STRUCTURES IN SEMICLASSICAL SPECTRA: A QUESTION OF SCALE

M.V. Berry

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

ABSTRACT

Theories of semiclassical bound state spectra for systems with N freedoms are reviewed, emphasizing the different features occurring on successively finer scales of energy E, measured in terms of $\hbar$, and attempting to correlate these with whether the underlying classical motion is regular or irregular.

1. INTRODUCTION

Substantial efforts are now being made [1-7] to unravel the semiclassical energy-level structure of bound Hamiltonian systems, motivated for example by the need to understand the high-lying states of vibrating molecules. The difficulty of the quantal problem reflects the complexity of the underlying classical motion [8-11]: for a system with N freedoms (i.e. with a 2N-dimensional phase space), where $N \geqslant 2$, motion may be regular (ordered, predictable), i.e. orbits wind smoothly round N-dimensional tori in phase space; or it may be irregular (chaotic, unpredictable), i.e. exploring 2N-1 dimensional regions in phase space with neighbouring orbits separating exponentially. What corresponds, quantum-mechanically, to these classical distinctions? My intention here is to give a brief discussion of this question, concentrating on the distribution of energy levels; a more substantial review, including an account of the morphologies of wave functions, will be published elsewhere [12].

Obviously one expects the nature of the classical motion to influence the quantal spectrum under semiclassical conditions,

S. Diner et al. (eds.), The Wave-Particle Dualism, 231–252.

i.e. when $\hbar$ is in some sense small (in comparison with classical quantities having the same dimensions). In fact when $\hbar \to 0$ the spectrum displays a richness of different types of behaviour on clearly separated energy scales. I shall describe this picture of the spectrum by starting in § 2 with the coarsest energy scales (larger than $O(\hbar)$), on which the spectrum is described by the mean level density $\bar{d}(E)$. In § 3 I shall explain how oscillatory corrections to $\bar{d}$, corresponding to level clustering on scales of order $\hbar$, are associated with classical closed orbits. § 4 is devoted to the 'quantization conditions' giving individual levels, whose spacing is of order $\hbar^N$. Then, in § 5 the texture of the spectrum on scales smaller than $\hbar^N$ will be discussed in terms of the distributions of spacings between neighbouring levels. Finally, in § 6 the most delicate features of all, on the finest scale-degeneracies between energy levels - will be considered.

It will be helpful to have in mind certain extreme types of classical motion. The extreme of regularity is displayed by integrable systems, where there are N global constants of motion (including the energy) and all orbits are confined to phase-space tori. Systems displaying extreme irregularity are ergodic: only the energy is conserved and almost all orbits (i.e. all except a set of zero measure) explore almost all of the 2N-1 dimensional 'energy surface'. But most Hamiltonians (including the anharmonically coupled oscillators representing vibrating molecules) are neither integrable nor ergodic: a finite volume of phase space is filled with N-tori, and a finite volume with 2N-1-dimensional chaotic orbits; often the proportions change as the energy increases, being nearly integrable ('quasi-integrable') at low energies and nearly ergodic at high energies.

2. SCALES LARGER THAN $\hbar$: THE MEAN LEVEL DENSITY

Consider a bound system with coordinates $\underset{\sim}{q} = \{q_1 \ldots q_N\}$ and momenta $\underset{\sim}{p} = \{p_1 \ldots p_N\}$, with classical Hamiltonian $H(\underset{\sim}{q},\underset{\sim}{p})$. Suppose the corresponding Hamiltonian $\hat{H} = H(\hat{\underset{\sim}{q}},\hat{\underset{\sim}{p}})$ generates the discrete spectrum of energies $E = E_1, E_2 \ldots = \{E_j\}$, labelled in order of increasing E. A complete description of this spectrum would be obtained from knowing the level density d(E), defined as

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

There is a beautiful semiclassical theory for d(E), whose principal architects were Gutzwiller [13-17] and Balian and Bloch [18-21]. This is based on representing d(E) in the form

$$d(E) = \bar{d}(E) + d_{osc}(E), \qquad (2)$$

where $\bar{d}$ is the mean level density and d_{osc} is a series of oscillatory corrections. The terms on the right side of (2) correspond to successive smoothings of the singular function $d(E)$. On the coarsest scale, that is after smoothing over energy ranges ΔE large enough to obliterate all traces of individual levels and all scales of level clustering, only $\bar{d}(E)$ survives. As ΔE is made smaller, more and more terms in $d_{osc}(E)$ contribute, with faster oscillations, until eventually they sum to give a series of delta functions at the energies of the levels. In simple terms, the representation (2) is a generalization of the following 'Poisson' expansion for a series of equally spaced delta functions:

$$\sum_{n=-\infty}^{\infty} \delta(E-n) = 1 + 2 \sum_{n=1}^{\infty} \cos 2\pi mE \qquad (3)$$

We begin by studying the mean level density $\bar{d}(E)$. This is given by the simple semiclassical rule [e.g. 22] that each quantum state is associated with a phase-space volume h^N, so that it predicts

$$\bar{d}(E) \approx \frac{1}{h^N} \int d\underset{\sim}{q} \int d\underset{\sim}{p}\, \delta(E-H(\underset{\sim}{q},\underset{\sim}{p})), \qquad (4)$$

i.e. the level density is proportional to the 'size' of the energy surface.

One way to obtain this formula - which is convenient in that it also leads to a theory for d_{osc} - is to write (2) in terms of the outgoing time-independent Green function:

$$G^+(\underset{\sim}{q},\underset{\sim}{q}';E) \equiv \left\langle q' \left| \frac{1}{E+i\varepsilon-\hat{H}} \right| q \right\rangle \qquad (5)$$

so that

$$d(E) = -\frac{1}{\pi} \operatorname{Im} \int d\underset{\sim}{q}\, G^+(\underset{\sim}{q},\underset{\sim}{q}';E)_{\underset{\sim}{q}' \to \underset{\sim}{q}} \qquad (6)$$

In view of the fact that G^+ is the probability amplitude for observing at $\underset{\sim}{q}$ particles emitted from $\underset{\sim}{q}'$ with energy E, it is not surprising that the semiclassical theory [13-21; see also the review 23] gives G^+ as a sum over all classical trajectories leading from $\underset{\sim}{q}'$ to $\underset{\sim}{q}$ with energy E. As $\underset{\sim}{q}' \to \underset{\sim}{q}$ this sum is dominated by the direct path going straight from $\underset{\sim}{q}'$ to $\underset{\sim}{q}$, and the contribution to (6) from this path alone gives precisely $\bar{d}$ as given by (4).

In the important special case where H takes the form

$$H(\underset{\sim}{q},\underset{\sim}{p}) = \frac{|\underset{\sim}{p}|^2}{2\mu} + V(\underset{\sim}{q}), \qquad (7)$$

(4) gives $\bar{d}$ as an integral over the classically accessible space, namely

$$\bar{d}(E) \approx \left[\frac{\mu}{2\pi\hbar^2}\right]^{\frac{N}{2}} \frac{1}{\Gamma(N/2)} \int d\underset{\sim}{q} \left[E-V(\underset{\sim}{q})\right]^{\frac{N}{2}-1} \Theta\left[E-V(\underset{\sim}{q})\right], \qquad (8)$$

where Θ denotes the unit step function. If N=2 this reduces to the simple formula

$$\bar{d}(E) = \frac{\mathcal{A}(E)\mu}{2\pi\hbar^2} \qquad (9)$$

where $\mathcal{A}(E)$ is the area of the classically accessible $\underset{\sim}{q}$-space. In the context of 'billiard' systems, where $V(\underset{\sim}{q})$ vanishes within a curve B and is infinite outside B, so that classical trajectories are straight lines reflecting specularly at B and quantal wave functions vanish on B, (9) is called the Weyl formula (for a review see [24]).

For numerical tests of (4) it is convenient to work not with the spiked function d(E) but with its integral, namely the mode number $\mathcal{N}(E)$ defined by

$$\mathcal{N}(E) \equiv \int_{-\infty}^{E} d(E')dE' = \sum_{j=1}^{\infty} \Theta(E-E_j). \qquad (10)$$

The procedure is to compute the levels E_j, construct the exact stepped curve $\mathcal{N}(E)$ and compare its trend $\bar{\mathcal{N}}(E)$ with what (4) would give. In one such test, Marcus and his coworkers [25,52] obtain very good agreement. Here I show another test, for the 'desymmetrized Sinai billiard', whose boundary B is shown on fig. 1. This system (which I studied in detail in [26]) was chosen because its classical trajectories were proved to be ergodic (Sinai [27]) whenever R > 0 (when R = 0 the system is trivially integrable). Fig. 2 shows $\mathcal{N}(E)$ (stepped curve) and $\bar{\mathcal{N}}(E)$ as given by (9) and (10) (full curve) for five values of R. Evidently the agreement is poor. The reason is that (4) is an asymptotic formula which is here being tested on low-lying states. It is necessary to include correction terms. These are not contributions to d_{osc} (which will be discussed in § 3), but are smooth terms of

lower order in $\hbar$, depending on aspects of billiard geometry [24, 18] other than the area $\mathcal{A}$. When these corrections are included, the agreement improves spectacularly, as the dashed curves in fig. 2 show.

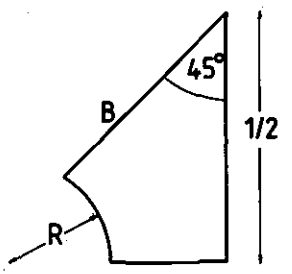

Figure 1. Desymmetrized Sinai billiard.

We draw two conclusions from the formula (4) for $\bar{d}$. The first is that the mean spacing of the levels, $\bar{d}^{-1}$, is proportional to $\hbar^N$. The second is that on these coarsest scales, embodied in $\bar{d}$, the spectrum is so smoothed as to no longer display any evidence of the regularity or irregularity of the underlying classical motion. To find such evidence, we must take an 'energy microscope' and inspect the spectrum more closely.

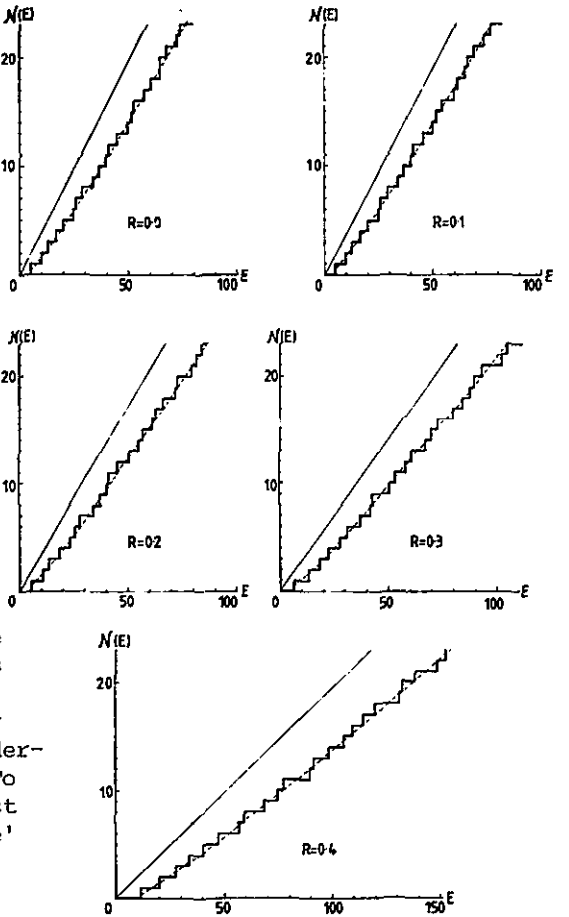

Figure 2. Mode number (stepped curve) for Sinai billiard, with uncorrected mean mode number (full curve) and corrected mean mode number (dashed curve).

3. SCALES OF ORDER $\hbar$: OSCILLATORY LEVEL CLUSTERING

The fact that the mean level spacing is of order $\hbar^N$ might

lead us to expect that the mean level density $\bar{d}(E)$ would be obtained simply by smoothing the delta-spikes of the exact $d(E)$ over an energy range of order $\hbar^N$. But such an expectation is mistaken, because as we shall see the corrections d_{osc} oscillate with energy 'wavelength' of order $\hbar$, which when $N > 1$ (i.e. in nontrivial cases) is infinitely larger than the mean spacing as $\hbar \to 0$. In this important respect the spectrum in the general case contrasts with those special cases (which actually correspond to N=1 as analyzed for example in [23]) represented by (3), in which the slowest-oscillating corrections have energy wavelength equal to the mean spacing.

The clearest route to understanding the nature of the contributions to d_{osc} is the Green function method based on (5) and (6). As already stated, $G^+(\underset{\sim}{q},\underset{\sim}{q}';E)$ is a sum over classical paths linking $\underset{\sim}{q}'$ and $\underset{\sim}{q}$. As $\underset{\sim}{q}' \to \underset{\sim}{q}$ these include not only the direct path giving $\bar{d}$, but paths looping back to $\underset{\sim}{q}$ after a finite excursion. A looping path need not be a closed orbit, because it may (and usually does) return to $\underset{\sim}{q}$ with momentum $\underset{\sim}{p}$ different from its initial momentum $\underset{\sim}{p}'$. But in the integration over $\underset{\sim}{q}$ in (6) such non-closed looping paths can be shown [16, 12, 23] to give negligible contributions.

The conclusion is that only closed classical orbits with energy E contribute to $d_{osc}(E)$; these include repetitions (labelled by p) of primitive (i.e. unrepeated) closed orbits (labelled by j). The resulting formula for d_{osc} will be written down first and then explained:

$$d_{osc}(E) = \sum_j \sum_{p=1}^{\infty} \frac{A_{j,p}(E)}{\hbar^{1+\ell_j/2}} \sin\left\{ \frac{pS_j(E)}{\hbar} + p\alpha_j \right\} \qquad (11)$$

In the phase, $S_j(E)$ is the action

$$S_j(E) = \oint \underset{\sim}{p}_j(\underset{\sim}{q};E).d\underset{\sim}{q} \qquad (12)$$

around the j'th path, and α_j depends on the focusing of trajectories near the j'th closed orbit.

As E varies, the action round the j'th path changes, and causes its contribution to d_{osc} to oscillate with energy wavelength ΔE given by

$$\frac{p}{\hbar}\frac{dS_j}{dE}\Delta E = 2\pi \quad \text{i.e.} \quad \Delta E = \frac{h}{pT_j} \qquad (13)$$

where T_j is the period of the j'th primitive orbit. Therefore the j^{th} oscillations are indeed on the scale $\hbar$ (much longer than $\hbar^N$ as $\hbar \to 0$) as asserted. Longer paths give faster

oscillations.

The amplitude in (11) depends on whether the j'th path is isolated or nonisolated, and whether it is stable or unstable. In (11) the integer ℓ_j measures the degree of isolation: the j'th orbit is embedded in an ℓ_j-parameter family of closed orbits. ℓ_j may vary from zero (for an isolated orbit) to N-1 (for the torus-filling closed orbits of an integrable system - taking account of the fact that each closed orbit occupies one dimension). As an example, the stadium billiard (fig. 3) which Bunimovich [28] proved to be ergodic, has both isolated closed orbits (e.g. the 'long-diameter' and 'rectangular' orbits in fig. 3) and non-isolated closed orbits (the 'short-diameter' orbits in fig. 3).

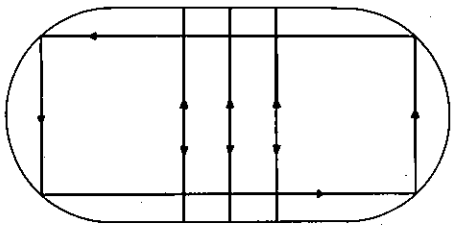

Figure 3. Stadium billiard, with two isolated closed orbits and three members of a family of nonisolated closed orbits.

In (11) the behaviour of the factor $A_{j,p}$ as a function of repetitions p depends on the stability of the j'th orbit. For isolated orbits, $A_{j,p}$ oscillates with p if the orbit is stable, and decays exponentially if the orbit is unstable - as expected in view of the repeated focusing or continued defocusing of beams of trajectories in the two cases. For integrable systems, where orbits are embedded in N-1 parameter families, Berry and Tabor [29, 30] showed that $A_{j,p}$ decreases as $p^{-(N-1)/2}$.

The Green function theory leading to (11) is reviewed by Rajaraman [31] and deWitt-Morette et al. [32]. In the original papers, Gutzwiller [16] emphasizes the question of stability, and Balian and Bloch [20] emphasize the question of isolation.

Now I want to dispose of a fallacy based on a misinterpretation of (11). Consider the terms with given j, i.e. those corresponding to all repetitions of a single closed orbit. The terms will interfere constructively if

$$S_j(E_m) = (2\pi m - \alpha_j)\hbar \qquad (14)$$

defining a series of energies E_m corresponding to integers m, at which the sum over p gives a contribution to d_{osc} with some sort

of singularity whose nature depends on $A_{j,p}$. The fallacy is to suppose that E_m must be eigenvalues of $\hat{H}$, i.e. that (14) is a semiclassical quantization condition associating individual quantum states with repetitions of individual closed orbits.

Why is this a fallacy? For a start, (14) gives energies with separations of order $\hbar$, whereas the true levels have separation $\hbar^N$. So there are too few levels. But why not superpose the level sequencies obtained from (14) with all topologically different orbits? Because this would give too many levels! An instructive example is motion in a rectangle, where the true levels are labelled by two quantum numbers (see § 4) and the false levels given by (14) are labelled by three quantum numbers and moreover have the 'infrared catastrophe' of existing with arbitrarily low energies.

Nevertheless, there are two circumstances where (14) does give semiclassical quantal levels correctly. The first (trivial) case is potential wells with N=1, where there is only one topology of closed orbit and the levels do have separation $\hbar$ (see e.g. [23]). The second case is when the orbits are isolated and stable. Then Miller [33] showed that by considering lowest-order fluctuations about the periodic orbit it was possible to generalize (14) into a condition with a full set of quantum numbers. But Voros [34] pointed out that isolated stable orbits are always surrounded by tori, and explained how this quantum condition is really an approximate version of the 'torus quantization' to be discussed in § 4.

In general, though, a single closed orbit gives not individual levels but a collective property of the spectrum, namely an oscillatory clustering with scale ΔE given by (13). Conversely, the determination of individual levels from (11) involves the close orbits collectively, and would require the summation over sufficiently many closed orbits for individual delta functions in $d(E)$ to emerge as the result of interference. Is this a feasible procedure for calculating individual levels? I shall argue that it is not.

To begin to see delta functions emerging from (11), it is necessary to include at least all orbits giving oscillations whose energy wavelength (13) exceeds the mean spacing $\bar{d}^{-1}$. Since the longer orbits give faster oscillations, it is necessary to sum over at least all closed orbits with period less than T_{max}, which is easily calculated to be

$$T_{max} = \frac{\int d\underset{\sim}{q} \int d\underset{\sim}{p} \ \delta(E-H(\underset{\sim}{q},\underset{\sim}{p}))}{h^{N-1}} \qquad (15)$$

As $\hbar \to 0$, T_{max} increases and it is necessary to include ever more closed orbits. How many? As discussed for example in [12] and [26], the number is of order $\hbar^{-N(N-1)}$ for

integrable systems, and of order exp{+const/ħ} for chaotic systems where closed orbits are vastly more numerous. This means that the number of operations required to determine levels by summing over closed orbits is much greater than in competing methods (see § 4).

Nevertheless, it is instructive to see delta functions emerging as more topologies are included, and I illustrate this with an integrable system with N=2, whose Hamiltonian involves a Morse potential:

$$\left.\begin{aligned} H &= \frac{p_x^2+p_y^2}{2\mu} + V_o\left[e^{-2\delta(r-r_o)} - 2e^{-\delta(r-r_o)}\right] \\ &(\mu=1 \text{ proton mass}, \; V_o = 0.2 \text{ eV}, \; r_o = 0.25 \text{ nm}, \; \delta = 10 \text{ nm}^{-1}); \\ & r \equiv (x^2+y^2)^{1/2} \end{aligned}\right\} \tag{16}$$

there are 166 bound levels. Fig. 4 shows some simple topologies of closed orbit (classified by the number of rotations and librations before closure), and fig. 5 shows the effect on d(E) of including increasingly many topologies. In the last frame, which includes several hundred closed orbits, delta functions are beginning to emerge very clearly (arrows mark exact levels, the chain curve is $\bar{d}(E)$). For more details see [29].

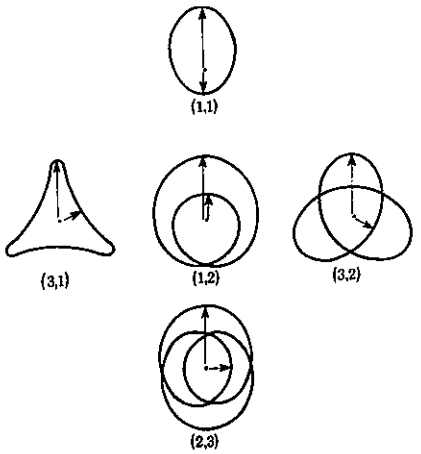

Figure 4. Some closed orbits in a central potential.

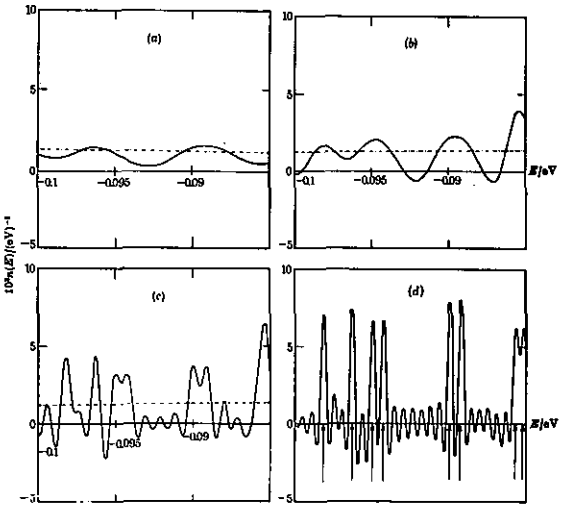

Figure 5. Emergence of δ functions in $d(E)$ as more oscillatory corrections, from closed orbits, are added to $\bar{d}(E)$.

For ergodic systems, direct summation over closed orbits appears hopeless as a means of locating eigenvalues. However, in a tour de force, Gutzwiller [35] has recently evaluated by an indirect (and not very economic) method the path sum for the (almost certainly ergodic) Kepler problem with anisotropic kinetic energy. He obtains the first 18 levels with an accuracy of a few percent.

It is therefore the case that, rather than being a means of calculating individual levels, the sum over closed orbits gives information about clustering on scales of order $\hbar$. According to (11), this clustering is much stronger for integrable systems (orbits nonisolated and neutrally stable) than for ergodic systems (most orbits isolated, and all unstable). This is the first evidence that systems whose classical motion is irregular have levels distributed more regularly than systems with regular classical motion; in § 5 we shall see that the same property holds down to the finest scales.

And finally, although I have presented the closed orbit expansions as semiclassical approximations, and this is how they arose in the Green function theories [13-20], it is encouraging that an exact correction is known between the spectrum and the closed orbits for nonintegrable systems. For the wave equation on a smooth compact Riemannian manifold (which need not have constant curvature), if the eigenwavenumbers are

$k_j (=2\mu E_j)^{1/2}/\hbar)$, then the function

$$Q(\lambda) = \sum_{j=1}^{\infty} e^{ik_j\lambda} \qquad (17)$$

was proved by Chazarain [36] to have singularities for each λ equal to the length of a closed geodesic ($Q(\lambda)$ is the Fourier transform of the k density of states). Balian and Bloch [21] obtained a similar exact result for the Schrödinger equation: the transform with respect to $\hbar^{-1}$ is singular at the action values of closed classical orbits.

4. SCALES OF ORDER $\hbar^N$: QUANTIZATION CONDITIONS

Now we step down from the scale $\hbar$ to the scale $\hbar^N$, and seek a semiclassical quantization formula giving the positions of individual energy levels with an accuracy that increases as $\hbar \to 0$. Such a formula exists for regular motion, that is for orbits lying on phase-space tori; it is known by various names: Bohr, Sommerfeld, Einstein [37], Brillouin, Keller [38], Maslov [39] ... I shall refer to the procedure, now to be described, simply as 'torus quantization'.

Let tori be labelled by the values I_i of the actions round their i'th irreducible circuits γ_i:

$$I_i \equiv \frac{1}{2\pi} \oint_{\gamma_i} \underset{\sim}{p}(\underset{\sim}{q}).d\underset{\sim}{q} \qquad (1 \leqslant i \leqslant N), \qquad (18)$$

Then the eigenstate with quantum numbers $\underset{\sim}{m} \equiv \{m_1 \ldots m_N\}$ is associated with the torus whose actions $\underset{\sim}{I} \equiv \{I_i \ldots I_N\}$ are quantized by

$$\underset{\sim}{I}_{\underset{\sim}{m}} = (\underset{\sim}{m} + \underset{\sim}{\alpha}/4)\hbar \qquad (19)$$

where $\underset{\sim}{\alpha} \equiv \{\alpha_1 \ldots \alpha_N\}$ is a set of N integers giving the number of real-space caustics encountered during each cycle γ_i of the torus. The energy levels $E_{\underset{\sim}{m}}$ are simply the energies of these quantized tori, i.e.

$$E_{\underset{\sim}{m}} = H(\underset{\sim}{I}_{\underset{\sim}{m}}) \qquad (20)$$

where H denotes the Hamiltonian in action representation (for elementary derivations of this torus quantization formula see [40] and [12]).

The tori picked out by (19), associated with individual

quantum states, almost never have commensurable frequencies and so do not contain closed orbits. On the other hand, we saw in § 3 that closed orbits can be made to generate the whole spectrum. This is apparently contradictory but actually no more than the fact that an irrational number can be approximated by a series of rationals.

The explicit formula (20) is exact as $\hbar \to 0$. For fixed E this corresponds to high states, but (20) often gives results accurate to a few percent even for the ground state (e.g. when N=1, in which case it is the familiar WKB level formula [23]); in a few cases (harmonic oscillator, Coulomb potential, rectangular boxes) it gives all levels exactly. For integrable systems, the whole phase space is filled with tori and (20) approximates all the levels. For quasi-integrable systems, the Kolmogorov-Arnol'd-Moser (KAM) theorem [8-11] guarantees the existence of tori filling part of phase space, and so (20) can be employed to find a finite proportion of the levels, provided the actions and energies of the tori can be determined. This is a difficult problem of classical mechanics, which has been tackled analytically by perturbation and iteration methods, and numerically by studying caustics and the Poincare surface of section. By these techniques, Marcus and his colleagues [1-4], Percival and Pomphrey [5], Jaffe and Reinhardt [6] and Chapman, Garrett and Miller [7] obtained eigenvalues for chemically interesting quasi-integrable systems in very good agreement with 'exact' computations.

What determines the levels if large regions of phase space are filled with irregular trajectories? The extreme is an ergodic system. Then torus quantization must fail, for there are no tori, no actions and so no quantum numbers. Nobody has so far succeeded in finding an explicit quantum condition to put in its place, i.e. to give the levels in what Percival [41] called the 'irregular spectrum'. For billiard systems with N=2, however, and in particular for the ergodic Sinai billiard (fig. 1), it is possible to devise an <u>implicit</u> quantum condition [26] in the form of a determinant, obtained by a Green function procedure, which is much more rapidly convergent than those obtained from conventional basis-set expansions. The determinant has effective size MxM, where

$$M = \frac{\text{perimeter of billiard boundary}}{\text{de Broglie wavelength } \lambda \text{ of state being studied}} \qquad (21)$$

Because λ is proportional to $\hbar$, the work of finding levels increases as $\hbar \to 0$, but less rapidly than in any other method known to me.

It is possible by (nontrivial) analytical manipulation of the explicit torus quantization condition (20), or the implicit determinantal condition for billiard eigenvalues, to give

alternative derivations [29,26] of the 'large-scale' formulae for $\bar{d}$ and d_{osc} already discussed in §§ 2 and 3.

5. SCALES SMALLER THAN $\hbar^N$: LEVEL SPACINGS

Now that we are down on the scale $\hbar^N$ corresponding to individual levels, the next step is a further application of the energy microscope, to study the fine-scale texture of the spectrum as embodied in the distribution P(S) of spacings S between neighbouring levels. S is measured as a fraction of the mean level spacing $[\bar{d}(E)]^{-1}$, and P(S) is defined by

$$P(S)dS = \text{probability that the spacing of a randomly chosen pair of neighbouring levels lies between S and S+dS.} \qquad (22)$$

The ensemble over which the probability is taken is defined semiclassically, as the infinitely many levels near any given E as $\hbar \to 0$. The function P(S) was introduced [42] in a non-semiclassical context, to describe the many-body spectrum of nuclei.

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

In § 6 I shall outline an argument strongly suggesting that for generic (i.e. 'almost all') systems,

$$P(S) \sim \text{const} \times S \text{ as } S \to 0 \qquad (23)$$

(the same result holds for eigenvalues of the ensembles of random matrices [42] considered in nuclear physics). Therefore generic systems are expected to display linear level repulsion. Fig. 6 shows a test of this prediction, obtained by calculating several hundred quantal eigenvalues [26] of the classically ergodic desymmetrized Sinai billiard (fig. 1), for a range of R between .20 and .44. It is clear that the levels do repel, and that the linear law gives a good fit. McDonald and Kaufman [43], and Casati et al. [44], in computations of P(S) for the desymmetrized stadium billiard (one quarter of the shape in fig. 3), also obtain level repulsion.

On the other hand, for systems with torus quantization Berry and Tabor [45] showed from (20) that in almost all cases (and if $N > 1$), P(S) has the universal form

$$P(S) = e^{-S} \qquad (24)$$

which is finite as $S \to 0$, in contrast to (23), and corresponds

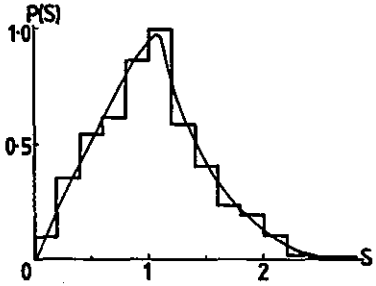

Figure 6. Level spacings distribution for desymmetrized Sinai billiard.

to levels arriving irregularly with a degree of clustering. Fig. 7 shows tests of this prediction for two integrable systems with N=2: (a) corresponds to a rectangle with side ratio 2, and (b) corresponds to a potential which is a square well in the x direction and a harmonic oscillator in the y direction. Evidently the negative exponential is a very good fit to the computed histograms.

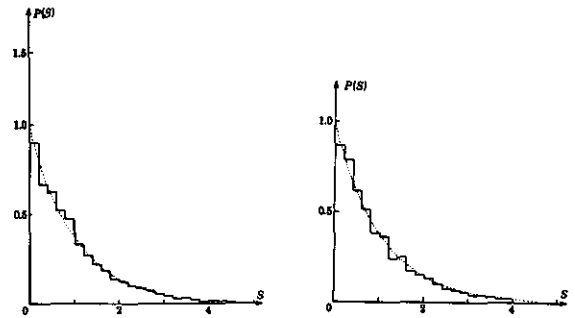

Figure 7. Level spacings for (a) rectangle with side ratio 1/2, and (b) square well along x and harmonic oscillator along y.

It is tempting on the basis of (23) and (24) to use P(S) to discriminate between a spectrum associated mainly with irregular motion and a spectrum associated mainly with regular motion. Indeed the implication

classically irregular motion in Hamiltonian without symmetry → level repulsion given by (23) (25)

is almost surely correct. But the reverse implication is wrong: the existence of level repulsion does not indicate classically irregular motion.

To see this, consider the delicate case of the 'desymmetrized square torus billiard' (fig. 8), which is one of a class of two-dimensional systems that Richens and Berry [46] showed to possess the 'pseudointegrable' property of having two constants of motion confining orbits not to tori but to phase-space surfaces with the topology of multiply-handled spheres (two handles in the case of fig. 8, as opposed to one for tori). If there is any chaos here, it is of a very rudimentary nature - not 2N-1 dimension filling, and with no exponential orbit separation. Nevertheless, P(S) shows clear level repulsion, as fig. 9 indicates.

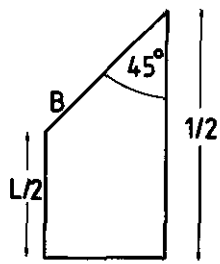

Figure 8. Desymmetrized square torus billiard.

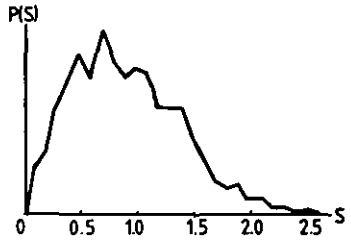

Figure 9. Level spacings distribution for desymmetrized square torus billiard.

Moreover, even for systems with tori the quantization condition (20) forming the basis of the level clustering law (24) is approximate rather than exact. For most systems with tori, and indeed for most integrable systems (where phase space is entirely filled by tori) I expect ([12] and § 6) that for very small values of S, of order $\exp\{-\text{const}/\hbar\}$, P(S) will be given not by (24) but will fall to zero like (23). So when does clustering occur in the exact (rather than semiclassical) spectrum? For the very special case where Schrödinger's equation is separable in the coordinates (as opposed to more general phase-space separability under the canonical trans-

formation to action-angle variables, whose existence is guaranteed by integrability).

It should also be pointed out that there are certain very special classes of system (all integrable) for which P(S) does not exist. For example, consider two equal harmonic oscillators with Hamiltonian

$$H = \frac{p_x^2 + p_y^2}{2\mu} + \tfrac{1}{2}\mu\omega^2\left(q_1^2 + q_2^2\right) \qquad (26)$$

The exact levels are

$$E_{m_1 m_2} = \hbar\omega(m_1+m_2+1) \qquad (0 \leqslant m_1, m_2 < \infty) \qquad (27)$$

and occur at energies

$$E = \hbar\omega p \qquad (1 \leqslant p < \infty) \qquad (28)$$

with degeneracies p. The mean level density is, from (4) (or directly)

$$\bar{d}(E) = \frac{E}{\hbar^2\omega^2} \qquad (29)$$

so that the mean spacing of states is of order $\hbar^2$ as expected for N=2. But the actual levels (28) have spacings $\hbar$ and arrive in ever-more-degenerate groups, so that P(S) does not exist - the clustering structure continually changes as $\hbar \to 0$.

6. THE FINEST SCALE: DEGENERACIES

The most delicate question we can ask concerning neighbouring levels is: under what circumstances do these coincide? In other words: when do degeneracies occur? We shall see that the answers to these questions will enable us to understand the spacings distributions P(S) just discussed.

If the Hamiltonian $\hat{H}$ has any symmetry, this may produce degeneracies, whose nature can be studied using group theory. I shall not consider degeneracies of this type, and so when $\hat{H}$ does have symmetry I shall consider only states which all have the same symmetry class. This procedure is equivalent to considering all the states in a suitably 'desymmetrized' Hamiltonian, and that is why I have used this term in connection with the billiards in figs. 1 and 8.

For a typical (generic) such Hamiltonian it seems clear that

degeneracy is infinitely improbable. But it might be expected that in a one-parameter family of Hamiltonians,

$$\hat{H} = H(\hat{q}, \hat{p}; A) \qquad (30)$$

the eigenvalue curves E(A) could cross, and so two levels could degenerate, for isolated values of the parameter A. But the surprising fact is that this picture is not correct: for typical systems with real eigenfunctions (the only ones considered here) it is necessary to vary two parameters, not one, in order to make two levels degenerate. This is the content of a theorem due originally to von Neumann and Wigner [47] and Teller [48] and later generalized by Arnol'd (ref. [9], appendix 10). The proof is based on a simple argument [12] employing degenerate perturbation theory.

Let the two parameters be A and B. Then the theorem just mentioned implies that the connection of eigenvalue surfaces $E = E_{\pm}(A,B)$ in E, A, B space takes the form of a double cone (diabolo) (fig. 10) with sheets joined at the 'diabolical point' E^*, A^*, B^*, where A^*, B^* are the parameters for which the degeneracy occurs. If only one parameter A is varied, the diabolical point will almost surely be missed and the curves $E_{\pm}(A)$ will avoid each other like branches of a hyperbola obtained by slicing the cone, rather than crossing.

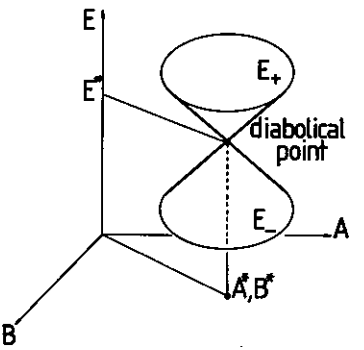

Figure 10. Diabolical degeneracy structure in space of energy E and parameters A, B.

In testing the resulting picture of the spectrum, it is plausible to assume that a one-parameter family of non-symmetric classically ergodic systems will be typical in the quantal sense, and so will produce no level crossings. Such a family is the desymmetrized Sinai billiard (fig. 1) with the radius R acting as parameter. The levels $E_j(R)$ were computed 'exactly' [26], and the resulting spectrum is shown in fig. 11. There are many near-degeneracies, but close examination shows that no two levels

actually cross, so this test is successful.

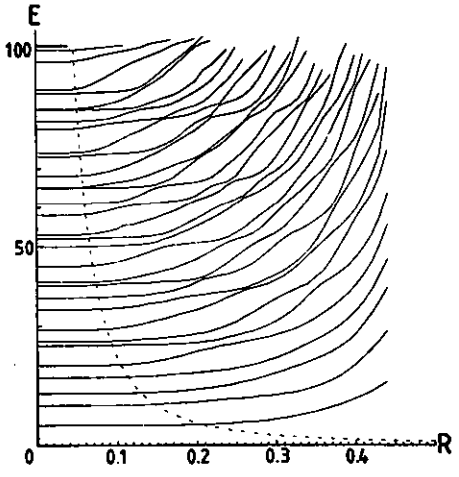

Figure 11. Spectrum of an ergodic system: levels E of desymmetrized Sinai billiard as functions of R.

The two-parameter rule does not hold for systems whose levels are given by the torus quantization rule (20). In these cases level crossings occur when only one parameter A is varied, thus reinstating the intuitive picture which as we have seen is wrong for the general case. To see why this is so, regard (20) as defining, for a given parameter A, a set of N-1 dimensional hypersurfaces $E=H(\underset{\sim}{I}_{\underset{\sim}{m}};A)$ in the N-dimensional space of quantum numbers $\underset{\sim}{m}$. The levels occur at energies whose hypersurfaces intersect a lattice point. Consider the lattice point $\underset{\sim}{m}^*$. Typically, the hypersurface through $\underset{\sim}{m}^*$ will not contain any other lattice point, and so the state $\underset{\sim}{m}^*$ will be nondegenerate. But on varying A the hypersurface will smoothly change its orientation and there will typically be values A^* where it cuts another lattice point, so that A^* corresponds to a pair of degenerate states. A simple example of this behaviour is the (integrable) family of rectangular boxes with sides 1, A, whose eigenvalues are

$$E_{m,n} = \frac{\pi^2\hbar^2}{2\mu}(m^2+A^2n^2) \qquad (31)$$

Degeneracies occur whenever A^2 is rational; for example, the states (2,1) and (1,3) degenerate when $A^2 = 3/8$.

One-parameter crossings will occur in the exact spectrum in the rare cases when torus quantization is exact. They will occur in separable systems for which torus quantization is not exact, because the basis for the conclusion still holds, namely that the energy can be written as a continuous function in the space of quantum numbers. But for nonseparable systems with tori I expect (20) to be modified by some multidimensional analogue of barrier penetration, causing the one-parameter degeneracies it predicts to be split by energies of order $\exp\{-1/\hbar^2\}$, so that the exact levels will show the typical property of two-parameter crossings.

There are some special integrable systems whose levels are all proportional to integers, and which are therefore not naturally embedded in continuous families, whose degeneracy structure is very strange. We have already mentioned the two equal harmonic oscillators (26), whose spectrum (28) shows increasing degeneracies. Another example is the 45° right triangle (desymmetrized square), whose levels are proportional to m^2+n^2. It can be shown [26] that as $\hbar \to 0$ the levels are increasingly multiply degenerate, with multiplicity proportional to $\ell n(\hbar^{-1})$, and are separated by energies of the order $\hbar^2 \ell n(\hbar^{-1})$ rather than $\hbar^2$.

Marcus and his coworkers [25, 52] suggest that the presence of many near-degeneracies (overlapping avoided crossings) in curves of energy levels as functions of a single parameter, is an indication that the corresponding quantum states are associated with classically chaotic motion. This is unlikely to be correct, because two-parameter degeneration is the generic case, i.e. exceptions such as classes of systems with one-parameter degeneracies (or some other number not equal to two) occur with measure zero and so are fragile (like nodal-line intersections of wave functions when N=2 [12, 49, 50]). On the other hand, regular motion such as occurs on tori is robust as guaranteed by the KAM theorem and does not occur with measure zero. Therefore most torus-associated spectra should show two-parameter degeneration, and hence avoided crossings, albeit on very fine scales; only separable systems should show true one-parameter crossings. Support for this idea that avoided crossings need not imply classical chaos is given by computations of Richens and Berry [46] on the 'pseudointegrable' billiard of fig. 8, where orbits are confined to a two-handled (two-dimensional) sphere in phase space: energy levels as a function of L show avoided crossings very similar to those in fig. 11. The further suggestion [25, 52] that the reverse implication holds, i.e. that the energy levels of quantum states associated with classical chaos will show many avoided crossings, is almost certainly correct.

To conclude, I shall now as promised outline the connection between degeneracy structure and the level spacings distribution P(S); for details and a more formal argument, see [12] and [26].

Consider first the case of a generic Hamiltonian, which can be embedded in a family with two parameters A, B and having degeneracies at diabolical points in E, A, B space as in fig. 10. Let the actual Hamiltonian under study have parameters A_o, B_o. Then the line in E, A, B space with $A = A_o$, $B = B_o$ will thread its way among cones, which if $\hbar$ is small will presumably be thickly distributed (according to an unknown, and in this context unimportant, law).

Now, P(S) is defined according to (22) as an average over spacings for fixed A, B, i.e. as an 'energy average'. But by hypothesis there is nothing special about the parameters A_o, B_o and so the energy average can be augmented by an ensemble average over a region of A, B near A_o, B_o. Thus can P(S) be expressed as an integral over a volume V in E, A, B space. For small S, the only contributions come from the neighbourhoods of diabolical points in V, and are proportional to the measure of the intersection of a diabolo by two parallel sheets separated by S. But this measure (length of intersection) is always proportional to S, whatever the orientation, ellipticity or cone angle of the diabolo, and this implies directly the linear level repulsion (23).

A similar argument for nongeneric systems which require m parameters to produce a degeneracy leads to

$$P(S) \sim \text{const} \times S^{m-1} \quad \text{as } S \to 0 \qquad (32)$$

We have seen that m=1 for systems with torus quantization (as opposed to 2 generically), and this leads to the 'clustering' prediction $P \to \text{const}$ as $S \to 0$, consistent with (24). For systems where degeneracies are strictly forbidden, such as finite one-dimensional systems, and two-dimensional harmonic oscillators with irrationally related frequencies, we can take $m = \infty$ and predict, in accord with observation [51, 12, 45], that P(S) vanishes faster than any power of S.

ACKNOWLEDGEMENTS

I thank the University of Utrecht for hospitality whilst this paper was written. The work was not supported by any military agency.

REFERENCES

[1] Noid, D.W., and Marcus, R.A.: 1977, J. Chem. Phys. 67, pp. 559-567.

[2] Eastes, W., and Marcus, R.A. : 1974, J. Chem. Phys. 61, pp. 4301-4306.

[3] Noid, D.W., and Marcus, R.A.: 1975, J. Chem. Phys. 62, pp. 2119-2124.
[4] Noid, D.W., Koszykowski and Marcus, R.A.: 1979, J. Chem. Phys. 71, pp. 2864-2873.
[5] Percival, I.C., and Pomphrey, N.: 1976, Mol. Phys. 31, pp. 97-114.
[6] Jaffe, C., and Reinhardt, W.P.: 1979, J. Chem. Phys. 71, pp. 1862-1869.
[7] Chapman, S., Garrett, B., and Miller, W.H.: 1976, J. Chem. Phys. 64, pp. 502-509.
[8] Ford, J.: 1975,in 'Fundamental Problems in Statistical Mechanics' (ed. E.G.D. Cohen) Vol. 111, pp. 215-255. North-Holland.
[9] Arnol'd, V.I., 'Mathematical Methods of Classical Dynamics', (Springer, New York 1978).
[10] Berry, M.V.: 1978, in 'Topics in Nonlinear Dynamics' (ed. S. Jorna), Am. Inst. Phys. Conf. Proc. 46, pp. 16-120.
[11] Helleman, R.: 1980, in 'Fundamental Problems in Statistical Mechanics (ed. E.G.D. Cohen), Vol. V, pp. 165-233, North-Holland.
[12] Berry, M.V.: 1981, 'Semiclassical Mechanics of Regular and Irregular Motion', Les Houches Summer School Lectures, to be published by North-Holland).
[13] Gutzwiller, M.C.: 1967, J. Math. Phys. 8, pp. 1979-2000.
[14] Gutzwiller, M.C.: 1969, J. Math. Phys. 10, pp. 1004-1020.
[15] Gutzwiller, M.C.: 1970, J. Math. Phys. 11, pp. 1791-1806.
[16] Gutzwiller, M.C.: 1971, J. Math.Phys., 12, pp. 343-358.
[17] Gutzwiller, M.C.: 1978, in 'Path Integrals and their Application in Quantum, Statistical and Solid State Physics' (eds.: G.J. Papadopoulos and J.T. Devreese), pp. 163-200, Plenum, N.Y.
[18] Balian, R., and Bloch, C.: 1970, Ann. Phys. N.Y. 60, pp. 401-447.
[19] Balian, R., and Bloch, C: 1971, Ann. Phys. N.Y. 64, pp. 271-307.
[20] Balian,R., and Bloch,C.:1972, Ann.Phys.N.Y.69,pp. 76-160.
[21] Balian, R., and Bloch, C: 1974, Ann. Phys. N.Y. 85, pp. 514-545.
[22] Landau, L.D., and Lifshitz, E.M.: 1965, in 'Quantum Mechanics (Nonrelativistic Theory)' (Oxford: Pergamon)
[23] Berry, M.V., and Mount, K.E.: 1972, Reps. Prog. Phys. 35, pp. 315-397.
[24] Baltes, H.P., and Hilf, E.R. 1978, 'Spectra of Finite Systems'(B-I Wissenschaftsverlag, Mannheim).
[25] Marcus, R.A.: 1980, in 'Nonlinear Dynamics', Ann. N.Y. Acad. Sci. 357, pp.169-182 (ed. R.H.G. Helleman).
[26] Berry, M.V.: 1981, Ann. Phys. N.Y. 131, pp. 163-216.
[27] Sinai, Ya. G.: 1970, Russ. Math. Surv. 25 No. 2, pp. 137-189.
[28] Bunimovich, L.A.: 1974, Funct. Anal. Appl. 8, pp. 254-255,

1979, Commun. Math. Phys. 65, pp. 295-312.
[29] Berry, M.V., and Tabor, M.: 1976, Proc. Roy. Soc. A349, pp. 101-123.
[30] Berry, M.V., and Tabor, M.: 1977, J. Phys. A.10, pp. 371-379.
[31] Rajaraman, R.: 1975, Phys. Rept. 21, pp. 227-313.
[32] DeWitt, C., Maheshwari, A., and Nelson, B.: 1979, Phys. Rept. 50, pp. 255-372.
[33] Miller, W.H.: 1975, J. Chem. Phys. 63, pp. 996-999.
[34] Voros, A.: 1976, Annls. Inst. H. Poincare XXIV, pp. 31-90.
[35] Gutzwiller, M.C.: 1980, Phys. Rev. Lett. 45, pp. 150-153.
[36] Chazarain, J.: 1974, Inv. Math. 24, pp. 65-82.
[37] Einstein, A.: 1917, Ver. Deut. Phys. Ges. 19, pp. 82-92.
[38] Keller, J.B.: 1958, Ann. Phys. N.Y. 4, pp. 180-188.
[39] Maslov, V.P., and Fedoriuk, M.V., in 'Semi-classical approximation in quantum mechanics' (D. Reidel: Dordrecht 1981) (original Russian edition 1965).
[40] Percival, I.C.: 1977, Advan. Chem. Phys. 36, pp. 1-61.
[41] Percival, I.C.: 1973, J. Phys. B6, pp. L229-232.
[42] Porter, C.F. (ed.): 1965, in 'Statistical Theory of Spectra: Fluctuations' (Academic Press, N.Y. 1965).
[43] McDonald, S.W., and Kaufman, A.N.:1979, Phys. Rev. Lett. 42, pp. 1189-1191.
[44] Casati, G., Valz-Gris, F., and Guarneri, I.: 1980, Nuovo Cimento Lett. 28, pp. 279-282.
[45] Berry, M.V., and Tabor, M.: 1977, Proc. Roy. Soc. A356, pp. 375-394.
[46] Richens, P.J., and Berry, M.V.: 1981, Physica 2D, pp. 495-512.
[47] Neumann, J. von, and Wigner, E.P.: 1929, Physik, Z.30, pp. 467-470.
[48] Teller, E.: 1937, J. Phys. Chem. 41, pp. 109-116.
[49] Pechukas, P.: 1972, J. Chem. Phys. 57, pp. 5577-5594.
[50] Uhlenbeck, K.: 1976, Amer. J. Math. 98, pp. 1059-1078.
[51] Pokrovskii, V.L.: 1966, JETP Lett. 4, pp. 96-99.
[52] Noid, D.W., Koszykowski, M.L., Tabor, M., and Marcus, R.A.: 1980, J. Chem. Phys. 72, pp. 6169-75.