the different asymptotic form of oscillatory corrections for nonisolated orbits (α-type) was also obtained. Subsequently, the α-type corrections for general integrable systems (where almost all closed orbits are nonisolated) were derived by Berry and Tabor [11]. All these authors employed general semiclassical arguments, in contrast to the present work where the formulae follow from the exact solution of a model selected because it is known to give ergodic classical motion.

On the basis of definition (6.1) of $d(E)$, it is expected that if sufficient closed orbits are included in the path sums (8.3) and (8.4) their oscillatory contributions would combine to give, at least approximately, a series of δ functions at the Sinai billiard energy levels, together with a smooth background, cancelling $\bar{d}(E)$. Certainly this occurs in numerical calculations for general integrable systems [11] (and in particular for the sphere [10]) where almost all closed orbits are nonisolated. In evaluating the path sums, it is tempting to sum first over all traversals p. For the nonisolated orbits, the resulting contributions to $d(E)$ have a series of square root divergences [10], representing not individual levels but clusters of levels. When R is small, these clusters are obscured in the subsequent summation over the large number of primitive paths p_0 permitted by (2.1), but for large R the levels should show prominent clustering on large scales (while of course repelling on small scales as explained in Section 5). For the isolated orbits, the sum over p converges rapidly if Δ_β is large, i.e., if the primitive orbit is very unstable. If β is not so unstable, so that Δ_β just exceeds unity, the sum over p gives a series of Lorentz resonances to $d(E)$ [12], but the number of isolated orbits β is so large that I expect this structure to be obscured in the subsequent summation over β.

On the basis of the asymptotic estimates (8.6), it might be supposed that in the path sum only the dominant contributions, i.e., $d_{\mathrm{osc}}^{(\alpha)}$, need be retained. But this would be a mistake, because the isolated orbits, although their individual contributions are smaller, are vastly more numerous than the nonisolated orbits, in a sense now to be explained.

From (8.3) and (8.4), a closed orbit of length L, whatever its type, contributes to $d(E)$ an oscillation with phase $kL = 2\pi L(E)^{1/2}$. This has energy "wavelength"

$$\Delta E = \frac{2(E)^{1/2}}{L}. \tag{8.7}$$

so that the longer paths give faster oscillations. In order to discriminate individual levels it is necessary to include in d_{osc} all those paths for which ΔE exceeds the mean level spacing $\bar{d}^{-1}$; that is, all paths for which $L < L_{\max}$ must be included, where

$$L_{\max} = 2(E)^{1/2}\,\bar{d} = 2\pi(E)^{1/2}(1 - \pi R^2). \tag{8.8}$$

In terms of $\hbar$, $L_{\max}$ is of order $\hbar^{-1}$, so that the path sum coverges very slowly in the semiclassical limit. This behaviour contrasts with that of one-dimensional systems, where $L_{\max}$ is of order $\hbar^0$ and the δ functions emerge clearly when only a few paths are included, as was first shown by Norcliffe and Percival [31].

In view of this slow convergence for two-dimensional systems, the question naturally arises: is the classical path sum a better method than the KKR determinant (3.16) for determining energy levels in the semiclassical limit? To answer this, it is necessary to estimate the number of operations that each method requires.

Consider first the classical path sum. All closed orbits for which $L < L_{\max}$ must be included. Let the number of such orbits be denoted by ν. For integrable systems, such as Sinai's billiard with $R = 0$, each path is represented by a lattice vector $\mathbf{p}$, so that

$$\nu = \pi L_{\max}^2 . \tag{8.9}$$

Of course in an integrable system the semiclassical eigenvalues are given by quantization of the tori and so the path sum would be a very inefficient way of computing them.

In Sinai's billiard with $R > 0$, the nonisolated orbits consist of all repetitions $p\mathbf{p}_0$ of a finite number of primitive lattice vectors satisfying the "clear view" condition (2.1). The isolated orbits are much more numerous; indeed, arguments given in Appendix H show that

$$\nu = Ae^{BL_{\max}}, \tag{8.10}$$

where A and B are constants. The rapid proliferation of isolated orbits is illustrated in Figs. 14(c–l) for the full and desymmetrized billiards; for comparison, the two shortest nonisolated orbits are shown on Figs. 14a and b. Calculation of the contributions (8.4) to $d_{\mathrm{osc}}^{(g)}$ requires the evaluation of the $s \times s$ determinants $D_{\pm}$ defined by (7.30), where s is the number of steps in one circuit of the path. For large $L_{\max}$, most of the paths are traversed only once, and as shown in Appendix H the average number of steps $\bar{s}$ is $CL_{\max}$, where C is a constant, indicating that most bounces take place between neighbouring or near-neighbouring discs. Evaluation of a determinant of large order n by Gaussian elimination requires $\mathcal{O}(n^3/3)$ operations. The total number ν_{tot} of operations necessary to compute the path sum to an accuracy sufficient to discriminate the levels is therefore

$$\begin{aligned}\nu_{\mathrm{tot}} &= \bar{s}^3\nu/3 \\ &\approx (CL_{\max})^3\, Ae^{BL_{\max}}/3 \\ &\sim E^{3/2}\exp[2\pi B(1 - \pi R^2)(E)^{1/2}],\end{aligned} \tag{8.11}$$

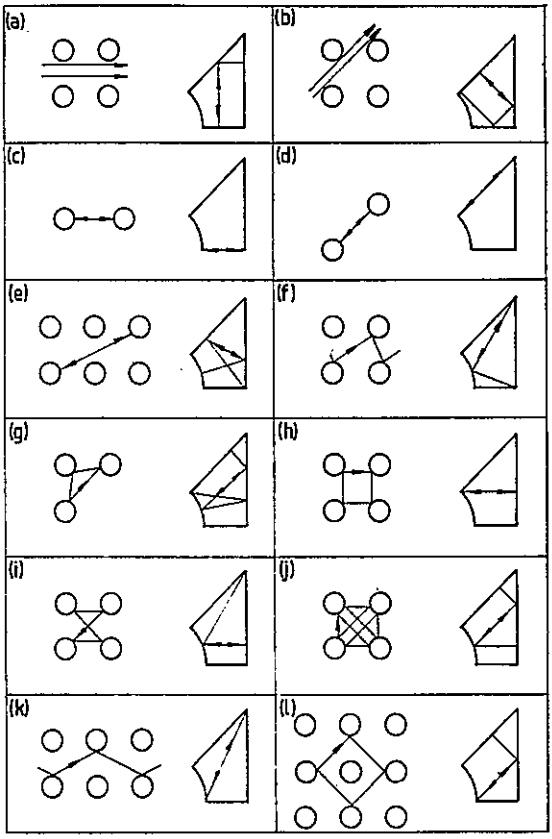

FIG. 14. Some of the shortest closed orbits in the unrestricted Sinai billiard (left-hand pictures) and the desymmetrized billiard (right-hand pictures): (a) and (b), nonisolated orbits; (c–l), isolated orbits. In the desymmetrized case the orbits (c) and (d) lie entirely along nodes of the wave function and do not contribute to the classical path sum.

where (8.8) has been used. Most of the difficulty in computing the path sum therefore lies in enumerating the isolated closed orbits.

Now consider the KKR determinant (3.16). Its elements are negligible if $|l| < 2\pi R(E)^{1/2}$, and so its effective order is $4\pi R(E)^{1/2}$. Its elements involve the structure constants (3.20), which are lattice sums converging when $\rho^2 \gtrsim E$. The total number of operators required to evaluate the determinant is therefore

$$\nu \approx \frac{(4\pi R(E)^{1/2})^3}{3} + \frac{4\pi R(E)^{1/2} \times \pi E}{4} \sim E^{3/2} \tag{8.12}$$

Comparison of (8.11) with (8.12) shows that calculation of the classical path sum involves many more operations than calculation of the KKR determinant. To deter-

mine the energy levels, the calculation must be carried out over a range of energies. Here the path sum has a slight advantage because the determinants $D_\pm$ are independent of energy, and need only be evaluated once, whereas the structure constants (3.17) must be computed afresh for each energy. On the other hand, in computing the spectrum for a range of disc radii R the KKR method has the advantage, because the structure constants are independent of R and need only be evaluated once, whereas $D_\pm$ must be computed afresh for each R.

On balance, however, the outcome of this analysis is that in calculating the complete spectrum of energy levels for all values of R the KKR method is vastly superior to the classical path sum. There is, however, an important class of circumstances in which the path sum is immediately informative, namely when what is required is the spectrum *smoothed on some given energy scale* δE, rather than the complete set of discrete levels. As first shown by Balian and Bloch [10], the smoothed spectrum is given not by $d(E)$ but in terms of its analytic continuation evaluated at $E + i\,\delta E$, which broadens the levels into resonances. In the classical path sum, each orbit acquires the weighting factor $\exp[-\pi L\,\delta E/(E)^{1/2}]$ so that the very long paths give negligible contributions. If the smoothing δE is greater than the level spacing $(\bar{d})^{-1}$ and less than the oscillation "wavelength" ΔE (Eq. (8.7)) for the shortest path, the clustering of the levels is determined essentially by a finite number of orbits. Then the estimates (8.6) show that on these scales the nonisolated orbits dominate the spectrum.

This seems a disappointing conclusion, because the characteristic new features introduced by ergodicity are the isolated unstable orbits, and these, as well as being too numerous to form a basis for computing individual levels, are individually too insignificant to be discerned in the smoothed spectrum. But the effect of the isolated orbits can be seen in the R-dependence of the levels. For these orbits, L depends on R, so that their contribution to d will oscillate as R varies. For the nonisolated orbits, L does not depend on R (the "measure" factors in $d_{\mathrm{osc}}^{(\alpha)}$ (Eq. (8.3)) do depend on R, but do not give rise to any oscillatory behaviour). It is plausible that the oscillatory behaviour originating in the isolated orbits is the cause of the oscillations of the individual energy levels with R, clearly visible on Fig. 4; their "wavelength" should be ΔR, given by

$$kL(R + \Delta R) - kL(R) = 2\pi, \qquad \text{i.e., } \Delta R = \left(\frac{\partial L}{\partial R}\,E^{1/2}\right)^{-1}. \tag{8.13}$$

For the shortest paths of the desymmetrized billiard (Fig. 14), $\partial L/\partial R \sim 2$, predicting $\Delta R \approx 1/2(E)^{1/2}$, in overall agreement with the trends on Fig. 4.

It is natural to ask whether the clustering of levels on fairly large scales can be reduced by choosing an ergodic system where all closed orbits are isolated, so that the contribution $d_{\mathrm{osc}}^{(\alpha)}$ is zero. The "stadium" studied numerically by McDonald and Kaufman [32] is not such a system, because it possesses one family of "transverse" nonisolated orbits. But geodesic motion on a closed surface with constant negative curvature is such a system; all closed orbits are isolated and unstable, and it follows from the celebrated Selberg trace formula, well reviewed by McKean [33], that they determine the spectrum of the Laplacian *exactly* (i.e., not merely asymptotically).

In this and similar systems, $\mathcal{N}_{\text{osc}} \sim \mathcal{O}(k^0)$. It is interesting to speculate on whether the level density oscillations can be further suppressed. I suggest that the modes of an enclosure ("auditorium") with a *fractal* boundary with Hausdorff dimension D (Mandelbrot [34]), where $1 < D < 2$, would cluster very weakly indeed, with $\mathcal{N}_{\text{osc}} \sim \mathcal{O}(k^{-D})$, so that the levels would be very regularly distributed; the argument is given in Appendix I.

9. Concluding Remarks

The "holy grail" of present-day semiclassical mechanics is a quantization condition for ergodic systems, analogous to the Bohr–Sommerfeld conditions quantizing the phase-space tori of integrable systems. In the case of Sinai's billiard, it seems from the detailed study presented here that this quantum condition is the *vanishing of the KKR determinant* (3.15), semiclassically approximated by truncation at $|l| = kR$ and replacement of Bessel functions by their asymptotic approximations. This condition expresses the constructive interference required to form a quantum state far more compactly than does the representation (8.1)–(8.4) in terms of a sum over classical closed orbits, in the same way that for integrable systems the Bohr–Sommerfeld conditions express constructive interference more compactly than the analogous path sum [11]. It is possible to generalize the KKR determinant to give a semiclassical quantization condition valid for any billiard system; I shall publish an account of this method separately.

Ergodic systems, like integrable systems, are special cases. A generic Hamiltonian system has its phase space divided into chaotic regions (surrounding unstable closed orbits) and regions filled with tori (surrounding stable closed orbits). However, my calculations support the view that quantum ergodic systems do represent the behaviour of generic quantum systems in that their levels do not degenerate when one parameter is varied (Section 4), and their level spacings are consistent with the level repulsion (Section 5) predicted in the generic case.

Similarly, I expect the eigenfunctions of ergodic systems to display the property, proved by Uhlenbeck [37] for the generic case, that the nodal lines do not intersect. Stratt *et al.* [38] have illustrated this theorem with computations for a quasi-integrable system. For ergodic systems, nodal noncrossing has been discovered computationally for the stadium by McDonald and Kaufman [32], and experimentally by Ede (unpublished) who studied the elastic vibrations of a metal plate cut to the shape of Fig. 3 to simulate the desymmetrized Sinai billiard.

Finally, I believe that the semiclassical multiple scattering theory developed in Section 7 could have an interest outside the immediate context of Sinai's billiard, in the theory of the smoothed density of electronic energy levels in condensed crystalline and disordered materials. The conventional expansion, based on (6.11) and (7.4), contains unphysical contributions from scatterings between distant discs, which carry almost the same weight as scatterings between neighbouring discs. In the semiclassical resummation (8.3) and (8.4), the unphysical contributions are cancelled byhigher-order

multiple scattering terms involving "ghost" discs, leaving only scatterings between mutually accessible discs. This cancellation is exact in the semiclassical limit and should hold to a good approximation for short waves.

APPENDIX A: LATTICE GREEN FUNCTION FORMULAE

The derivation of (3.11) from (3.10), with G given by (3.5), depends on a double application of the formula [14]

$$\exp[il\phi_{\mathbf{a}+\mathbf{b}}]\, H_l^{(1)}(|\,\mathbf{a}+\mathbf{b}\,|) = \sum_{l'=-\infty}^{\infty} H_{l-l'}^{(1)}(|\,\mathbf{b}\,|)\, J_{l'}(|\,\mathbf{a}\,|) \exp[i\{(l-l')\,\phi_{\mathbf{b}} + l'\phi_{\mathbf{a}}\}], \quad \text{(A.1)}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vectors in the plane with polar angles $\phi_{\mathbf{a}}$ and $\phi_{\mathbf{b}}$, satisfying $|\,\mathbf{b}\,| > |\,\mathbf{a}\,|$. In Eq. (3.10), $|\,\mathbf{r}'+\boldsymbol{\rho}\,| > r$ for all $\boldsymbol{\rho}$ including $\boldsymbol{\rho}=0$ (cf. Fig. 2). Therefore

$$\sum_{\boldsymbol{\rho}} G(\mathbf{r}, \mathbf{r}'+\boldsymbol{\rho}) = \frac{-i}{4} \sum_{\boldsymbol{\rho}} \sum_{l} J_l(kr)\, H_{-l}^{(1)}(k\,|\,\mathbf{r}'+\boldsymbol{\rho}\,|)(-1)^l \times \exp[i(l\phi - l\phi_{\mathbf{r}'+\boldsymbol{\rho}})]. \quad \text{(A.2)}$$

Using $H_{-l}\exp(i\pi l) = H_l$ for the term $\boldsymbol{\rho}=0$, and transforming the terms $\boldsymbol{\rho}\neq 0$, bearing in mind that $\rho > r'$, gives

$$\sum_{\boldsymbol{\rho}} G(\mathbf{r}, \mathbf{r}'+\boldsymbol{\rho}) = \frac{-i}{4} \sum_{l} e^{il\phi} J_l(kr) \Big\{ e^{-il\phi'} H_l^{(1)}(kr') + \sum_{\boldsymbol{\rho}}{}' (-1)^l \sum_{l'} H_{-l-l'}(k\rho)\, J_{l'}(kr') \exp[i\{-(l+l')\,\phi_{\boldsymbol{\rho}} + l'\phi'\}] \Big\}, \quad \text{(A.2')}$$

from which (3.11) follows on replacing l' by $-l'$ and letting $r' \to r \to R$.

The derivation of (3.16) from (3.15) proceeds by first splitting $S_{l-l'}$ (Eq. (3.14)) into its real and imaginary parts, and noting that the matrices in the two equations are identical apart from the terms

$$T_{ll'} \equiv -i\delta_{ll'} - i \sum_{\boldsymbol{\rho}}{}' J_{l-l'}(k\rho) \exp[i(l-l')\,\phi_{\boldsymbol{\rho}}]. \quad \text{(A.3)}$$

On writing $\sum_{\boldsymbol{\rho}}' = \sum_{\boldsymbol{\rho}}$ − term in $\boldsymbol{\rho}=0$, this becomes

$$T_{ll'} = -i \sum_{\boldsymbol{\rho}} J_{l-l'}(k\rho) \exp[i(l-l')\,\phi_{\boldsymbol{\rho}}]. \quad \text{(A.4)}$$

The Bessel function can be represented as an integral over the angles $\phi_{\mathbf{k}}$ of a vector $\mathbf{k}$ whose length is k, as follows:

$$T_{ll'} = \frac{-i}{2\pi} \sum_{\boldsymbol{\rho}} \int_0^{2\pi} d\phi_{\mathbf{k}} \exp[i\{(l-l')\,\phi_{\mathbf{k}} + \mathbf{k}\cdot\boldsymbol{\rho}\}]. \quad \text{(A.5)}$$

Now the Poisson summation formula gives

$$\sum_{\rho} e^{i\mathbf{k}\cdot\boldsymbol{\rho}} = (2\pi)^2 \sum_{\mathbf{g}} \delta(\mathbf{k} - \mathbf{g}), \tag{A.6}$$

where $\mathbf{g}$ are the vectors $\mathbf{g} = 2\pi(m, n)$ of the reciprocal lattice. Therefore

$$\begin{aligned} T_{ll'} = 0 \qquad & \text{unless } k^2 = (2\pi)^2(m^2 + n^2), \\ & \text{i.e.,} \quad E = m^2 + n^2, \end{aligned} \tag{A.7}$$

so that the two determinants (3.15) and (3.16) are equivalent provided the energy does not equal one of the eigenvalues of the "unperturbed" ($R = 0$) quantum billiard.

APPENDIX B: Transformation of KKR Structure Constants

In deriving formula (3.20) for $S_l(E)$, the first step is to write the Hankel functions in (3.14) as integrals over momentum space $\mathbf{q}$:

$$-iH_l^{(1)}(k\rho)\exp[il\phi_\rho] = \frac{1}{\pi^2 k^l}\iint d\mathbf{q}\,\frac{(q_x + iq_y)^l}{k^2 + i\epsilon - q^2}\,e^{i\mathbf{q}\cdot\boldsymbol{\rho}}. \tag{B.1}$$

This relation is easily verified with the aid of standard Bessel-function formulae [16]. Next, the substitution

$$\frac{1}{k^2 + i\epsilon - q^2} = -\int_0^{i\infty} d\xi \exp[(k^2 + i\epsilon - q^2)\,\xi] \tag{B.2}$$

is made, with the contour of integration as shown on Fig. B1, where X is an arbitrary real positive number.

The Ewald procedure consists in substituting (B.2) into (B.1) and (B.1) into (3.14), splitting the ξ integration into two ranges as on Fig. B1 and employing the Poisson relation (A.6) for the summation over the terms with ξ-range X to $i\infty$. This gives

$$S_l(E) = S_{l1}(E) + S_{l2}(E) + S_{l3}(E), \tag{B.3}$$

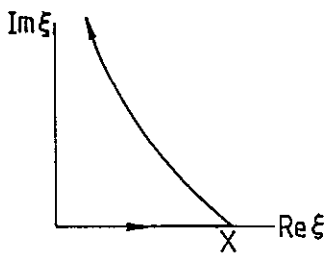

Fig. B1. Integration path in Ewald summation of structure constants.

where

$$S_{l1}(E) = -\frac{4}{k^l}\iint d\mathbf{q}\sum_{\mathbf{g}}\delta(\mathbf{q}-\mathbf{g})(q_x+iq_y)^l\int_X^{i\infty}d\xi\exp[(k^2+i\epsilon-q^2)\,\xi]$$

$$= \frac{4}{k^l}\sum_{\mathbf{g}}\frac{(g_x+ig_y)^l}{k^2+i\epsilon-g^2}\exp[(k^2+i\epsilon-g^2)\,X], \tag{B.4}$$

$$S_{l2}(E) = \frac{1}{\pi^2k^l}\iint d\mathbf{q}(q_x+iq_y)^l\int_X^{i\infty}d\xi\exp[(k^2+i\epsilon-q^2)\,\xi]$$

$$= \left\{-\frac{1}{\pi}Ei(k^2X)+i\right\}\delta_{l0}\,, \tag{B.5}$$

and

$$S_{l3}(E) = -\frac{1}{\pi^2k^l}\int_0^X d\xi\sum_{\boldsymbol{\rho}}{}'\iint d\mathbf{q}(q_x+iq_y)^l\exp[i\mathbf{q}\cdot\boldsymbol{\rho}+\xi(k^2+i\epsilon-q^2)]. \tag{B.6}$$

I shall discuss these three terms separately. Scaling the g-lattice by 2π into the $\boldsymbol{\rho}$-lattice transforms S_{l1} into

$$S_{l1}(E) = \frac{1}{\pi^2}\sum_{\boldsymbol{\rho}}\frac{\exp[(l/2)\ln(\rho^2/E)+4\pi^2XE(1-\rho^2/E)]}{E+i\epsilon-\rho^2}\cos[l\phi_{\boldsymbol{\rho}}]. \tag{B.7}$$

The singularities occur where $\rho^2 = E$ and it is natural to choose X to ensure that this value of ρ^2 corresponds to the maximum value of the exponential factor, so that for any nonzero l the summation is dominated by lattice vectors with ρ near $(E)^{1/2}$. This is achieved with

$$4\pi^2XE = l/2, \tag{B.8}$$

whereupon $S_{l1}(E)$ reduces exactly to the first sum in (3.20). For $l = 0$ this procedure fails, and it is simplest to choose

$$4\pi^2XE = Q, \tag{B.9}$$

whereupon $S_{l1}(E)$ and the real part of $S_{l2}(E)$ (Eq. (B.5)) reduce exactly to the corresponding terms in (3.20).

Thus (3.20) is justified provided $S_{l3}(E)$ can be neglected. To show this, (B.6) is first rewritten as

$$S_{l3}(E) = -\frac{1}{\pi^2k^l}\sum_{\boldsymbol{\rho}}{}'\int_0^X d\xi\,e^{k^2\xi}\left(\frac{\partial}{\partial\rho_x}+\frac{i\partial}{\partial\rho_y}\right)^l\iint d\mathbf{q}\exp[i\mathbf{q}\cdot\boldsymbol{\rho}-q^2\xi]$$

$$= -\frac{1}{\pi k^l}\sum_{\boldsymbol{\rho}}{}'\int_0^X\frac{d\xi}{\xi}\,e^{k^2\xi}\left(\frac{\partial}{\partial\rho_x}+\frac{i\partial}{\partial\rho_y}\right)^l e^{-\rho^2/4\xi}$$

$$= -\frac{1}{\pi}\sum_{\boldsymbol{\rho}}{}'\cos[l\phi_{\boldsymbol{\rho}}]\int_0^X\frac{d\xi}{\xi}\left(\frac{\rho}{2k\xi}\right)^l\exp[-\rho^2/4\xi+k^2\xi], \tag{B.10}$$

where the last step made use of the identity

$$\left(\frac{\partial}{\partial\rho_x}+\frac{i\partial}{\partial\rho_y}\right)^l f(\rho^2)=2^l\rho^l\exp[il\phi_\rho]\frac{d^l}{d(\rho^2)^l}f(\rho^2). \tag{B.11}$$

Next (assuming $l\neq 0$) the new integration variable

$$s\equiv -\ln[\xi/X] \tag{B.12}$$

is introduced, together with the parameter

$$Z_\rho\equiv 4\pi^2\rho^2E/l^2. \tag{B.13}$$

Then, with the choice (B.8) for X, $S_{l3}(E)$ becomes

$$S_{l3}(E)=-\sum_\rho{}'\cos[l\phi_\rho]\,T_l(Z_\rho), \tag{B.14}$$

where

$$T_l(Z)\equiv\frac{Z^{l/2}}{\pi}\int_0^\infty ds\exp\left[l\left(s+\frac{1}{2}e^{-s}-\frac{Z}{2}e^{+s}\right)\right]. \tag{B.15}$$

In assessing the importance of these terms, it is necessary only to consider $E\geqslant 5$, because 5 is the ground state energy when $R=0$ (Eq. (3.26)) and for larger R all eigenstates must have higher energy because of the smaller area accessible to the waves. The simplest case is $l=0$. Then (B.10) depends on the value of

$$\begin{aligned} I&=\int_0^{Q/4\pi^2E}\frac{d\xi}{\xi}\exp[-\rho^2/4\xi+4\pi^2E\xi]\\ &=\int_{Q^{-1}}^\infty\frac{du}{u}\exp[u^{-1}-\rho^2\pi^2Eu]. \end{aligned} \tag{B.16}$$

This decreases as ρ and E increase. When Q is not too large, I can be well approximated by

$$I\approx\frac{\exp[-\rho^2\pi^2E/Q+Q]}{(\rho^2\pi^2E/Q+Q)}. \tag{B.17}$$

The choice $Q=3$, employed in the numerical computations reported in Section 3, lies well within the range of validity of this formula. Then in the worst case ($\rho=1$, $E=5$), $I\approx 7.4\times 10^{-8}$, implying that $S_{l3}(E)$ is always negligible when $l=0$.

Analysis of S_{l3} for $l\neq 0$ depends on the terms $T_l(Z)$ as given by (B.15). The value of these terms depends on Z_ρ. Consider first those l for which the phase shift factor $\tan\eta_{4n}$ in (3.23) is not negligible, namely $l<4n_{\max}$ where $n_{\max}$ is given by (3.29). In the structure constants in (3.23) the corresponding l satisfy $l<8n_{\max}$, the extreme

corresponding to diagonal elements in the determinant. For such l, Z_ρ as defined by (B.13) satisfies

$$Z_\rho > \frac{4\pi^2\rho^2 E}{(8n_{\max})^2} = \frac{l^2}{4R^2} \geqslant 1, \tag{B.18}$$

since $R \leqslant \frac{1}{2}$. When $Z > 1$ the exponent in the integrand in (B.15) has no real stationary points, and for large l the main contribution comes from the endpoint $s = 0$, giving the asymptotic approximation

$$T_l(Z) \approx \frac{2\exp[-(l/2)(Z - \ln Z - 1)]}{\pi(Z-1)} \qquad (Z > 1, l \gg 0). \tag{B.19}$$

This gets rapidly smaller as E (and hence Z) increases for fixed l. Even for $E = 5$, when $\rho = 1$ and l is as large as possible ($8n_{\max}$), (B.19) gives $T_l \sim 0.015$ when $R = 0.3$.

Consider next the value of S_{l3} when $l \gg 4n_{\max}$, i.e., when the phase-shift factor $\tan \eta_{4n}$ is very small. It is necessary to study this in order to establish the convergence of the determinant in (3.23). In these circumstances Z_ρ (Eq. (B.12)) can be less than unity. Then the exponent in (B.15) has a maximum when

$$s = \ln[Z^{-1} + (Z^{-1} - 1)^{1/2}], \tag{B.20}$$

and for large l the integral for T_l can be evaluated by the method of steepest descent, giving the asymptotic approximation

$$T_l(Z) \approx \left(\frac{2}{\pi l}\right)^{1/2} \frac{\exp[l\{\ln[Z^{-1/2} + (Z^{-1} - 1)^{1/2}] - (1 - Z)^{1/2}\}]}{(1 - Z)^{1/4}} \qquad (Z < 1, l \gg 0). \tag{B.21}$$

In the limit $l \to \infty$ with E and ρ fixed, this formula, combined with (B.13), gives

$$T_l \to \left(\frac{2}{\pi l}\right)^{1/2} \left(\frac{l}{e\pi\rho(E)^{1/2}}\right)^l \qquad \text{as } l \to \infty. \tag{B.22}$$

Therefore the structure constants diverge as $l \to \infty$, and convergence of (3.23) depends on the phase shifts $\tan \eta_l$.

From (3.13) and standard Bessel-function asymptotic formulae for the case argument $\ll$ order [14],

$$\tan \eta_l \to -\frac{1}{2}\left(\frac{e\pi R(E)^{1/2}}{l}\right)^{2l} \qquad \text{as } l \to \infty. \tag{B.22'}$$

The diagonal terms of the determinant (3.23) therefore involve the asymptotic dependence

$$\tan \eta_{4n}(E)\, T_{8n}(E) \to -\frac{1}{4((\pi n)^{1/2})}\left(\frac{2R}{\rho}\right)^{8n}. \tag{B.23}$$

This vanishes as $n \to \infty$ because $\rho \geqslant 1$ and $R < \frac{1}{2}$, thus ensuring the convergence of the KKR determinant. It is interesting that the KKR method makes sense only when $R < \frac{1}{2}$, i.e., when the discs do not overlap.

One case has been left out of the foregoing asymptotics, namely the "marginal" case $n \sim n_{\max}$ for $R \to 0.5$, when $Z_\rho \sim 1$ (Eq. (B.18)). Then the exponent in (B.15) has a double stationary point at $s = 0$, giving the asymptotic approximation

$$T_l \approx \frac{1}{3\pi}\left(\frac{6}{l}\right)^{1/3} \Gamma\left(\frac{1}{3}\right) = \frac{0.516506}{l^{1/3}} \qquad (Z = 1,\ l \to \infty). \tag{B.24}$$

The phase shift factor is

$$\tan \eta_l = \frac{J_l(l)}{Y_l(l)} \approx -\frac{1}{3^{1/2}} \qquad \text{as } l \to \infty \tag{B.25}$$

so that in the determinant the important terms are

$$\tan \eta_{4n_{\max}} T_{8n_{\max}} \approx \frac{0.517}{3^{1/2}(2\pi(E)^{1/2})^{1/3}}. \tag{B.26}$$

For $E = 10$, this equals 0.11, while for $E = 100$ it equals 0.075, so that even in this most unfavourable case the term S_{l3} ((B.4) and (B.5)) is small in comparison with S_{l1}.

The conclusion is that for energies ($E \geqslant 5$) of interest in the quantum billiard problem the structure constants can always be approximated to high accuracy by the formulae (3.20).

APPENDIX C: Perturbation Theory for Small Discs

As $R \to 0$ the phase shifts (3.13) vanish as follows:

$$\tan \eta_l(E) \approx \frac{-\pi(\frac{1}{2}kR)^{2l}}{\Gamma(l)\,\Gamma(l+1)}. \tag{C.1}$$

In a formal expansion of the desymmetrized determinant (3.23) in powers of R, the leading term is $n = n' = 1$, giving

$$1 + \tan \eta_4(E)\{S_0(E) - S_8(E)\} = 0 \tag{C.2}$$

as the equation for eigenvalues E. Because $\tan \eta_4$ is small the roots must lie near the singularities of the structure constants as given by (3.20). These singularities are the free-particle eigenvalues (3.26).

Considering the level(s) corresponding to lattice vectors of length ρ and defining the lowest-order correction Δ by

$$E = \rho^2 + \Delta, \tag{C.3}$$

Eq. (C.2) becomes, on using (3.20) and (C.1),

$$1 - \frac{\pi(\pi R\rho)^8 \times 8}{\Gamma(4)\,\Gamma(5)\,\pi^2 \Delta} \sum_{\phi_\rho} (1 - \cos[8\phi_\rho]) = 0. \tag{C.4}$$

The summation is over all lattice vectors with length ρ lying within the octant $0 < \phi_\rho < \pi/4$ of the ρ plane; for nondegenerate unperturbed states there is one such vector. Solution of (C.4) gives

$$\Delta = \frac{\pi^7 \rho^8 R^8}{18} \sum_{\phi_\rho} (1 - \cos[8\phi_\rho]). \tag{C.5}$$

This perturbation theory shows that for given R the corrections increase rapidly with E, as is clear from Fig. 4. Of course the theory is valid only when Δ is small in comparison with the unperturbed level spacings. Since this is of order $\pi/8$ (Section 6), R is thus restricted by

$$R < \left(\frac{18}{8\pi^6}\right)^{1/8} \times \frac{1}{E^{1/2}} \sim \frac{0.47}{E^{1/2}}. \tag{C.6}$$

This is plotted as the dashed line on Fig. 4, and corresponds fairly well to the limit of smooth variation of the eigenvalues.

Equation (C.5) predicts that all members of a group of degenerate states suffers the same correction, proportional to R^8. At this level of perturbation theory there is no splitting of degenerate levels. To study the splitting of levels with multiplicity two is is sufficient to include terms of one order smaller in R. Instead of (C.2), the determinant now becomes

$$\begin{vmatrix} 1 + \tan \eta_4(S_0 - S_8) & \tan \eta_4(S_4 - S_{12}) \\ \tan \eta_8(S_4 - S_{12}) & 1 + \tan \eta_8(S_0 - S_{16}) \end{vmatrix} = 0, \tag{C.7}$$

and includes terms through order R^{16}.

Introduction of the correction Δ from (C.3) and the limiting forms (C.1), and multiplication by Δ^2, gives a quadratic equation for Δ with two solutions Δ_1 and Δ_2. Up to terms of lowest order in R, Δ_1 is given by (C.5), and Δ_2 by

$$\Delta_2 = \frac{\left(\pi^{15}\rho^{16}R^{16}\{(\sum_{\phi_\rho}(1 - \cos[8\phi_\rho]))(\sum_{\phi_\rho}(1 - \cos[16\phi_\rho])) - (\sum_\rho(\cos[4\phi_\rho] - \cos[12\phi_\rho]))^2\}\right)}{(8!)^2 \sum_{\phi_\rho}(1 - \cos[8\phi_\rho])}. \tag{C.8}$$

If this equation is applied to a nondegenerate state it predicts $\Delta_2 = 0$ by virtue of the identity

$$(1 - \cos 2\theta)(1 - \cos 4\theta) - (\cos\theta - \cos 3\theta)^2 = 0. \tag{C.9}$$

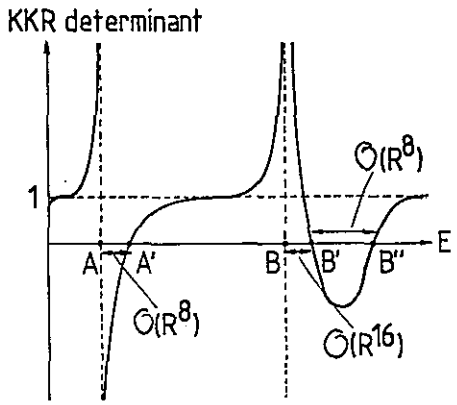

FIG. C1. Behaviour of the KKR determinant for small R near a nondegenerate unperturbed level (A) and a degenerate unperturbed level of multiplicity two (B). The perturbed levels are A', B' and B''.

This does not correspond to an extra level, because $\Delta = 0$ is not a solution of (C.7) although it satisfies the quadratic equation derived from it. When applied to a state with multiplicity two it shows that the degeneracy splits by $\mathcal{O}(R^8)$ with the lower level deviating from its unperturbed value by $\mathcal{O}(R^{16})$; this behaviour can be seen in the two examples on Fig. 4, where $E = 65$ and 85.

Figure C.1 shows the behaviour of the KKR determinant (3.23) as a function of E for small R near a nondegenerate level and near a level with multiplicity two.

To understand the splitting under perturbation of levels with multiplicity N would require analysis of an $N \times N$ truncated KKR determinant.

APPENDIX D: EIGENFUNCTIONS NEAR A DEGENERACY

Consider a family of systems with Hamiltonian $\hat{H}(A, B)$, and let there be a degeneracy at A^*, B^* where two orthogoual eigenstates $|u\rangle$, $|v\rangle$ have the same energy E^*, i.e.,

$$\hat{H}(A^*, B^*)\,|u\rangle = E^*\,|u\rangle, \qquad \hat{H}(A^*, B^*)\,|v\rangle = E^*\,|v\rangle. \tag{D.1}$$

Consider eigenstates $|\psi\rangle$ corresponding to parameters close to A^*, B^*; these satisfy

$$\hat{H}(A, B)\,|\psi\rangle = E(A, B)\,|\psi\rangle. \tag{D.2}$$

Define

$$\left.\begin{aligned} &\Delta A \equiv A - A^*, \qquad \Delta B \equiv B - B^*, \qquad \Delta E(A, B) \equiv E(A, B) - E^*, \\ &\Delta H \equiv \Delta A \frac{\partial \hat{H}}{\partial A}(A^*, B^*) + \Delta B \frac{\partial \hat{H}}{\partial B}(A^*, B^*) \\ &\quad \approx H(A, B) - H(A^*, B^*). \end{aligned}\right\} \tag{D.3}$$

When ΔA and ΔB are small, $|\psi\rangle$ will be a normalized linear combination of $|u\rangle$ and $|v\rangle$, which can be written as

$$|\psi\rangle = \cos\chi(A, B)\,|u\rangle - \sin\chi(A, B)\,|v\rangle. \tag{D.4}$$

The problem is to see how the eigenvalues ΔE and eigenfunctions ψ vary as A, B is taken round a small circuit C surrounding A^*, B^* (cf. Fig. 5).

On substituting (D.3) and (D.4) into (D.2), and using (D.1), Schrödinger's equation becomes

$$\cos\chi\,\Delta\hat{H}\,|u\rangle - \sin\chi\,\Delta\hat{H}\,|v\rangle = \Delta E(\cos\chi\,|u\rangle - \sin\chi\,|v\rangle). \tag{D.5}$$

Multiplication by $\langle u|$ and $\langle v|$ gives

$$\begin{aligned}\cos\chi\langle u|\Delta\hat{H}|u\rangle - \sin\chi\langle u|\Delta\hat{H}|v\rangle &= \Delta E\cos\chi,\\ \cos\chi\langle v|\Delta\hat{H}|u\rangle - \sin\chi\langle v|\Delta\hat{H}|v\rangle &= -\Delta E\sin\chi.\end{aligned} \tag{D.6}$$

These equations have two solutions for ΔE, namely

$$\begin{aligned}\Delta E(A, B) = \tfrac{1}{2}\{&\langle u|\Delta\hat{H}|u\rangle + \langle v|\Delta\hat{H}|v\rangle \pm [(\langle u|\Delta\hat{H}|u\rangle - \langle v|\Delta\hat{H}|v\rangle)^2\\ &+ 4(\langle u|\Delta\hat{H}|v\rangle)^2]^{1/2}\}.\end{aligned} \tag{D.7}$$

Since the matrix elements of $\Delta\hat{H}$ are linear combinations of ΔA and ΔB, $\Delta E(A, B)$ does indeed correspond to two sheets of a double cone, as claimed in Section 4 and sketched in Fig. 5.

The values of χ satisfying (D.6) are given by

$$\tan[2\chi(A, B)] = \frac{2\langle u|\Delta\hat{H}|v\rangle}{\langle v|\Delta\hat{H}|v\rangle - \langle u|\Delta\hat{H}|u\rangle}. \tag{D.8}$$

For each (A, B) this has two solutions χ, differeing by $\pi/2$ and giving the orthogonal eigenfunctions (D.4) coresponding to the two sheets of the cone. Now define

$$\Delta A \equiv r\cos\phi, \qquad \Delta B \equiv r\sin\phi, \tag{D.9}$$

so that ϕ is an angle parameterising C on Fig. 5. Use of (D.3) gives

$$\tan[2\chi(\phi)] = \frac{2(\langle u|\partial\hat{H}/\partial A|v\rangle + \tan\phi\langle u|\partial\hat{H}/\partial B|v\rangle)}{\langle v|\partial\hat{H}/\partial A|v\rangle - \langle u|\partial\hat{H}/\partial A|u\rangle + \tan\phi(\langle v|\partial\hat{H}/\partial B|v\rangle - \langle u|\partial\hat{H}/\partial B|u\rangle)}. \tag{D.10}$$

which does not involve r. Sufficiently close to A^*, B^*, then, ψ depends only on the angular distance round C. For a half-circuit, ϕ changes by π and (D.10) shows that χ, which is a smooth function of ϕ, changes by $\pi/2$, so that the two solutions ψ have become interchanged. For a complete circuit, ϕ changes by 2π and χ changes by π, from which it follows that each eigenfunction (Eq. (D.4)) changes sign, as claimed in Section 4.

APPENDIX E: ASYMPTOTIC PROBABILITY THAT A NUMBER IS THE SUM OF TWO SQUARES

The following argument is due to Dr. J. H. Hannay (private communication). It is known [21] that an integer E is the sum of two squares if and only if its prime decomposition contains no odd powers of primes of the form -1 mod 4. Such primes will be denoted by p; the first few p are 3, 7, 11, 19, 23. The required probability $\mathscr{P}(E)$ is the fraction of numbers near E that contain no odd powers of any p, so that

$$\mathscr{P}(E) = \prod_{p=3}^{\infty} \mathscr{P}(E, p), \tag{E.1}$$

where $\mathscr{P}(E, p)$ is the probability that E contains no odd powers of p. Obviously $\mathscr{P}(E, p) = 1$ when $p > E$. The probability that E does not contain p as a factor is $1 - p^{-1}$. This excludes too many numbers, and those containing p^2 must be reinstated. But now numbers containing p^3 must be re-excluded, and then numbers containing p^4 must be re-reinstated, the procedure stopping when the factor exceeds E. Therefore

$$\mathscr{P}(E, p) = \sum_{n=1}^{n(E,p)} \frac{(-1)^n}{p^n}, \qquad \text{where } n(E, p) = \log E/\log p. \tag{E.2}$$

The sum is elementary, and gives

$$\mathscr{P}(E, p) = \frac{1 + (-1)^{n(E,p)}/p^{n(E,p)+1}}{1 + p^{-1}}. \tag{E.3}$$

$\mathscr{P}(E)$ now becomes

$$\mathscr{P}(E) = \prod_{p=3}^{E} \frac{1 + (-1)^{n(E,p)}/p^{n(E,p)+1}}{1 + p^{-1}} \tag{E.4}$$

For large E the numerators are close to unity, and so

$$\mathscr{P}(E) \approx \exp\left[-\sum_{p=3}^{\infty} \ln(1 + p^{-1})\right]. \tag{E.5}$$

The sum can be replaced by an integral using the fact that numbers of type p constitute approximately half of all primes, and that the probability of any given number being prime is $1/\ln p$ [21]. Therefore

$$\begin{aligned}\mathscr{P}(E) &\approx \exp\left[-\frac{1}{2}\int_3^E dp\, \frac{\ln(1 + p^{-1})}{\ln p}\right] = \text{constant} \times \exp\left[-\frac{1}{2}\int_3^{\infty} \frac{dp}{p \ln p}\right] \\ &\approx \text{const}/(\ln E)^{1/2},\end{aligned} \tag{E.6}$$

as claimed in Section 4.

APPENDIX F: ASYMPTOTICS OF PHASE SHIFT SUM

The sum to be evaluated is the second term in (6.11), which will be denoted by $\sum(E)$. Asymptotically, (i.e., for large k) the phase shifts defined by (3.13) may be approximated using the Debye formula (7.12) for the Bessel functions, giving

$$\begin{aligned} \eta_l &\approx |l| \arccos\frac{l}{kR} - (k^2R^2 - l^2)^{1/2} - \pi/4 \qquad (|l| < kR), \\ &\approx 0 \qquad\qquad\qquad\qquad\qquad\qquad (|l| > kR). \end{aligned} \tag{F.1}$$

(An alternative way to obtain this result would be to apply the WKB method [28] to the differential equations for the radial wave functions.) The Poisson sum formula (7.8) now gives

$$\begin{aligned} \sum(E) &= \frac{1}{\pi}\sum_{l=-\infty}^{\infty}\eta_l(E) \\ &\approx \frac{kR}{\pi}\sum_{m=-\infty}^{\infty}\int_{-1}^{1} dx\, e^{2\pi i m k R x}\{kR(|x| \arccos x - (1-x^2)^{1/2}) - \pi/4\}. \end{aligned} \tag{F.2}$$

The steady (nonoscillatory) contribution $\overline{\sum}$ comes from the term $m = 0$, which is easily evaluated to give

$$\overline{\sum}(E) = -\pi^2R^2E - \pi R(E)^{1/2}, \tag{F.3}$$

whose dominant term is the desired result (6.13).

Oscillatory correctijns $\mathcal{N}_{\text{osc}}^{(2)}$ (Eq. (7.9)) to $\overline{\sum}$ are contained in the Poisson sum terms with $m \neq 0$, and come from the regions near $|x| = 1$ in the integrals. Estimated on the basis of (F.2) the corrections would be

$$\mathcal{N}_{\text{osc}}^{(2)} \approx -\sum_{m=1}^{\infty}\frac{\sin 2\pi kRm}{m} + \mathcal{O}(k^{-1/2}). \tag{F.4}$$

These are of the same order of magnitude as the contributions (7.31) to $\mathcal{N}_{\text{osc}}^{(3)}$ from orbits bouncing off discs, and would therefore have to be retained. However, the estimate (F.4) results from the discontinuity in the WKB phase shifts (F.1) at $|l| = kR$, whereas the true phase shifts (3.13) tend smoothly to zero over a range of order $(kR)^{1/3}$ centred on $|l| = kR$. This observation leads to an improved estimate of the asymptotic behaviour of $\mathcal{N}_{\text{osc}}^{(2)}$, namely

$$\mathcal{N}_{\text{osc}}^{(2)} \approx \sum_{m=1}^{\infty} A_m(kR)^{1/3}\sin[2\pi mkR]\exp[-Bm(kR)^{1/3}], \tag{F.5}$$

where A_m and B are constants. The physical meaning of these terms is that they represent "creeping rays" [29, 30] winding m times round the circumference of each disc whilst radiating tangentially and so attenuating exponentially. As $k \to \infty$ the contributions vanish by comparison with those from real orbits, justifying their neglect in Section 7.

APPENDIX G: STATIONARY VALUE OF THE PHASE IN (7.23)

Let this phase be denoted by kL, and substitute the WKB phase shifts $\bar{\eta}(b)$ as given by (7.13) and (F.1). Then

$$L = \sum_{\nu=1}^{n} \Big\{ 2b_\nu \arccos \frac{b_\nu}{R} - 2(R^2 - b_\nu^2)^{1/2} + [\rho_\nu^2 - (b_\nu - b_{\nu+1})^2]^{1/2} - (b_\nu - b_{\nu+1}) \Big(\arccos \Big[\frac{b_\nu - b_{\nu+1}}{\rho_\nu} \Big] + \phi_\nu \Big) \Big\}. \tag{G.1}$$

The problem is to show that L is the length of the orbit. It is clear from Fig. 11 that the length L_ν of path between discs separated by lattice vector ρ_ν, i.e., between scatterings with b_ν and $b_{\nu+1}$, is

$$L_\nu = [\rho_\nu^2 - (b_{\nu+1} - b_\nu)^2]^{1/2} - (R^2 - b_\nu^2)^{1/2} - (R^2 - b_{\nu+1}^2)^{1/2}. \tag{G.2}$$

The total path length is $\sum_{\nu=1}^{n} L_\nu$. On the other hand, the classical relation (7.21) between scattering angles, together with (7.20), gives

$$2b_\nu \arccos \frac{b_\nu}{R} - b_\nu \phi_\nu - b_\nu \arccos \Big[\frac{b_\nu - b_{\nu+1}}{\rho_\nu} \Big] = -b_\nu \phi_{\nu-1} - b_\nu \operatorname{rc}\cos \Big[\frac{b_{\nu-1} - b_\nu}{\rho_{\nu-1}} \Big]. \tag{G.3}$$

Substitution into (G.1) gives

$$L - \sum_{\nu=1}^{n} L_\nu = \sum_{\nu=1}^{n} \Big\{ b_{\nu+1}\phi_\nu - b_\nu \phi_{\nu-1} + b_{\nu+1} \arccos \Big[\frac{b_\nu - b_{\nu+1}}{\rho_\nu} \Big] - b_\nu \arccos \Big[\frac{b_{\nu-1} - b_\nu}{\rho_{\nu-1}} \Big] \Big\},$$

which vanishes on account of the periodicity of the orbit.

APPENDIX H: CLOSED ORBITS WHOSE LENGTH IS LESS THAN $L_{\max}$

Let the number of such orbits be ν. First ν will be calculated in the limit $R \to 0$. Any walk between lattice points is then a closed orbit on the torus, because slight adjustments at the ends of each step will ensure specularity at each intermediate disc and the equality of position and direction at the discs at the ends of the path. Let such a walk have s steps consisting of lattice vectors $\rho_1 \cdots \rho_s$. Since successive discs must be mutually accessible, all these lattice vectors must be primitive, that is their (integer) coordinates must be mutually prime. The length of this path is

$$L = \sum_{j=1}^{s} \rho_j . \tag{H.1}$$

Summing over all such paths, and using a step function Θ to exclude those with $L > L_{\max}$, gives

$$\nu = \sum_{s=2}^{\infty} \sum_{\rho_1} \cdots \sum_{\rho_s} \Theta\left(L_{\max} - \sum_{j=1}^{s} \rho_j\right) \tag{H.2}$$

This is an exact expression. To get an approximation for large $L_{\max}$, the lattice sums can be replaced by integrals as follows:

$$\sum_{\rho} \to 2\pi\gamma \int_0^{\infty} \rho \, d\rho, \tag{H.3}$$

where γ is the probability that a lattice vector is primitive. γ is obtained by subtracting from unity the fractions of pairs of numbers (coordinates of ρ) divisible by successive primes p. This gives

$$\gamma = 1 - \frac{1}{2^2} - \left(1 - \frac{1}{2^2}\right)\frac{1}{3^2} \cdots = \prod_{p} (1 - p^{-2}) = \{\zeta(2)\}^{-1} = \frac{6}{\pi^2} . \tag{H.4}$$

(H.2) now becomes

$$\nu = \sum_{s=2}^{\infty} \left(\frac{12L_{\max}^2}{\pi}\right)^s \int_0^{\infty} x_1 \, dx_1 \cdots \int_0^{\infty} x_s \, dx_s \, \Theta\left(1 - \sum_{j=1}^{s} x_s\right)$$

$$= \sum_{s=2}^{\infty} \left(\frac{12L_{\max}^2}{\pi}\right)^s \frac{1}{(2s)!} = \cosh\left[L_{\max}\left(\frac{12}{\pi}\right)^{1/2}\right] - \frac{6L_{\max}^2}{\pi} - 1 \tag{H.5}$$

$$\to \frac{1}{2} \exp\left[L_{\max}\left(\frac{12}{\pi}\right)^{1/2}\right]. \tag{H.6}$$

The mean number of steps is

$$\bar{s} = \frac{1}{\nu} \sum_{s=2}^{\infty} \left(\frac{12L_{\max}^2}{\pi}\right)^s \frac{s}{(2s)!} \to L_{\max} \left(\frac{3}{\pi}\right)^{1/2}. \tag{H.7}$$

Therefore s is only slightly less than $L_{\max}$, so that most steps are very short and in fact link nearest-neighbour discs.

This result suggests that when R is not zero the exponential dependence of ν on $L_{\max}$ will survive, because in spite of the inaccessibility of distant discs resulting from finite R the neighbouring discs can still be reached. Even on the most extreme restrictive assumption, that an orbit can strike only nearest-neighbour discs, there are three choices at each step, and at most $L_{\max}/(1 - 2R)$ steps, so that

$$\nu \approx \sum_{s=1}^{L_{\max}/(1-2R)} 3^s \to \exp[L_{\max} \ln 3/(1 - 2R)]. \tag{H.8}$$

(The exponent for all the possible glancing collisions between discs lying in two neighbouring rows is somewhat smaller.)

To see that most of the ν paths have been traversed only once, let $\nu^*(L)\, dL$ be the number of singly traversed paths with lengths between L and $L + dL$. Then

$$\nu(L_{\max}) \approx \int_1^{L_{\max}} dL \left\{\nu^*(L) + \frac{1}{2}\nu^*\left(\frac{L}{2}\right) + \frac{1}{3}\nu^*\left(\frac{L}{3}\right) + \cdots\right\}. \tag{H.9}$$

I do not know the general solution of this functional equation for $\nu^*(L)$ given $\nu(L_{\max})$, but taking $\nu^* \sim Le^{AL}$ gives

$$\nu(L_{\max}) \approx \int_1^{L_{\max}} dL \sum_{t=1}^{t} \frac{L}{t^2} \exp[AL/t] \approx \frac{1}{A} \int_1^{L_{\max}} dL \int_0^{AL} dx\, e^x$$

$$\approx \frac{1}{A} \exp[AL_{\max}], \tag{H.10}$$

which has the form already established. The mean number of traversals $\bar{t}$ is therefore

$$\bar{t} \approx \frac{\sum_{t=1}^{L_{\max}} t\nu^*(L_{\max}/t)}{\sum_{t=1}^{L_{\max}} \nu^*(L_{\max}/t)} = \frac{\sum_{t=1}^{L_{\max}} \exp[AL_{\max}/t]}{\sum_{t=1}^{L_{\max}} t^{-1} \exp[AL_{\max}/t]}$$

$$\to 1 \qquad \text{as } L_{\max} \to \infty. \tag{H.11}$$

APPENDIX I: Level Density Oscillations in a Fractal "Auditorium"

By "auditorium" I mean here a two-dimensional enclosed space which shares with real auditoriums the property that its boundary B is partially absorbing, so that the "reverberation" or exponential decay time is finite. This property can be (imper-

fectly) modelled by taking a perfectly reflecting boundary and an imaginary part for the frequency, and has the effect of changing eigenfrequencies into broad resonances. This broadening means that neighbouring eigenfrequencies cannot be distinguished, but is not so large as to smooth away clusters of resonances. Part of the science of auditorium design is an attempt to choose a boundary shape that suppresses the level density oscillations.

It is clear from the asymptotic estimates (8.5) that nonisolated orbits must be avoided because of their stronger oscillations. In an ergodic auditorium with only isolated closed orbits, $\mathcal{N}_{\text{osc}} \sim \mathcal{O}(k^0)$. But this is not the best that can be achieved. Let B be a "fractal" [34], that is a continuous but nondifferentiable curve with Haussdorff dimension D. For such an auditorium, reflection of trajectories is not defined, but of course modes of vibration exist and presumably possess a limiting distribution for high frequencies. An approximate argument leading to an estimate of the strength of level clustering will now be given.

To a mode with wave number k, details of B on scales less than the wavelength $2\pi/k$ are imperceptible, and B can be replaced with a "k-smoothed boundary" B_k. Then the mode number $\mathcal{N}(k)$ can be approximated by $\bar{\mathcal{N}}$ plus an oscillatory contribution consisting of a sum over all closed orbits bouncing off B_k. These orbits will all be isolated and unstable, and the dominant one will be the longest, approximately diametral path with two reflections, traversed once. Its contribution will be given by a formula like (7.32) with R representing the radius of curvature $R(k)$ of B_k at the specular points and $(\rho - 2R)$ the distance L between reflections. For such a highly convoluted surface, $R(k) \ll L$, so that this dominant oscillatory contribution can be written as

$$\mathcal{N} \sim \frac{R(k)}{\pi L} \sin 2kL. \tag{I.1}$$

As k increases, B_k becomes more convoluted as the fractality of B is revealed. Therefore $R(k)$ decreases. To find the law governing this decrease, let B be modelled locally by the Weierstrass–Mandelbrot function $f(x)$, studied by Berry and Lewis [35]. This is defined as

$$f(x) \equiv A \sum_{n=-\infty}^{\infty} \frac{\sin[\gamma^n x]}{\gamma^{n(2-D)}} \qquad (1 < D < 2, \gamma > 1), \tag{I.2}$$

and is a function whose graph is a fractal curve with dimension D and which depends on a parameter γ. k-smoothing simply means omitting terms for which $\gamma^n > k$, so that the fastest-varying term in $f(x)$ is

$$f_k(x) \approx \frac{A \sin kx}{k^{2-D}}, \tag{I.3}$$

and the curvature is

$$\frac{1}{R(k)} \approx \left| \frac{d^2 f_k}{dx^2} \right| \approx k^D A \sin kx, \qquad \text{i.e., } R(k) \sim k^{-D}. \tag{I.4}$$

From (I.1), this implies that for a fractal "auditorium,"

$$\mathcal{N} \sim \frac{\text{const}}{k^D} \sin 2kL. \tag{I.5}$$

It therefore appears that fractal boundaries produce an almost complete suppression of the oscillations in level density, so that the levels themselves should be very regularly spaced. Of course these arguments are hardly rigorous, and it is desirable to check their conclusions by calculating the spectrum, and the density oscillations, for an exactly soluble model. To my knowledge no such model exists to date. One system for which an exact solution might be found is the "auditorium" whose boundary is the Koch snowflake curve [34].

REFERENCES

1. V. I. ARNOL'D, "Mathematical Methods of Classical Mechanics," Sections 49 and 50, Springer-Verlag, New York, 1978 (original Russian edition, 1974).
2. A. EINSTEIN, *Verh. Deut. Phys. Ges.* **19** (1917), 82–92.
3. J. B. KELLER, *Ann. Phys. (N.Y.)* **4** (1958), 180–188.
4. V. P. MASLOV, "Théorie des Perturbations et des Méthodes Asymptotiques," Dunod, Paris, 1972 (original Russian edition, 1965).
5. YA. G. SINAI, *Russ. Math. Surv.* **25**, No. 2 (1970), 137–189.
6. J. KORRINGA, *Physica* **13** (1947), 392–400.
7. W. KOHN AND N. ROSTOKER, *Phys. Rev.* **94** (1954), 1111–1120.
8. J. VON NEUMANN AND E. P. WIGNER, *Physik Z.* **30** (1929), 467–470.
9. H. P. BALTES AND E. R. HILF, "Spectra of Finite Systems," B-I Wissenschaftsverlag, Mannheim, 1978.
10. R. BALIAN AND C. BLOCH, *Ann. Phys. (N.Y.)* **69** (1972), 76–160.
11. M. V. BERRY AND M. TABOR, *Proc. Roy. Soc. Ser. A* **349** (1976), 101–123; *J. Phys. A* **10** (1977), 371–379.
12. M. C. GUTZWILLER, *J. Math. Phys.* **12** (1971), 343–358; *in* "Stochastic Behavior in Classical and Quantum Hamiltonian Systems" (G. Casati and J. Ford, Eds.), Lecture Notes in Physics No. 93, pp. 316–325, Springer-Verlag, Berlin/New York, 1979; *in* "Path Integrals and their Applications in)+uantum Statistical and Solid State Physics" (G. J. Papadopoulos and J. T. Devreese, Eds.), pp. 163–200, Plenum, New York, 1978.
13. Ref. [1], Appendix 1.
14. M. ABRAMOWITZ AND I. A. STEGUN, "Handbook of Mathematical Functions," U. S. National Bureau of Standards, Washington, D.C., 1964.
15. A. M. OZORIO DE ALMEIDA, *Acta Cryst. A* **31** (1975), 435–442, 442–445.
16. I. S. GRADSHTEYN AND I. M. RYZHIK, "Table of Integrals, Series and Products," Academic Press, New York, 1965.
17. E. TELLER, *J. Phys. Chem.* **41** (1937), 109–116.
18. Ref. [1], Appendix 10.
19. H. C. LONGUET-HIGGINS, *Proc. Roy. Soc. Ser. A* **344** (1975), 147–156.
20. B. F. BUXTON AND M. V. BERRY, *Phil. Trans. Roy. Soc.* **282** (1976), 485–525.
21. W. SIERPINSKI, "Elementary Theory of Numbers," Math. Mon. Pol. Acad. No. 42, Warsaw, 1964.
22. C. F. PORTER (Ed.), "Statistical Theory of Spectra: Fluctuations," Academic Press, New York, 1965.
23. M. V. BERRY AND M. TABOR, *Proc. Roy. Soc. Ser. A* **356** (1977), 375–394.

24. V. L. POKROVSKII, *JETP Lett.* 4 (1966), 96–99.
25. G. M. ZASLAVSKII, *Sov. Phys. JETP* 46 (1977), 1094–1098.
26. R. A. MARCUS, *in* "Proc. Third International Congress of Quantum Chemistry, 1979, Kyoto" (B. Pullman, Ed.), Reidel, Dordrecht.
27. P. LLOYD, *Proc. Phys. Soc.* 90 (1967), 207–216.
28. M. V. BERRY AND K. E. MOUNT, *Rep. Progr. Phys.* 35 (1972), 315–390.
29. H. M. NUSSENZWEIG, *Ann. Phys. (N.Y.)* 34 (1965), 23–95.
30. R. M. LEWIS, N. BLEISTEIN, AND D. LUDWIG, *Comm. Pure Appl. Math.* 20 (1967), 295–328.
31. A. NORCLIFFE AND I. C. PERCIVAL, *J. Phys. B* 1 (1968), 774–783.
32. S. W. MCDONALD AND A. N. KAUFMAN, *Phys. Rev. Lett.* 42 (1979), 1189–1191.
33. H. P. MCKEAN, *Comm. Pure Appl. Math.* 25 (1972), 225–246.
34. B. B. MANDELBROT, "Fractals," Freeman, San Francisco, 1977.
35. M. V. BERRY AND Z. V. LEWIS, *Proc. Roy. Soc. Ser. A* 370 (1980), 459–484.
36. J. H. HANNAY AND M. V. BERRY, *Physica* 1D (1980), 267–290.
37. K. UHLENBECK, *Amer. J. Math.* 98 (1976), 1059–1078.
38. R. M. STRATT, N. C. HANDY, AND W. H. MILLER, *J. Chem. Phys.* 71 (1979), 3311–3322.
39. M. PINSKY, *SIAM J. Math. Anal.* 11 (1980), 819–827.

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium