Proc. R. Soc. Lond. A **400**, 229–251 (1985)
Printed in Great Britain

Semiclassical theory of spectral rigidity

By M. V. Berry, F.R.S.

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

(*Received* 20 *February* 1985)

The spectral rigidity $\Delta(L)$ of a set of quantal energy levels is the mean square deviation of the spectral staircase from the straight line that best fits it over a range of L mean level spacings. In the semiclassical limit ($\hbar \rightarrow 0$), formulae are obtained giving $\Delta(L)$ as a sum over classical periodic orbits. When $L \ll L_{\text{max}}$, where $L_{\text{max}} \sim \hbar^{-(N-1)}$ for a system of N freedoms, $\Delta(L)$ is shown to display the following universal behaviour as a result of properties of very long classical orbits: if the system is classically integrable (all periodic orbits filling tori), $\Delta(L) = \frac{1}{15}L$ (as in an uncorrelated (Poisson) eigenvalue sequence); if the system is classically chaotic (all periodic orbits isolated and unstable) and has no symmetry, $\Delta(L) = \ln L/2\pi^2 + D$ if $1 \ll L \ll L_{\text{max}}$ (as in the gaussian unitary ensemble of random-matrix theory); if the system is chaotic and has time-reversal symmetry, $\Delta(L) = \ln L/\pi^2 + E$ if $1 \ll L \ll L_{\text{max}}$ (as in the gaussian orthogonal ensemble). When $L \gg L_{\text{max}}$, $\Delta(L)$ saturates non-universally at a value, determined by short classical orbits, of order $\hbar^{-(N-1)}$ for integrable systems and $\ln(\hbar^{-1})$ for chaotic systems. These results are obtained by using the periodic-orbit expansion for the spectral density, together with classical sum rules for the intensities of long orbits and a semiclassical sum rule restricting the manner in which their contributions interfere. For two examples $\Delta(L)$ is studied in detail: the rectangular billiard (integrable), and the Riemann zeta function (assuming its zeros to be the eigenvalues of an unknown quantum system whose unknown classical limit is chaotic).

1. Introduction

Several statistical measures of the regularity of sequences of eigenvalues were introduced to describe the energy levels of many-particle systems such as nuclei (Porter 1965). Recently these spectral measures have been employed for bound systems with few freedoms, to explore the ways in which the distribution of quantal energy levels reflects integrability or chaos in the underlying classical trajectories. It was expected, and found, that classically integrable systems have levels that are locally uncorrelated and well described by a Poisson distribution; in contrast, classically chaotic systems have levels with strong local repulsion, well described by the eigenvalues of matrices drawn randomly from appropriate ensembles (for reviews see Berry 1983, 1984; Bohigas & Giannoni 1984).

My purpose in this paper is to explain the semiclassical origin and the limits of validity of these two types of spectral universality, by deriving theoretical expressions for one of the spectral statistics, namely the rigidity. This will be defined in (5) in terms of the spectral staircase

$$\mathscr{N}(E) \equiv \sum_n \Theta(E-E_n), \tag{1}$$

where $E_n = E_1, E_2 \ldots$ is the eigenvalue sequence and Θ denotes the unit step function. We also require the spectral density

$$d(E) \equiv \mathrm{d}\mathscr{N}(E)/\mathrm{d}E = \sum_n \delta(E-E_n). \tag{2}$$

For a system with N freedoms the local averages of these functions are

$$\langle \mathscr{N}(E)\rangle = \Omega(E)/h^N; \quad \langle d(E)\rangle = (d\Omega(E)/dE)/h^N, \tag{3}$$

where $\Omega(E)$ is the classical phase-space volume enclosed by the surface with energy E, given in terms of the Hamiltonian $H(q,p)$ by

$$\Omega(E) = \int \mathrm{d}^N q \int \mathrm{d}^N p\, \Theta(E-H(q,p)). \tag{4}$$

In the cases of interest here, $N \geqslant 2$.

The averages denoted by $\langle \quad \rangle$ in (3) refer to all levels in an energy range that is classically small, i.e. small in comparison with E, but semiclassically large, i.e. large in comparison with the mean level spacing $\langle d\rangle^{-1} \sim \hbar^N$. This mean level spacing will play an important role in what follows, and will be called the inner energy scale.

The rigidity $\Delta(L)$ is now defined as the local average of the mean square deviation of the staircase from the best fitting straight line over an energy range corresponding to L mean level spacings, namely

$$\Delta(L) \equiv \left\langle \min_{(A,B)} \frac{\langle d(E)\rangle}{L} \int_{-L/2\langle d\rangle}^{L/2\langle d\rangle} \mathrm{d}\epsilon\, [\mathscr{N}(E+\epsilon)-A-B\epsilon]^2 \right\rangle. \tag{5}$$

This function was introduced by Dyson & Mehta (1963) (they called it Δ_3 to distinguish it from two less useful statistics). Minimizing over A and B leads to

$$\Delta(L) = \left\langle \left\{ \frac{\langle d\rangle}{L} \int_{-L/2\langle d\rangle}^{L/2\langle d\rangle} \mathrm{d}\epsilon\, \mathscr{N}^2(E+\epsilon) - \left[\frac{\langle d\rangle}{L} \int_{-L/2\langle d\rangle}^{L/2\langle d\rangle} \mathrm{d}\epsilon\, \mathscr{N}(E+\epsilon)\right]^2 - 12\left[\frac{\langle d\rangle^2}{L^2} \int_{-L/2\langle d\rangle}^{L/2\langle d\rangle} \mathrm{d}\epsilon\, \epsilon\, \mathscr{N}(E+\epsilon)\right]^2 \right\} \right\rangle. \tag{6}$$

When $L \ll 1$, the fact that $\mathscr{N}(E)$ is a staircase leads to the limit $\Delta \to \frac{1}{15}L$ whatever distribution the levels have (provided this is non-singular). Therefore the spectral rigidity gives no information about the very finest scales corresponding to the spacings between neighbouring levels. Its usefulness lies in the way it describes correlations over level sequences longer than the inner energy scale (which corresponds to $L = 1$).

We shall demonstrate the existence of two universality classes of rigidity, extending from $L \sim 1$, corresponding to the inner energy scale, to a value $L_{\max}$

corresponding to an outer energy scale $h/T_{\min}$, where $T_{\min}$ is the period of the shortest classical closed orbit. Thus

$$L_{\max} \equiv h\langle d\rangle / T_{\min} \sim \hbar^{-(N-1)} \tag{7}$$

greatly exceeds unity, in spite of the fact that the outer energy scale is of order $\hbar$ and hence classically small. Now we can be more precise about the local averages denoted by $\langle \quad \rangle$: these correspond to energy ranges much larger than the outer scale but still classically small, for example energy ranges of order $\hbar^{\frac{1}{2}}$.

The first universality class occurs for classically integrable systems. In these, as will be shown in §3, the Poisson form $\frac{1}{15}L$ extends from $L = 0$ to $L_{\max}$. For $L > L_{\max}$, $\Delta(L)$ reaches a saturation value (not universal), Δ_∞, which can be regarded as a measure of the totality of the spectral fluctuations on all scales. This extends earlier work (Berry & Tabor 1977*a*) in which Poisson statistics were shown to describe local fluctuations, and, as will be shown explicitly in §4 it explains recent numerical results of Casati *et al.* (1985) on the rectangular billiard.

The second universality class occurs for classically chaotic systems. In these, as will be shown in §6, $\Delta(L)$ increases only logarithmically in the range $1 \lesssim L < L_{\max}$, which indicates long-range rigidity in the level distribution. The fact that for systems with time-reversal symmetry the coefficient of the logarithm is twice what it is when there is no such symmetry is given a simple semiclassical explanation in §8. For $L > L_{\max}$, $\Delta(L)$ reaches a saturation value (not universal), Δ_∞, much smaller than in the integrable case. When applied to the 'level sequence' consisting of the imaginary parts of the zeros of the Riemann zeta function (§7), our results are consistent with what little is known and conjectured about the asymptotics of this sequence.

In deriving these results from (6), semiclassical methods are essential. The reason is that for local spectral statistics to attain well defined limiting values, classically small energy ranges must contain many levels, and this happens only as $\hbar \to 0$ (for scaling systems such as billiards, this is equivalent to $E \to \infty$). The semiclassical technique employed here (in §2) is the representation of the spectral density $d(E)$ as a sum over all the periodic orbits of the classical system, introduced and developed by Gutzwiller (1967, 1969, 1970, 1971, 1978) and Balian & Bloch (1972, 1974).

When applied to the calculation of spectral rigidity, the periodic-orbit technique requires two further ingredients. The first is a classical sum rule for the orbit intensities, recently discovered by Hannay & Ozorio de Almeida (1984), and the second is a new semiclassical sum rule derived in §5.

The arguments and conclusions of this paper complement those of Pechukas (1983). He obtained all the spectral statistics, but made use of a statistical assumption about the wavefunction. I make no such statistical assumption, but discuss only $\Delta(L)$.

2. Rigidity in terms of periodic orbits

Periodic-orbit theory gives the semiclassical spectral density as

$$d(E) = \langle d(E)\rangle + d_{osc}(E), \tag{8}$$

where $d_{osc}(E)$ is a sum over classical periodic orbits, each of which contributes an oscillatory function and whose combined effect is to produce, by constructive interference, a sequence of singularities (such as the δ functions in (2) or approximations to them) minus $\langle d\rangle$. We shall write d_{osc} in a form appropriate to systems that are completely integrable or completely chaotic. In an integrable system, closed orbits are not isolated but form $(N-1)$–parameter families filling N-dimensional phase-space tori. A chaotic system we define as one that is ergodic and for which, in addition, all closed orbits are isolated and therefore unstable. The oscillatory contribution in (8) is

$$d_{osc}(E) = \frac{1}{\hbar^{\mu+1}} \sum_j A_j(E) \exp\{iS_j(E)/\hbar\}. \tag{9}$$

Detailed descriptions and discussions of this formula were given by Berry (1983, 1984); here we need the following facts. In the sum, j labels all distinct periodic orbits including all multiple traversals, positive and negative (but not zero). It will be of crucial importance that negative traversals correspond to retracings, where the orbit is followed backwards in time, and not to time-reversed orbits (the latter exist only for systems with time-reversal symmetry). The exponent μ is $\frac{1}{2}(N-1)$ for integrable systems and zero for chaotic ones. This difference is important and for integrable systems the periodic orbits on a torus combine coherently to produce much stronger spectral oscillations than an isolated orbit of a chaotic system, as well as giving the first hint that the level statistics will be different too. The (real) amplitudes A_j will be discussed later. In the exponent, the phase contains the action $S_j(E)$, defined for m traversals as

$$S_j(E) \equiv m\left\{\oint \boldsymbol{p}\cdot d\boldsymbol{q} + \alpha\hbar\right\}, \tag{10}$$

where the integral is over a single traversal and $\alpha\hbar$, which will play no part in what follows, gives focusing corrections such as Maslov indices. Because of the negative traversals, d_{osc} is real.

In (9) the energy dependence of the oscillations is determined by the orbit periods $T_j(E)$ because

$$T_j(E) = dS_j(E)/dE. \tag{11}$$

The longest oscillation, giving the largest scale of spectral fluctuations, comes from the shortest orbit and has 'wavelength' given by the outer scale h/T_{min}, already defined. Thus the constructive interference that gives δ functions, whose mean spacing is the inner scale $\langle d\rangle^{-1} \approx \hbar^N$, is determined by very long orbits, with periods $T \sim \hbar^{-(N-1)}$.

To incorporate the oscillations (9) into the rigidity formula (6) we use the fact that the energy range $L/\langle d\rangle$ is classically small (although it may be semiclassically large) to write

$$S_j(E+\epsilon) \approx S_j(E) + \epsilon T_j(E) \tag{12}$$

and ignore the ϵ dependences of A_j and $\langle d\rangle$. Thus the spectral staircase is

$$\mathcal{N}(E) = \langle \mathcal{N}(E)\rangle + \mathcal{N}_{\rm osc}(E),$$

where

$$\mathcal{N}_{\rm osc}(E+\epsilon) = \frac{-i}{\hbar^{\mu}} \sum_j \frac{A_j}{T_j} \exp\{i(S_j+\epsilon T_j)/\hbar\}. \tag{13}$$

The integrals in (6) are now elementary and give

$$\Delta(L) = \left\langle \frac{1}{\hbar^{2\mu}} \sum_i \sum_j \frac{A_i A_j}{T_i T_j} \exp\{i(S_i - S_j)/\hbar\} \times [F(y_i - y_j) - F(y_i)F(y_j) - 3F'(y_i)F'(y_j)] \right\rangle, \tag{14}$$

where

$$y_j \equiv LT_j/2\hbar\langle d\rangle, \tag{15}$$

$$F(y) \equiv \sin y/y, \tag{16}$$

and primes denote differentiation.

To arrive at (14) we made use of the result

$$\langle \exp\{iS_j/\hbar\}\rangle \to 0 \quad \text{as} \quad \hbar \to 0, \tag{17}$$

which holds because the local averaging is over an energy range much greater than the outer scale. It is tempting to think that the same principle of destructive interference will eliminate the non-diagonal terms $i \neq j$ in (14). This is the case for integrable systems (as will be shown in §5), but for chaotic systems the proliferation of pairs of very long orbits with action differences $(S_i - S_j) < \hbar$ is important and we shall see that local averaging does not diagonalize the sum.

However, the restriction to pairs with $(S_i - S_j)/\hbar < 1$ does have the effect that in the functions F in (14) we can set $y_i = y_j$. The reason is that, for long orbits,

$$S_j \to T_j N\Omega/(\mathrm{d}\Omega/\mathrm{d}E) \tag{18}$$

(Hannay & Ozorio de Almeida 1984), and together with (3) this implies

$$|y_i - y_j| \to \frac{|S_i - S_j|}{\hbar} \frac{L}{2N\mathcal{N}}, \tag{19}$$

which vanishes because $L \ll \mathcal{N}$. Thus the rigidity becomes

$$\Delta(L) = \frac{2}{\hbar^{2\mu}} \int_0^\infty \frac{\mathrm{d}T}{T^2} \phi(T)\, G(LT/2\langle d\rangle\hbar), \tag{20}$$

where

$$G(y) \equiv 1 - F^2(y) - 3(F'(y))^2 \tag{21}$$

and

$$\phi(T) \equiv \left\langle \sum_i \sum_j{}^{+} A_i A_j \cos\{(S_i - S_j)/\hbar\}\, \delta\{T - \tfrac{1}{2}(T_i + T_j)\} \right\rangle \tag{22}$$

in which the + on the summation denotes restriction to positive traversals (i.e. $T_j > 0$).

Figure 1 shows the important function $G(y)$. Because G is small if $y \ll 1$, it selects from the sum (22) only those pairs of orbits whose average period exceeds $2\langle d\rangle\hbar/L$; therefore G will be called the orbit selection function. Such selection is physically reasonable because $\Delta(L)$ is defined by (5) in terms of deviations from a linear approximation to the staircase over an energy range $L/\langle d\rangle$: the periodic-orbit sum (13) shows that this linear approximation is determined by orbits with $T_j < \langle d\rangle\hbar/L$, and deviations from it by orbits with $T_j > \langle d \rangle \hbar / L$. The effects of this orbit selection depend strongly on the value of L in comparison with unity and L_{max} (equation 7). For the shortest closed orbit, (7) shows that when $L = L_{\text{max}}$ the argument of G is $y = \pi$.

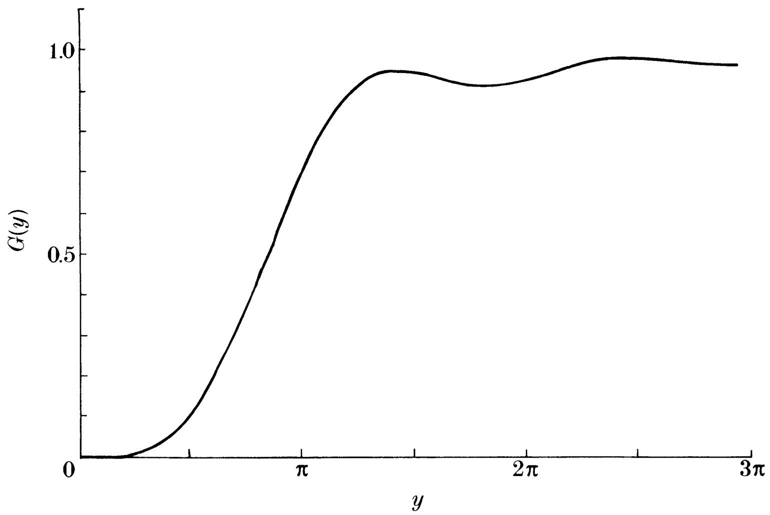

FIGURE 1. Orbit selection function defined by (21).

It will be important to know the large-T limiting form of the diagonal sum in (22), namely

$$\phi_{\text{D}}(T) \equiv \langle \textstyle\sum^{+} A_j^2 \, \delta(T - T_j) \rangle . \tag{23}$$

This is the number density of orbits with periods near T, weighted with intensities A_j^2. As $T \to \infty$, the density of periodic orbits increases and the intensities decrease: as a power-law for integrable systems and exponentially for chaotic ones. The results of the competition between these two tendencies was calculated by Hannay & Ozorio de Almeida (1984) as a consequence of their extension of the important idea that very long periodic orbits are uniformly distributed in phase space (see, for example, Parry and Pollicot 1983; Parry 1984). They found that

$$\phi_{\text{D}}(T) \to (\mathrm{d}\Omega / \mathrm{d}E)/(2\pi)^{N+1} \quad \text{(integrable)} \tag{24}$$

and

$$\phi_{\text{D}}(T) \to T/4\pi^2 \quad \text{(chaotic)}. \tag{25}$$

(The integrable result (24) had previously been found by Berry & Tabor (1977*a*), in a different way.)

These formulae correspond to the amplitudes A_j combining incoherently as in (23), which is the correct procedure if no symmetry enforces strict degeneracy among the orbit actions S_j. When such degeneracy does exist, the appropriate amplitudes must be combined coherently, and we shall see in §8 that this has important repercussions for time-reversal symmetry. In writing (25) we have, for simplicity, ignored multiple traversals ($M > 1$ in 10), because these give contributions to ϕ_{D} that vanish as $T \to \infty$ (this will be illustrated in §7).

3. Integrable systems

For integrable systems, global action-angle variables exist, with each set of actions $\boldsymbol{I} = \{I_1 \ldots I_N\}$ denoting a phase-space torus. The Hamiltonian can be written $H(\boldsymbol{I})$ and the frequencies $\boldsymbol{\omega} = \{\omega_1 \ldots \omega_N\}$ on the torus $\boldsymbol{I}$ are given by $\boldsymbol{\omega} = \nabla_{\boldsymbol{I}} H(\boldsymbol{I})$. The periodic orbits at energy E are knots on the torus, with winding numbers $\boldsymbol{M} = \{M_1 \ldots M_N\}$ for the N irreducible cycles. These winding numbers constitute the label j in the periodic-orbit sums of the previous section. Each $\boldsymbol{M}$ defines a resonant torus $\boldsymbol{I}_{\boldsymbol{M}}$, which is one whose frequencies are commensurable, and hence a period $T_{\boldsymbol{M}}$, by

$$\boldsymbol{\omega}(\boldsymbol{I}_{\boldsymbol{M}}) = 2\pi \boldsymbol{M}/T_{\boldsymbol{M}}; \quad H(\boldsymbol{I}_{\boldsymbol{M}}) = E. \tag{26}$$

A convenient form for the orbit amplitudes $A_{\boldsymbol{M}}$ is that given by Berry & Tabor (1977*b*):

$$A_{\boldsymbol{M}}^2 = \frac{(2\pi)^{N-1}}{T_{\boldsymbol{M}}^N \,|\, \boldsymbol{\omega} \cdot \partial \boldsymbol{I}_{\boldsymbol{M}}/\partial T_{\boldsymbol{M}} \det\{\partial\omega_i/\partial I_j\}_{\boldsymbol{M}} \,|}. \tag{27}$$

For the rigidity, (20)–(22), together with the diagonal average to be justified later, give the topological sum

$$\Delta(L) = \frac{2}{\hbar^{N-1}} \sum_{\boldsymbol{M}}^{+} \frac{A_{\boldsymbol{M}}^2}{T_{\boldsymbol{M}}^2} G(LT_{\boldsymbol{M}}/2\langle d \rangle \hbar). \tag{28}$$

Thus $\Delta(L)$ is a sum of weighted scaled orbit selection functions (figure 1), one for each resonant torus.

When $L \ll L_{\max}$, G selects only long orbits and the topological sum can be evaluated by using the continuum limit (24). This gives

$$\begin{aligned}\Delta(L) &= \frac{\hbar}{(2\pi\hbar)^N \pi} \frac{\mathrm{d}\Omega}{\mathrm{d}E} \int_0^\infty \frac{\mathrm{d}T}{T^2} G(LT/2\langle d \rangle \hbar) \\ &= \frac{L}{2\pi} \int_0^\infty \frac{\mathrm{d}y}{y^2} G(y).\end{aligned} \tag{29}$$

The integral, from (21) and (16), equals $\frac{2}{15}\pi$, so that

$$\Delta(L) = \tfrac{1}{15} L \quad (L \ll L_{\max}). \tag{30}$$

This supports the claim that the local spectra of classically integrable systems belong to the universality class of uncorrelated level sequences; of course 'local' means $L \ll L_{\max}$.

When $L \gg L_{\max}$, the orbit selection function in (28) is unity for all orbits $\boldsymbol{M}$, and Δ attains a saturation value given by the convergent sum

$$\Delta_\infty = \frac{2}{\hbar^{N-1}} \sum_{\boldsymbol{M}}^{+} \frac{A_{\boldsymbol{M}}^2}{T_{\boldsymbol{M}}^2}. \tag{31}$$

Although this is semiclassically large, the r.m.s. fluctuations $\Delta_\infty^{\frac{1}{2}}$ in the staircase are still much smaller than the mean height $\langle \mathcal{N} \rangle$ of the staircase itself (equation (3)), by a factor $\hbar^{\frac{1}{2}(N+1)}$. Saturation of the rigidity has been observed in numerical calculations by Seligman *et al.* (1985) (for an integrable polynomial Hamiltonian) and Casati *et al.* (1985) (for an integrable billiard (see the next section)).

In the crossover region $L \sim L_{\max}$, (28) predicts a few weak oscillations as L increases to reveal the contribution of the shortest orbit with period $T_{\min}$. The slowest oscillations in $\Delta(L)$ have an L-wavelength of $L_{\max}$.

4. An example: billards in a rectangle

For a particle of mass m moving freely within a rectangle with sides a, b and impenetrable walls, the quantal energy levels are

$$E_{l,\,n} = \hbar^2 \pi^2 (l^2 \alpha^{-\frac{1}{2}} + n^2 \alpha^{\frac{1}{2}})/2mab, \tag{32}$$

where

$$\alpha \equiv a^2/b^2. \tag{33}$$

We assume without loss of generality that $a \geqslant b$, i.e. $\alpha \geqslant 1$. Classically, the action Hamiltonian for this two-dimensional integrable system is

$$H(I_1, I_2) = \pi^2\,(I_1^2/a^2 + I_2^2/b^2)/2m \tag{34}$$

with frequencies

$$\boldsymbol{\omega} = \pi^2 (I_1/a^2, I_2/b^2)/m. \tag{35}$$

It follows from (26) that the period of the closed orbit with topology $\boldsymbol{M} = (M_1, M_2)$ is

$$T_{\boldsymbol{M}} = [2m(M_1^2 a^2 + M_2^2 b^2)/E]^{\frac{1}{2}}. \tag{36}$$

Three of these orbits are illustrated in figure 2. The resonant tori whose orbits have these periods have actions

$$\boldsymbol{I_M} = 2m\,(M_1 a^2, M_2 b^2)/\pi T_{\boldsymbol{M}}. \tag{37}$$

In terms of these quantities it is easy to calculate the torus amplitudes (27)

$$A_{\boldsymbol{M}}^2 = m^2 a^2 b^2/\pi^3 E T_{\boldsymbol{M}}. \tag{38}$$

It might be thought that time-reversal symmetry of the periodic orbits might cause them to contribute twice to the sum (28) for the rigidity. But this is not the case, because although each orbit on the torus $\boldsymbol{I_M}$ is distinct from its time-reverse if neither M_1 nor M_2 is zero, both these orbits are included on the same torus, which is a four-sheeted geometric object with inversion symmetry about $\boldsymbol{p} = 0$ in momentum space, and hence time-reversal symmetry. If either M_1 or M_2 is zero (for example the orbit $(0, 1)$ in figure 2) the orbit is self-retracing and hence is its own time-reverse. The corresponding torus has only two sheets rather than four and so counts $\frac{1}{2}$ in amplitude and $\frac{1}{4}$ in intensity (for a detailed discussion of this phenomenon see Appendix C of Richens & Berry (1981)).

It is natural to express $\Delta(L)$ in terms of the scaled energy

$$\mathscr{E} \equiv E\langle d\rangle\ (= \langle \mathscr{N}(E)\rangle), \tag{39}$$

corresponding to a mean level spacing of unity, which by (28) gives

$$\Delta(L) = \frac{\mathscr{E}^{\frac{1}{2}}}{\pi^{\frac{5}{2}}} \sum_{M_1=0}^{\infty} \sum_{M_2=0}^{\infty} \frac{\delta_{\boldsymbol{M}}\, G(y_{\boldsymbol{M}})}{(M_1^2 \alpha^{\frac{1}{2}} + M_2^2 \alpha^{-\frac{1}{2}})^{\frac{3}{2}}}, \tag{40}$$

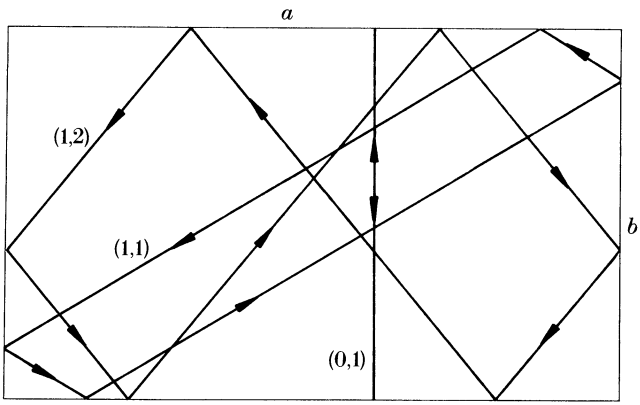

FIGURE 2. Rectangular billiard showing closed orbits with three different pairs of winding numbers (M_1, M_2). The orbit $(0, 1)$ is self-retracing; the other two are not.

where
$$\delta_M \equiv \begin{cases} 0 & \text{if } M_1 = M_2 = 0, \\ \frac14 & \text{if one of } M_1 \text{ and } M_2 \text{ is zero,} \\ 1 & \text{otherwise} \end{cases} \tag{41}$$

and
$$y_M = L\{\pi(M_1^2\alpha^{\frac12} + M_2^2\alpha^{-\frac12})/\mathscr{E}\}^{\frac12}. \tag{42}$$

The shortest periodic orbit has winding numbers $(0, 1)$ and (7) leads to
$$L_{\max} = (\pi\mathscr{E})^{\frac12}\alpha^{\frac14}. \tag{43}$$

When $L \gg L_{\max}$, $G \approx 1$ for all $\boldsymbol{M}$ and so the saturation rigidity is
$$\Delta_\infty = \frac{\mathscr{E}^{\frac12}}{\pi^{\frac52}} \sum_{M_1=0}^{\infty} \sum_{M_2=0}^{\infty} \frac{\delta_M}{(M_1^2\alpha^{\frac12} + M_2^2\alpha^{-\frac12})^{\frac32}}. \tag{44}$$

This convergent series is easy to sum numerically. For α close to unity it varies slowly with α and so can be approximated by its value when $\alpha = 1$, which is (Zucker 1974)
$$\Delta_\infty \overset{(\alpha=1)}{=} \mathscr{E}^{\frac12}\pi^{-\frac52}[\zeta(\tfrac32)\beta(\tfrac32) - \tfrac12\zeta(3)] = 0.0947\,\mathscr{E}^{\frac12}, \tag{45}$$

where ζ and β denote the number-theoretic series
$$\zeta(s) \equiv \sum_{n=1}^{\infty} \frac{1}{n^s}; \quad \beta(s) \equiv \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)^s}. \tag{46}$$

These results will now be compared with quantal calculations of $\Delta(L)$ by Casati *et al.* (1985). They used the levels $\alpha l^2 + n^2$ and so our $\mathscr{E}$ is their energy multiplied by $\pi/4\alpha^2$. Figure 3 shows the comparison for two energies, between $\Delta(L)$ computed from (40) for $\alpha = 1$ and their quantal calculations for an ensemble of α-values between 0.9 and 1.2 (a range in which the different theoretical Δ-curves are almost indistinguishable). The agreement is good in the universal crossover and saturation ranges of L, although their oscillations are slightly stronger (this might be the result of their different averaging procedure).

Casati *et al.* also show (non-local) averages of $\Delta(L)$ over the whole energy range from zero to $\mathscr{E}$, for $\alpha = \frac{1}{3}\pi$. From (45), their curves ought to saturate at $\frac{2}{3} \times 0.0947\, \mathscr{E}^{\frac{1}{2}}$, and comparison shows that they do.

Finally, Casati *et al.* show a graph of $\Delta(L)$ in which L consists of the range from the ground state to the Lth level. In our notation this corresponds to choosing $\mathscr{E} = \frac{1}{2}L$ so that L is always in the saturation range and (45) predicts

$$\Delta = 0.0947\, L^{\frac{1}{2}}/\sqrt{2} = 0.067 L^{\frac{1}{2}}. \tag{47}$$

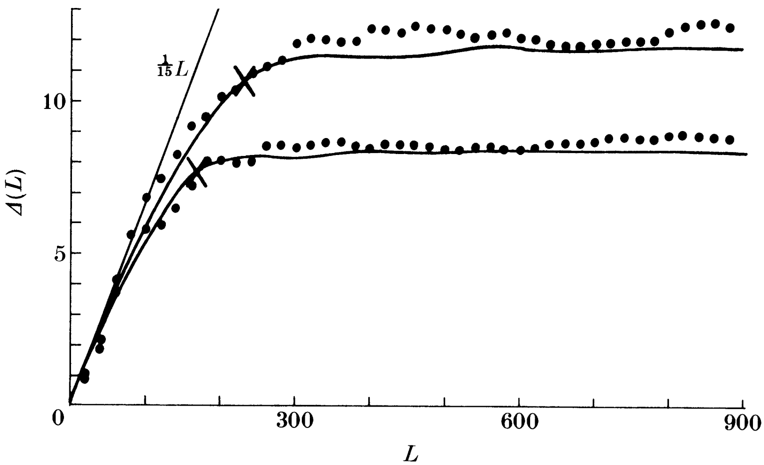

FIGURE 3. The full curves show $\Delta(L)$ computed from (40) for $\alpha = 1$ and $\mathscr{E} = 10500 \times \frac{1}{4}\pi$ (lower curve) and $\mathscr{E} = 20500 \times \frac{1}{4}\pi$ (upper curve) by using 1250 closed orbits. The circular points show data from Casati *et al.* (1985). The theoretical crossover values $L_{\max}$ are 161 and 225 and are indicated by crosses. The straight line shows the local universal Poisson rigidity $\frac{1}{15}L$.

Their curve is fitted by $0.063 L^{\frac{1}{2}}$, which is a close agreement when one considers that the theoretical formulae have here been applied to values of L that are certainly not classically small.

5. A SEMICLASSICAL SUM RULE

The periodic-orbit sum (9) can be at best conditionally convergent because it represents the spectral density, with δ singularities at the energy levels. The delicate conspiracy of amplitudes A_j and phases $S_j/\hbar$ by which this is achieved is not fully understood, but there is some theoretical evidence that the semiclassical approximation (9) is in fact capable of reproducing singularities when infinitely long orbits are included. For integrable systems, we refer to Norcliffe & Percival (1968), Balian & Bloch (1972), and Berry & Tabor (1976). For chaotic systems, we refer to the Selberg identity (reviewed by Hejhal (1976) and McKean (1972)), which shows that for manifolds of constant negative curvature (9) is exact, and also to a study by Gutzwiller (1980) and an example to be given in §7 here. The purpose of the present section is to derive an identity that must be satisfied by the function $\phi(T)$ (equation (22)) that appears in the rigidity formula (20), to ensure that the periodic-orbit sum (9) has the correct density of singularities.

By analytically continuing the energy to $E \rightarrow E + i\eta$ and using the representation of the spectral density (2) as the imaginary part of the causal Green function, $d(E)$ can be expressed as the limit $\eta \rightarrow 0$ of the Lorentzians

$$d_\eta(E) = -\frac{1}{\pi}\,\mathrm{Im}\sum_n \frac{1}{E - E_n + i\eta}. \tag{48}$$

If

$$\eta\,\langle d\rangle \ll 1 \tag{49}$$

the Lorentzians do not overlap, and so

$$d_\eta^2(E) = \frac{\eta^2}{\pi^2}\sum_n \frac{1}{[(E-E_n)^2+\eta^2]^2}. \tag{50}$$

It now follows from

$$\frac{2\eta^3}{\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}x}{(x^2+\eta^2)^2} = 1 \tag{51}$$

that

$$d(E) = \lim_{\eta\rightarrow 0} 2\pi\eta\, d_\eta^2(E). \tag{52}$$

By taking local averages and using the representation (8),

$$\langle d(E)\rangle = \lim_{\eta\rightarrow 0} 2\pi\eta\,\langle d^2_{\mathrm{osc},\,\eta}(E)\rangle. \tag{53}$$

The semiclassical formula for $d_{\mathrm{osc},\,\eta}(E)$, analogous to (9) for $d_{\mathrm{osc}}(E)$, involves only positive traversals of periodic orbits, and gives

$$\langle d(E)\rangle = \lim_{\eta\rightarrow 0}\frac{4\pi\eta}{\hbar^{2\mu+2}}\Big\langle\sum_i \sum_j{}^{+} A_i A_j \cos\{(S_i - S_j)/\hbar\}\exp\{-\eta(T_i+T_j)/\hbar\}\Big\rangle \tag{54}$$

$$= \lim_{\eta\rightarrow 0}\frac{4\pi\eta}{\hbar^{2\mu+2}}\int_0^\infty \mathrm{d}T\,\phi(T)\exp\{-2\eta T/\hbar\}, \tag{55}$$

where (12) has been used with $\epsilon = i\eta$ and where ϕ is defined by (22).

Asymptotic inversion of the Laplace transform and use of (49) now gives

$$\phi(T) \rightarrow \langle d\rangle\,\hbar^{2\mu+1}/2\pi \quad \text{if} \quad T \gg \hbar\langle d\rangle. \tag{56}$$

This is the semiclassical sum rule. It guarantees that the amplitudes and phases of very long orbits generate the mean level density, and hence shows how late terms in the representation (8) and (9) determine the first term: an 'analytic bootstrap' reminiscent of that introduced in one dimension by Voros (1983).

If we define a new variable to measure time in relation to the inner energy scale, i.e.

$$\tau \equiv T/h\langle d\rangle, \tag{57}$$

and write

$$\phi(T) \equiv \langle d\rangle \hbar^{2\mu+1} K(\tau)/2\pi. \tag{58}$$

then the sum rule (56) gives

$$K(\tau)\to 1 \quad \text{when} \quad \tau \gg 1. \tag{59}$$

The physical interpretation of $K(\tau)$ is that this function is the spectral form factor, defined as the Fourier transform of the correlation function of the spectral density:

$$K(\tau) = \langle d\rangle^{-2}\int_{-\infty}^{\infty} \mathrm{d}L\, \langle d(E-L/2\langle d\rangle)\, d(E+L/2\langle d\rangle)\rangle \exp\{2\pi i L\tau\}. \tag{60}$$

(apart from a δ function at $\tau = 0$). In terms of K, the pair correlation function of the levels is

$$g(L) = 1-\frac{1}{\pi L}\int_0^{\infty} \mathrm{d}\tau \sin\{2\pi L\tau\}\, K'(\tau). \tag{61}$$

For *integrable* systems, (56) can be written

$$\phi(T)\to(\mathrm{d}\Omega/\mathrm{d}E)/(2\pi)^{N+1} \quad \text{if} \quad T \gg \hbar\langle d\rangle, \tag{62}$$

which is identical to the asymptotic value (24) of the diagonal sum. Thus neglect of off-diagonal terms in ϕ is indeed justified for these systems, as asserted previously. Moreover, the form factor $K(\tau)$ is unity not only when $\tau \gg 1$ (as in (59)), but also down to the much smaller value

$$\tau_{\min} = T_{\min}/\hbar\langle d\rangle \quad (\ll 1). \tag{63}$$

It then follows from (61) that the pair correlation is unity, which implies lack of level correlation, for sequences of length $L \ll L_{\max}$, consistent with the behaviour already found for $\Delta(L)$ in §3.

6. Chaotic systems without time-reversal symmetry

For chaotic systems the diagonal approximation to $\phi(T)$, which must be valid for any given $T \gg T_{\min}$ if $\hbar$ is small enough, is (25). Together with (59) and (63) this implies that the form factor $K(\tau)$ defined by (57) and (58) has the behaviour

$$K(\tau)\to\begin{cases}\tau & (\tau_{\min} \ll \tau \ll 1),\\ 1 & (\tau \gg 1).\end{cases} \tag{64}$$

The rigidity is given by (20) as

$$\Delta(L) = \frac{1}{2\pi^2}\int_0^{\infty} \frac{\mathrm{d}y}{y}\,\frac{K(y/\pi L)}{y/\pi L}\, G(y). \tag{65}$$

When $L \ll 1$, K can be replaced by unity (because of (64)) to give for $\Delta(L)$ the correct limiting form $\frac{1}{15}L$ (cf. (29) and (30) and the discussion below (6)). This result could not have been obtained without the semiclassical sum rule.

When $1 \ll L \ll L_{\max}$, it is possible to divide the integration range of (65) into two parts by choosing a value of Y that satisfies $1 \ll Y \ll L$, so that, from (64) and figure 1,

$$\left.\begin{aligned}\frac{K(y/\pi L)}{y/\pi L} &\approx 1 \quad \text{if} \quad y < Y\\ \text{and} \qquad G(y) &\approx 1 \quad \text{if} \quad y > Y.\end{aligned}\right\} \tag{66}$$

Thus
$$\Delta(L) = \frac{1}{2\pi^2}\left[\int_0^Y \frac{\mathrm{d}y}{y}\,G(y) + \int_Y^\infty \frac{\mathrm{d}y}{y}\,\frac{K(y/\pi L)}{y/\pi L}\right]. \tag{67}$$

The definitions (16) and (21) lead to
$$\int_0^Y \mathrm{d}y\, G(y)/y = \ln Y + \gamma + \ln 2 - \tfrac{9}{4}, \tag{68}$$

where γ is the Euler constant 0.577.... Integration by parts gives
$$\int_Y^\infty \frac{\mathrm{d}y}{y}\,\frac{K(y/\pi L)}{y/\pi L} = -\ln(Y/\pi L) - \int_0^\infty \mathrm{d}\tau\,\ln\tau\,\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{K(\tau)}{\tau}\right). \tag{69}$$

Thus the rigidity is
$$\Delta(L) = (\ln L)/2\pi^2 + D \quad (1 \ll L \ll L_{\max}), \tag{70}$$

where
$$D = \frac{1}{2\pi^2}\left[\ln 2\pi + \gamma - \tfrac{9}{4} - \int_0^\infty \mathrm{d}\tau\,\ln\tau\,\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{K(\tau)}{\tau}\right)\right]. \tag{71}$$

Equation (70) is precisely the asymptotic rigidity of the gaussian unitary ensemble (g.u.e.) of random-matrix theory (Mehta 1967), i.e. $\Delta(L)$ averaged over the spectra of large Hermitian matrices whose elements are gaussian random variables with statistics invariant under unitary transformations. The logarithmic dependence and correct prefactor are a direct consequence of the diagonal sum rule (25) given by Hannay & Ozorio de Almeida (1984), which applies when all closed orbits are isolated and unstable, with no action degeneracies. Without the semiclassical sum rule, however, the additive constant D would be infinite (because (69) would then be illegitimate).

Without specifying $K(\tau)$ (or what is equivalent, $\phi(T)$) more closely than (64), the value of the constant D cannot be determined, and I know no direct semiclassical arguments based on the definition (22) by which this can be achieved. However, with the simplest interpolation, namely
$$K_0(\tau) = \begin{cases} \tau & (\tau \leqslant 1), \\ 1 & (\tau \geqslant 1), \end{cases} \tag{72}$$

the integral in (71) is -1 and
$$D = (\ln 2\pi + \gamma - \tfrac{5}{4})/2\pi^2 = 0.0590, \tag{73}$$

which is exactly the correct constant given by random-matrix theory for the g.u.e.! The reason is that (72) is the exact form factor of the g.u.e. (Mehta 1967). However, this must be regarded as a remarkable coincidence, because where there is time-reversal symmetry we shall see (§8) that the simplest interpolation fails to give the exact result. In any case, D is small and not very sensitive to $K(\tau)$. To illustrate this, the interpolations
$$K_1(\tau) = \tau/(1+\tau), \quad K_2(\tau) = \tau/(1+\tau^2)^{\frac{1}{2}}, \quad K_3(\tau) = 2\tau\pi^{-1}\arctan(\pi/2\tau) \tag{74}$$

give
$$D_1 = 0.0083, \quad D_2 = 0.0434, \quad D_3 = 0.0312. \tag{75}$$

The discontinuity in slope of the 'correct' form factor (71) is very surprising. It implies that as $\hbar \to 0$ the double sum (22) for $\phi(T)$ has an abrupt transition at $T = h\langle d \rangle$, between the diagonal (25) and that given by the semiclassical sum rule (56). The origin of this 'semiclassical phase phase-transition' is at present obscure. In the next section we shall present a curious example of it. (It should also be remarked that the g.u.e. repulsion between *neighbouring* levels, which causes $g(L)$ (equation (61)) to vanish as L^2 when $L \to 0$, cannot be obtained from the semiclassical arguments leading to (64), but is of course implicit in (72).)

The preceding results explain local universality of the rigidity when $L \ll L_{\max}$, and the logarithmic behaviour (70) has recently been observed in computations for a chaotic system without time-reversal symmetry by Seligman *et al.* (1985).

When $L \gg L_{\max}$, short orbits, and hence $\tau \sim \tau_{\min}$ (equation (63)) give important contributions to the saturation value, which is given by (20) as

$$\Delta_\infty = 2\int_0^\infty \mathrm{d}T\,\phi(T)/T^2. \tag{76}$$

This can be expressed in terms of the short (non-universal) orbits and the (universal) density of the long ones by introducing an 'intermediate' period T_{I} such that

$$T_{\min} \ll T_{\mathrm{I}} \ll h\langle d \rangle. \tag{77}$$

Then

$$\begin{aligned}\Delta_\infty &= 2\sum_{T_j<T_{\mathrm{I}}} \frac{A_j^2}{T_j^2} + \frac{1}{2\pi^2}\int_{T_{\mathrm{I}}/h\langle d\rangle}^\infty \frac{\mathrm{d}\tau}{\tau^2}\,K(\tau) \\ &= 2\sum_{T_j<T_{\mathrm{I}}} \frac{A_j^2}{T_j^2} + \frac{1}{2\pi^2}\ln\left\{\frac{h\langle d\rangle}{T_{\mathrm{I}}}\right\} - \frac{1}{2\pi^2}\int_0^\infty \mathrm{d}\tau\,\ln\tau\,\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{K(\tau)}{\tau}\right).\end{aligned} \tag{78}$$

A useful approximation to Δ_∞, valid up to an additive constant, can be obtained by replacing T_{I} by $T_{\min}$, thereby extrapolating the continuous approximation (25) down to the shortest orbits. This gives (by also using (7), (72) and (3))

$$\Delta_\infty \approx \frac{1}{2\pi^2}\ln\{\mathrm{e}\,L_{\max}\} = \frac{(N-1)}{2\pi^2}\ln\left\{\frac{1}{\hbar}\left(\frac{e}{T_{\min}}\frac{\mathrm{d}\Omega}{\mathrm{d}E}\right)^{1/(N-1)}\right\}, \tag{79}$$

and shows that for chaotic systems the semiclassical spectral fluctuations increase only logarithmically with $\hbar^{-1}$ and so are much weaker than for integrable systems (cf. equation (31)).

7. Example: Riemann's zeta function

According to the Riemann hypothesis (Edwards 1974), the non-trivial zeros of the function $\zeta(s)$, defined in (46), all lie on the line $\mathrm{Re}\,s = \frac{1}{2}$. It is a natural conjecture, apparently first made by Hilbert and Polya, that the imaginary parts of these zeros are the eigenvalues of a linear operator (for a discussion, see Hejhal 1976). In this section I shall present evidence supporting the view that if this operator is regarded as the Hamiltonian of some (unknown) bound quantum-mechanical system, then in the classical limit the corresponding (unknown) dynamical system has trajectories that are chaotic and without time-reversal symmetry. By applying the semiclassical theory developed in this paper it will then be possible: to explain the local g.u.e. statistics that the Riemann zeros display (as was originally conjectured by Montgomery (1973) and as observed in computations reported by Bohigas & Giannoni (1984) and attributed by them to Odlyzko); to obtain an interesting formula involving prime numbers; and to predict the mean square fluctuations of the Riemann staircase.

Pavlov & Fadeev (1975) (see also Lax & Phillips (1976) and Gutzwiller (1983)) have discovered a scattering system (a leaky surface of constant negative curvature) whose phase shifts are given by the zeta function with Re $s = 1$, and whose classical limit is chaotic. It is not clear what relation, if any, exists between their work and that described here.

The starting point is the formula for $\zeta(s)$ as a product over primes p:

$$\zeta(s) = \prod_p (1-p^{-s})^{-1}. \tag{80}$$

Defining
$$s \equiv \tfrac{1}{2}+iE, \quad \zeta(\tfrac{1}{2}-iE) \equiv D(E), \tag{81}$$

we have
$$\ln D(E) = -\sum_p \ln(1-\exp\{iE\ln p\}/p^{\frac{1}{2}}). \tag{82}$$

Assuming the Riemann hypothesis, and noting that $\ln D$ vanishes as $\operatorname{Im} E \to +\infty$, we see that this function jumps by $-i\pi$ as each Riemann zero $E = E_n$ is traversed from above, so that the oscillatory part of the spectral staircase is

$$\begin{aligned}\mathcal{N}_{\text{osc}}(E) &= -\pi^{-1} \lim_{(\eta\to 0)} \operatorname{Im} \ln D(E+i\eta) \\ &= -\frac{1}{\pi} \operatorname{Im} \sum_p \sum_{m=1}^{\infty} \frac{\exp\{im\ln p\}}{mp^{\frac{1}{2}m}}.\end{aligned} \tag{83}$$

The oscillatory part of the spectral density is therefore

$$d_{\text{osc}}(E) = -\frac{1}{\pi}\sum_p \frac{\ln p[\cos(E\ln p)-p^{-\frac{1}{2}}]}{1-2\cos(E\ln p)/p^{\frac{1}{2}}-p^{-1}} \tag{84}$$

or
$$d_{\text{osc}}(E) = -\frac{1}{2\pi}\sum_p \sum_{m=-\infty}^{\infty}{}' \ln p \exp\{-|m|\ln p/2\}\exp\{iEm\ln p\}, \tag{85}$$

where the prime on the second summation means that $m = 0$ is omitted.

In the form (85) the analogy with the periodic-orbit formula (9) is clear: each primitive orbit corresponds to a prime p and is traversed m times; the action is

$$S_{m,p} = Em\ln p. \tag{86}$$

If E is regarded as the energy, the period is

$$T_{m,p} = \mathrm{d}S/\mathrm{d}E = m\ln p \tag{87}$$

and the amplitude is

$$A_{m,p} = -(\ln p\exp\{-|m|\ln p/2\})/2\pi. \tag{88}$$

Equation (86) shows that in this unknown dynamical system $\hbar^{-1}$ scales with E, so that we can set $\hbar = 1$ and regard $E \to \infty$ as the semiclassical limit. The amplitudes decay exponentially with period, as they must for a chaotic system (in contrast to an integrable one). Moreover the diagonal part $\phi_D(T)$ of the sum (22) is, by the prime number theorem,

$$\begin{aligned}\phi_D(T) &= \sum_{m=1}^{\infty}\sum_p A_{m,p}^2\,\delta(T-T_{m,p}) \\ &= \frac{1}{4\pi^2}\sum_{m=1}^{\infty}\sum_p \frac{\ln^2 p}{p^m}\,\delta(T-m\ln p) \\ &\xrightarrow{(T\to\infty)} \frac{T}{4\pi^2}\sum_{m=1}^{[T/\ln 2]}\exp\{-T(1-m^{-1})\}/m^2 \approx \frac{T}{4\pi^2},\end{aligned} \tag{89}$$

which agrees exactly with the diagonal sum rule (25) for periodic orbits in a chaotic non-degenerate system.

The local average of the Riemann staircase is (Montgomery 1976)

$$\langle \mathcal{N}(E)\rangle = \frac{E}{2\pi}\left(\ln\left\{\frac{E}{2\pi}\right\}-1\right)-\frac{7}{8}, \tag{90}$$

corresponding to the average density

$$\langle d(E)\rangle = \frac{1}{2\pi}\ln\left\{\frac{E}{2\pi}\right\}. \tag{91}$$

The inner energy scale is thus $2\pi/\ln\{E/2\pi\}$, which vanishes in the semiclassical limit $E\to\infty$, in contrast with the outer energy scale $2\pi/T_{\min} = 2\pi/\ln 2$.

Figure 4*a*, *b* illustrates how well the periodic-orbit sum, with (equation (84)) and without (equation (85)) with $|m| = 1$) repetitions, is capable of reproducing δ functions at the Riemann zeros and $-\langle d(E)\rangle$ between them. (I am not claiming that (84) or (85) is an efficient way to calculate the positions of the zeros, because it is not: to roughly locate a zero near E requires about $(E/2\pi)/\ln\{E/2\pi\}$ primes, compared to $(E/2\pi)^{\frac{1}{2}}$ terms of the Riemann–Siegel formula (Edwards 1974), which would locate the same zero much more accurately.)

If on the basis of the identifications (86)–(88) and the diagonal sum (89) the existence of a chaotic non-degenerate dynamical system underlying the Riemann zeros is accepted, the arguments of preceding sections can be applied, and show that the zeros have the spectral rigidity of the g.u.e. In particular, for the spectral form factor (58), the semiclassical analysis gives the limits in (64). These limits agree with what was proved (for $\tau < 1$) and conjectured (for $\tau > 1$) by Montgomery (1973), namely that the Riemann zeros have a $K(\tau)$ that tends to the g.u.e. form (72) as $E\to\infty$.

When expressed in terms of closed orbits (primes and powers of primes) by using (58) and (22) this leads to a formula involving prime numbers that can be tested numerically. The test is most easily made by using not $K(\tau)$ but its integral, which from (72) is expected to be

$$S(\tau) \equiv 2\int_0^{\tau} d\tau'\,K_0(\tau') = \begin{cases}\tau^2 & (\tau \leqslant 1),\\ 2\tau-1 & (\tau \geqslant 1).\end{cases} \tag{92}$$

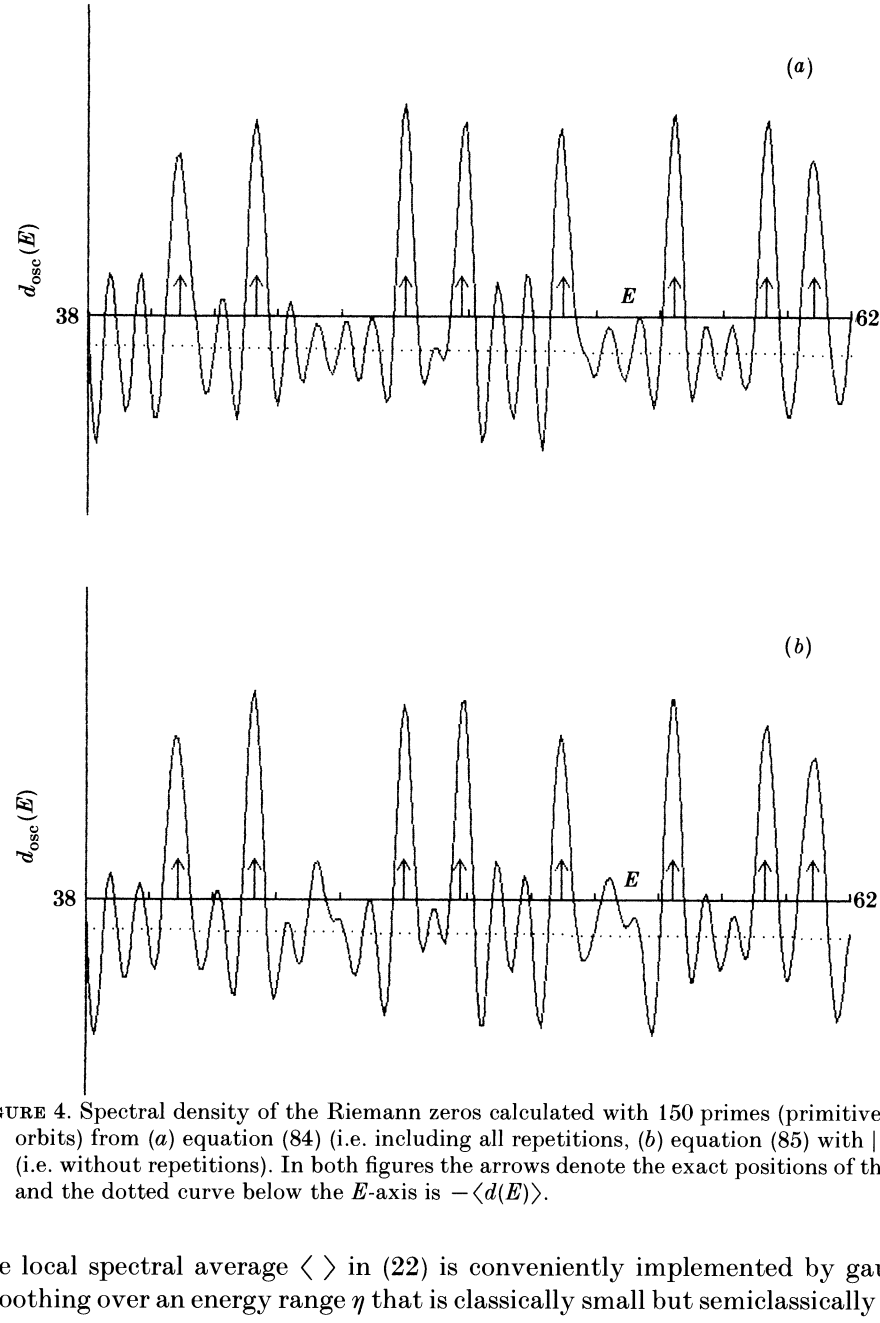

FIGURE 4. Spectral density of the Riemann zeros calculated with 150 primes (primitive closed orbits) from (*a*) equation (84) (i.e. including all repetitions, (*b*) equation (85) with $|m| = 1$ (i.e. without repetitions). In both figures the arrows denote the exact positions of the zeros and the dotted curve below the E-axis is $-\langle d(E)\rangle$.

The local spectral average $\langle\ \rangle$ in (22) is conveniently implemented by gaussian smoothing over an energy range η that is classically small but semiclassically large, i.e.

$$2\pi |\ln 2 \ll \eta \ll E. \tag{93}$$

The closed-orbit formula can now be written down as

$$S(\tau) = \lim_{(E\to\infty)} \frac{2}{(\ln\{E/2\pi\})^2} \sum_{p_1,\, m_1,\, p_2,\, m_2}^{p_1^{m_1} p_2^{m_2} < (E/2\pi)^{2\tau}} \sum \frac{\ln p_1 \ln p_2}{p_1^{\frac{1}{2}m_1} p_2^{\frac{1}{2}m_2}} \cos(E \ln\{p_1^{m_1}/p_2^{m_2}\}) \times \exp(-\eta^2 \ln^2\{p_1^{m_1}/p_2^{m_2}\}). \tag{94}$$

Of particular interest is the 'phase phase-transition' that (92) predicts at $\tau = 1$, where incoherence causes the double sum to depart from its diagonal form τ^2 and adopt the linear form $2\tau - 1$ dictated by the semiclassical sum rule. Figure 5 shows a preliminary numerical test of (94) without repetitions, supporting the view that this phase transition does exist.

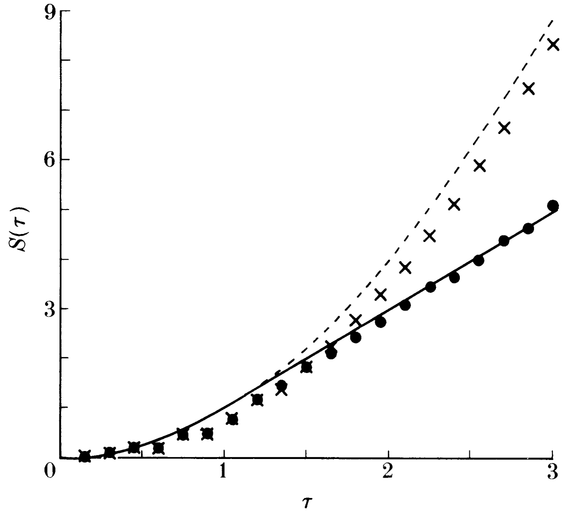

FIGURE 5. Integrated spectral form factor $S(\tau)$ for the Riemann zeros. The circular points are computed from (94) without repetitions, i.e. $m_1 = m_2 = 1$, for $E = 90$, and by using primes $p < 3517$ and a smoothing of $\eta = 4\pi/\ln 2 = 18.3$; the full line is the prediction (92). The crosses show the diagonal sum in (94) and the broken line is the prediction τ^2.

The rigidity $\Delta(L)$ should rise logarithmically according to (70) and (73) until $L \sim L_{\max}$ where, from (7), (87) and (91)

$$L_{\max} = \ln(E/2\pi)/\ln 2. \tag{95}$$

When $L \gg L_{\max}$, the semiclassical analogy predicts that $\Delta(L)$ saturates at a value given by (78) and (88) as

$$\begin{aligned}\Delta_\infty &= \left(\sum_{p<p_{\mathrm{I}}} p^{-1} + \ln\ln\{E/2\pi\} - \ln\ln p_{\mathrm{I}} + 1\right)/2\pi^2 \\ &= (\ln\ln\{E/2\pi\} + 1.2615)/2\pi^2. \end{aligned}\tag{96}$$

This shows that the r.m.s. staircase fluctuations grow as $(\ln\ln E)^{\frac{1}{2}}$ (it has been conjectured (Montgomery 1976, 1977) that the largest fluctuations grow as $(\ln E/\ln\ln E)^{\frac{1}{2}}$).

The saturation fluctuations predicted by (96) grow very slowly: for $E = 10^3$, $\Delta_\infty = 0.146$ (and $L_{\max} = 7.3$); for $E = 10^6$, $\Delta_\infty = 0.190$ (and $L_{\max} = 17.3$). To reach $\Delta_\infty = 0.5$ requires $E = 10^{2379}$, and to reach $\Delta_\infty = 1$ – an r.m.s. staircase deviation from $\langle \mathcal{N}(E)\rangle$ of only one level – requires $E = 10^x$ where $x \approx 5\times 10^7$. This slow growth, indicating very slow approach to the semiclassical limit, would seem to rule out any direct test of (96).

8. Chaotic systems with time-reversal symmetry

When there is time-reversal symmetry, every orbit that is not self-retracing must be combined coherently with its time-reverse in the sum (22), because both orbits have the same action and period. The counterpart of the chaotic classical diagonal sum rule (25) must incorporate this coherence. Thus instead of $A_j^2+A_j^2 = 2A_j^2$ we have $(A_j+A_j)^2 = 4A_j^2$, so the right-hand side of (25) must be multiplied by two. This means that the form factor $K(\tau)$ defined by (58) has the limiting form 2τ, rather than τ, when $\tau_{\min} \ll \tau \ll 1$. The semiclassical sum rule is, however, unaffected by the symmetry, so that instead of (6) the conditions on $K(\tau)$ are

$$K(\tau) \rightarrow \begin{cases} 2\tau & (\tau_{\min} \ll \tau \ll 1), \\ 1 & (\tau \gg 1). \end{cases} \tag{97}$$

The rigidity $\Delta(L)$ is given by (65) with this new form factor, and the argument parallels that in §6.

When $L \ll 1$, the limit $\frac{1}{15}L$ is regained, and is as before a consequence of the semiclassical sum rule.

When $1 \ll L \ll L_{\max}$, the analogues of (67), (70) and (71) are

$$\begin{aligned} \Delta(L) &= \frac{1}{\pi^2}\left[\int_0^Y \frac{\mathrm{d}y}{y}\, G(y) + \frac{1}{2}\int_Y^\infty \frac{\mathrm{d}y}{y}\,\frac{K(y/\pi L)}{y/\pi L}\right] \\ &= \frac{1}{\pi^2}\ln L + E \end{aligned} \tag{98}$$

with $$E = \frac{1}{\pi^2}\left[\ln 2\pi + \gamma - \frac{9}{4} - \frac{1}{2}\int_0^\infty \mathrm{d}\tau \ln\tau \,\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{K(\tau)}{\tau}\right)\right]. \tag{99}$$

Equation (98) is precisely the asymptotic rigidity of the gaussian orthogonal ensemble (g.o.e.) of random-matrix theory (Dyson & Mehta 1963), that is $\Delta(L)$ averaged over the spectra of large real symmetric matrices whose elements are gaussian random variables with statistics invariant under orthogonal transformations. The logarithmic dependence and correct prefactor again follow from the Hannay & Ozorio de Almeida (1984) sum rule when this is modified to include the orbital degeneracy.

Without specifying $K(\tau)$ more closely than (97), the additive constant E cannot be determined and, as in §6, there is no obvious semiclassical way of doing this. Moreover, the simplest interpolation, analogous to (72), namely

$$K_1(\tau) = \begin{cases} 2\tau & (\tau \leqslant \frac{1}{2}), \\ 1 & (\tau \geqslant \frac{1}{2}), \end{cases} \tag{100}$$

gives $$E = 0.067, \tag{101}$$

which is not the correct g.o.e. constant. To obtain this, it is necessary to use the correct form factor, which is (Mehta 1967)

$$K_0(\tau) = \begin{cases} 2\tau - \tau \ln\{1+2\tau\} & (\tau \leqslant 1), \\ 2 - \tau \ln\left\{\dfrac{1+2\tau}{2\tau-1}\right\} & (\tau \geqslant 1), \end{cases} \tag{102}$$

which gives $$E = (\ln 2\pi + \gamma - \tfrac{5}{4} - \tfrac{1}{8}\pi^2)/\pi^2 = -0.00695. \tag{103}$$

For comparison, the smooth interpolation

$$K_2(\tau) = 2\tau/(1+2\tau) \tag{104}$$

gives $$E = -0.054. \tag{105}$$

The 'correct' form factor (102) is also discontinuous, but in contrast to (72) the discontinuity is only in the third derivative. Therefore, the sum (22) for $\phi(t)$ also has a 'semiclassical phase phase-transition' when there is time-reversal symmetry, with a higher order than the phase transition when there is no such symmetry.

The result (98) explains the local universality observed in many numerical experiments on chaotic systems with time-reversal symmetry (see the review by Bohigas & Giannoni 1984), which holds for $L \ll L_{\max}$. When $L \gg L_{\max}$, the rigidity saturates non-universally, at a value approximately given by the analogue of (79), namely

$$\begin{aligned}\Delta_\infty &= \pi^{-2} \ln\{\mathrm{e}L_{\max}\} - \tfrac{1}{8} \\ &= \frac{(N-1)}{\pi^2} \ln\left\{\frac{1}{\hbar}\left(\frac{\mathrm{e}}{T_{\min}}\frac{\mathrm{d}\Omega}{\mathrm{d}E}\right)^{1/(N-1)}\right\} - \frac{1}{8}.\end{aligned} \tag{106}$$

9. Discussion

The study reported here suggests a number of questions and directions for further investigation.

We have considered only second-order statistics ($\Delta(L)$ and $K(\tau)$). It is natural to enquire about higher statistics such as many-level distributions and the distribution of spacings between neighbouring levels. Direct generalization of the methods used here would involve multiple sums analogous to (22), and diagonal and partially diagonal analogues of the sum rules (24), (25) and (56). At present it is not clear how this could be accomplished.

Even the rigidity has been studied only in the integrable and chaotic extremes. In view of the recent interest in spectral statistics of systems that show a transition to chaos as a parameter is varied (Robnik 1984, Seligman *et al.* 1985; Meyer *et al.* 1984; Berry & Robnik 1984), it is desirable to extend the theory to cover these more general cases. In the range $L \ll L_{\max}$, it follows as a rough approximation from the periodic orbit theory that $\Delta(L)$ is the sum of two contributions, one from the isolated unstable orbits in the chaotic region of phase space and one from the closed orbits in the region where there are tori (because of the finite resolution imposed by $\hbar$, orbits in this region can be considered as filling resonant 'tori' even though in reality they are isolated with near-marginal stability). The conjecture that $\Delta(L)$ will be approximately additive has also been made by Seligman *et al.* (1985). As L increases, the contribution of the orbits in the integrable component should increase in relative importance, and ought to completely dominate fluctuations in the saturation régime $L \gg L_{\max}$.

Another interesting case is that of non-integrable systems with non-isolated periodic orbits. One important example is the stadium billiard of Bunimovich

(1974) (see also Berry 1981*a*), a chaotic system in which all periodic orbits are isolated with the exception of those bouncing perpendicularly between the straight sides, which form a one-parameter family. This single orbit (and its repetitions) will not spoil the logarithmic universality of $\Delta(L)$ for $L \ll L_{\text{max}}$ because the orbit selection function $G(y)$ (figure 1) will eliminate it. But its contribution will increase rapidly with L and it will dominate the saturation régime $L \gg L_{\text{max}}$; arguments from §4 then give, for a stadium whose parallel sides have length a and are separated by b,

$$\Delta_{\infty} = (2\pi E)^{\frac{1}{2}} b^2 \zeta(3)/8\hbar\pi^3 a \tag{107}$$

(the factor $\zeta(3)$ incorporates repetitions). Analogous behaviour is predicted for the Sinai billiard (Berry 1981*b*), for which there are non-isolated orbits whose finite number depends on the radius of the central disc. The emergence and dominance of particular non-isolated orbits as L increases through L_{max} is a good example of transition from universal to non-universal spectral behaviour.

Another example, whose theoretical treatment is less clear, is billiards in irrational-angled polygons. These systems have been conjectured to be ergodic (Hobson 1976), but they have zero Kolmogorov entropy (Sinai 1976) and so are not chaotic. This behaviour stems from their closed orbits, which are almost all non-isolated and moreover marginally stable. The arguments of this paper strongly suggest that the spectral fluctuations should be much stronger than for the chaotic systems considered in §§6, 8, i.e. $\Delta(L)$ should rise faster than logarithmically, in spite of a numerical study of the level spacings distribution of triangles (Berry & Wilkinson 1984) suggesting g.o.e. behaviour.

It is at first sight surprising that for the chaotic systems considered in §§6, 8 the Kolmogorov entropy S did not appear in the rigidity formulae. S is a non-universal quantity and so might be expected to contribute to $\Delta(L)$ when $L \sim L_{\text{max}}$. However, S^{-1}, being the time for the separation between two nearby orbits to grow by a factor e, is of the same order of magnitude as T_{min} (for billiards the ratio is a geometrical factor), which does appear in the formulae (cf. (79) and (106)). Of course there remains the possibility that S might contribute directly to a higher-order statistic.

Another problem is the derivation of the semiclassical sum rule (56) directly from the definition (22). The derivation in §5 was based on (53), which is the condition for the spectral density to contain the correct density of singularities. A direct derivation from (22) would require detailed knowledge, at present lacking, of correlations between the actions of long orbits, which conspire in the off-diagonal terms of the sum to cancel the growth in the diagonal terms and cause $\phi(T)$ to saturate at the value (56). Such knowledge might also explain the 'phase phase-transitions' in the form factors (72) and (102).

The semiclassical sum rule is only the first in an infinite hierarchy of similar relations, obtained as the result of generalizing (53) by expressing a δ function as the limit of its analytic continuation raised to a power, namely

$$\langle d(E) \rangle = \lim_{\eta \to 0} \frac{(4\pi\eta)^{l-1} \Gamma^2(l)}{\Gamma(2l-1)} \langle d^l_{\text{osc},\eta}(E) \rangle. \tag{108}$$

These higher-order sum rules might help in understanding higher-order spectral statistics. Their existence is related to a more fundamental and very remarkable 'bootstrap' property of the periodic-orbit sum (9): as a consequence of the spectrum being determined by the singularities of d_{osc}, which in turn are generated by very long classical periodic orbits, any finite number of terms may be deleted from (9) without destroying the spectral information it contains.

I thank Dr J. Hannay and Dr M. Robnik for helpful discussions.

Note added in proof (4 *June* 1985). Dr A. Voros has pointed out to me that (80), on which the closed-orbit sum (85) for the density of Riemann zeros is based, does not converge if $\text{Re}\, s < 1$. Therefore the ability of (85) to discriminate individual zeros, as illustrated in figure 4, might deteriorate as E increases, and could fail altogether when $E > 2\pi \exp(4\pi) \sim 2 \times 10^6$.

References

Balian, R. & Bloch, C. 1972 *Ann. Phys.* **69**, 76–160.
Balian, R. & Bloch, C. 1974 *Ann. Phys.* **85**, 514–545.
Berry, M. V. 1981*a* *Eur. J. Phys.* **2**, 91–102.
Berry, M. V. 1981*b* *Ann. Phys.* **131**, 163–216.
Berry, M. V. 1983 Semiclassical mechanics of regular and irregular motion. In *Chaotic behavior of deterministic systems* (Les Houches Lectures, vol. XXXVI, ed. G. Iooss, R. H. G. Helleman & R. Stora, pp. 171–271. Amsterdam: North-Holland.
Berry, M. V. 1984 Structures in semiclassical spectra: a question of scale. In *The wave–particle dualism* (ed. S. Diner, D. Fargue, G. Lochak & F. Selleri), pp. 231–252. Dordrecht: D. Reidel.
Berry, M. V. & Tabor, M. 1976 *Proc. R. Soc. Lond.* A **349**, 101–123.
Berry, M. V. & Tabor, M. 1977*a* *Proc. R. Soc. Lond.* A **356**, 375–394.
Berry, M. G. & Tabor, M. 1977*b* *J. Phys.* A **10**, 371–379.
Berry, M. V. & Robnik, M. 1984 *J. Phys.* A **17**, 2413–2421.
Berry, M. V. & Wilkinson, M. 1984 *Proc. R. Soc. Lond.* A **392**, 15–43.
Bohigas, O. & Giannoni, M. J. 1984 Chaotic motion and random-matrix theories. In *Mathematical and computational methods in nuclear physics* (ed. J. S. Dehesa, J. M. G. Gomez & A. Polls). *Lecture Notes in Physics* vol. 209, pp. 1–99. New York: Springer-Verlag.
Bunimovich, L. A. 1974 *Funct. Anal. Appl.* **8**, 254–255.
Casati, G., Chirikov, B. V. & Guarneri, I. 1985 *Phys. Rev. Lett.* **54**, 1350–1353.
Dyson, F. J. & Mehta, M. L. 1963 *J. Math. Phys.* **4**, 701–712.
Edwards, H. M. 1974 *Riemann's Zeta Function.* New York and London: Academic Press.
Gutzwiller, M. C. 1967 *J. Math. Phys.* **8**, 1979–2000.
Gutzwiller, M. C. 1969 *J. Math. Phys.* **10**, 1004–1020.
Gutzwiller, M. C. 1970 *J. Math. Phys.* **11**, 1791–1806.
Gutzwiller, M. C. 1971 *J. Math. Phys.* **12**, 343–358.
Gutzwiller, M. C. 1978 In *Path integrals and their applications in quantum, statistical and solid-state physics* (ed. G. J. Papadopoulos & J. T. Devreese), pp. 163–200. New York: Plenum.
Gutzwiller, M. C. 1980 *Phys. Rev. Lett.* **45**, 150–153.
Gutzwiller, M. C. 1983 *Physica* **7**D, 341–355.
Hannay, J. H. & Ozorio de Almeida, A. M. 1984 *J. Phys. A* **17**, 3429–3440.
Hejhal, D. A. 1976 *Duke math. J.* **43**, 441–482.
Hobson, A. 1976 *J. Math. Phys.* **16**, 2210–2214.
Lax, P. D. & Phillips, R. S. 1976 *Scattering theory for automorphic functions.* Princeton University Press.
McKean, H. P. 1972 *Communs pure. appl. Math.* **25**, 225–246.

Mehta, M. L. 1967 *Random matrices and the statistical theory of energy levels.* New York and London: Academic Press.
Meyer, H.-D., Haller, E., Köppel, H. & Cederbaum, L. S. *J. Phys.* A **17**, L831–836.
Montgomery, H. L. 1973 *Proc. Symp. pure Math.* **24**, 181–193.
Montgomery, H. L. 1976 *Proc. Symp. pure Math.* **38**, 307–310.
Montgomery, H. L. 1977 *Communs Math. Helv.* **52**, 511–523.
Norcliffe, A. & Percival, I. C. 1968 *J. Phys.* B **1**, 774–83.
Parry, W. 1984 *Ergod. Th. Dynam. Syst.* **4**, 117–134.
Parry, W. & Pollicott, M. 1983 *Ann. of Math* **118**, 573–591.
Pavlov, B. S. & Fadeev, L. D. 1975 *Soviet Math.* **3**, 522–548.
Pechukas, P. 1983 *Phys. Rev. Lett.* **51**, 943–946.
Porter, C. E. 1965 *Statistical theories of spectra: fluctuations.* New York: Academic Press.
Richens, P. J. & Berry, M. V. 1981 *Physica* **1***D*, 495–512.
Robnik, M. 1984 *J. Phys.* A **17**, 1049–1074.
Seligman, T. H. & Verbaarschot, J. J. M. 1985 *Phys. Lett.* A **108**, 183–187.
Seligman, T. H., Verbaarschot, J. J. M. & Zirnbauer, M. R. 1985 *J. Phys.* A (In the press.)
Sinai, Ya. G. 1976 *Introduction to ergodic theory.* Princeton University Press.
Voros, A. 1983 *Ann. Inst. H. Poincaré* **39**, 211–338.
Zucker, I. J. 1974 *J. Phys.* A **7**, 1568–1575.