RIEMANN'S ZETA FUNCTION: A MODEL FOR QUANTUM CHAOS?

M.V.Berry
H.H.Wills Physics Laboratory, Tyndall Avenue,
Britol BS8 1TL, U.K.

1 INTRODUCTION

The celebrated hypothesis of Riemann[1] is that all the complex zeros of his function $\zeta(z)$ have real part 1/2, so that the quantities $\{E_j\}$ defined by

$$\zeta\left(\tfrac{1}{2} - iE_j\right) = 0 \tag{1}$$

are all real. There is evidence supporting the hypothesis: the first few million E_j have been computed and are all real, and it has been proved that uncountably many E_j are real. My purpose in this speculative paper is to extend the old suggestion that the E_j are real because they are eigenvalues of some Hermitian operator $\hat{H}$. The extensions are that if $\hat{H}$ is regarded as the Hamiltonian of a quantum-mechanical system then

(i) $\hat{H}$ has a classical limit

(ii) the classical orbits are all chaotic (unstable)

(iii) the classical orbits do not possess time-reversal symmetry.

To make these assertions plausible I will combine two sorts of evidence. The first (section 2) concerns largely numerical results connecting $\{E_j\}$ with the spectra of infinite random complex Hermitian matrices. The second (section 3) concerns analogies between a (divergent) representation of the number of Riemann zeros with $0<E_j<E$ and an asymptotic formula expressing the number of quantum energy levels in any given interval as a sum over classical closed orbits. Finally (section 4), I will give a semiclassical interpretation of the Riemann-Siegel formula (the basis of a powerful method for computing the $\{E_j\}$[1]) leading to a conjectured generalization that would be a quantization formula for classically chaotic systems.

The most useful formulae for $\zeta(z)$ will be the familiar ones: as a product over primes p or a sum of inverse powers of integers n, namely

$$\zeta(z) = \prod_p \left(1 - p^{-z}\right)^{-1} \qquad (\mathrm{Re}\ z > 1) \tag{2a}$$

$$= \sum_{n=1}^{\infty} n^{-z} \qquad (\mathrm{Re}\, z > 1) \tag{2b}$$

even though neither representation converges on the line Rez=1/2 where the zeros are. Of course there are many representations which are valid for Rez=1/2, such as this resummation of (2b):

$$\zeta(z) = \frac{1}{1-2^{1-z}} \sum_{n=1}^{\infty} (-1)^{n+1} n^{-z} \qquad (\mathrm{Re}\, z > 0) \tag{3}$$

The symmetry [1] relating $\zeta(z)$ and $\zeta(1-z)$ implies that each of the zeros (1) with real E_j has a counterpart with $-E_j$, and with this understanding we henceforth regard $\{E_j\}$ as a set of positive numbers with $E_1<E_2<E_3\ldots$.

2 RIEMANN ZEROS AND RANDOM MATRICES

A useful separation of the $\{E_j\}$ into an average part and a fluctuating part can be achieved by means of the <u>Riemann staircase.</u> This is defined as

$$N_R(E) \equiv \sum_{j=1}^{\infty} \Theta(E-E_j) \tag{4}$$

where Θ denotes the unit step function. $N_R(E)$ is simply the number of zeros with $E_j<E$. The average $\langle N_R(E)\rangle$ is a smooth approximation to the staircase, whose form is known [1] to be

$$\langle N_R(E)\rangle = \frac{E}{2\pi}\left(\ln\left\{\frac{E}{2\pi}\right\} - 1\right) + \frac{7}{8} . \tag{5}$$

Fig.1a shows just how close an approximation this is, even for small E. Figs 1b,c show that it is necessary to include the term 7/8.

Deviations from $\langle N_R(E)\rangle$ constitute the fluctuations in $\{E_j\}$. The statistics of these fluctuations can be studied numerically, and it is found that with high accuracy they coincide with the statistics of the eigenvalues of a typical member of the 'Gaussian unitary ensemble' (GUE)[2] of complex Hermitian matrices whose elements are Gauss -distributed in a way that is invariant under unitary transformations.

One such statistic is the probability distribution of the normalized spacings $\{S_j\}$ between adjacent zeros; these are defined by

$$S_j \equiv (E_{j+1}-E_j)\Big/\langle d_R((E_j+E_{j+1})/2)\rangle \tag{6}$$

where $\langle d_R(E)\rangle$ is the average density of zeros, given by (5) as

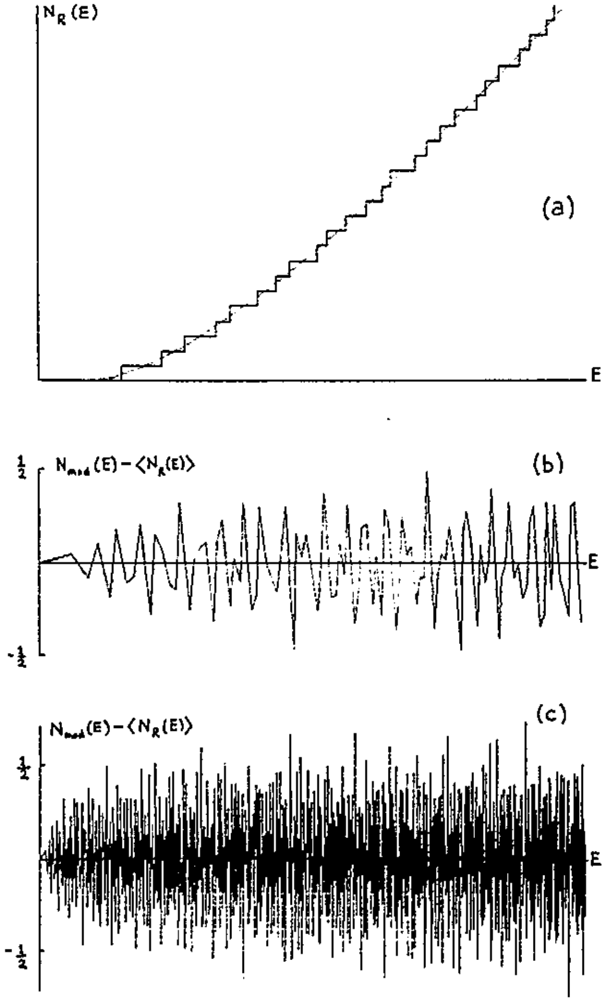

Fig.1 a) Riemann staircase $N_R(E)$, and (dotted) its average $\langle N_R(E)\rangle$, for the lowest 25 zeros; b) deviation $N_{mod}(E)-\langle N_R(E)\rangle$ for the lowest 100 zeros; c) $N_{mod}(E)-\langle N_R(E)\rangle$ for the lowest 1000 zeros ($N_{mod}(E)$ is $N_R(E)$ made continuous by replacing the N'th step by the straight line joining E_N, $N-1/2$ and E_{N+1}, $N+1/2$).

$$\langle d_R(E)\rangle = \frac{d}{dE}\langle N_R(E)\rangle = \frac{1}{2\pi}\ln\left\{\frac{E}{2\pi}\right\} \tag{7}$$

The spacings distribution $P(S)$ is shown in fig.2, together with $P_{GUE}(S)$, for which random-matrix theory [2] gives the close approximation

$$P_{GUE}(S) = \frac{32}{\pi^2}S^2\exp\left\{-4S^2/\pi\right\}. \tag{8}$$

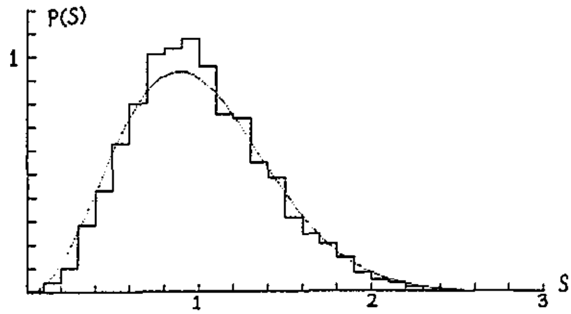

Fig.2 Histogram of distribution P(S) for the first 5000 Riemann zeros, together with $P_{GUE}(S)$ dotted.

Far more extensive computations have been performed by Odlyzko; these are unpublished, but preliminary reports have been given by Dyson [3] and Bohigas and Giannoni [4]. Odlyzko calculated sequences of up to 10^5 Riemann zeros, reaching to the 10^{11}th, and studied not only P(S) but also correlation functions betwen pairs, triplets and quartets of zeros. Apart from one apparent exception, to which we will return in section 3, his statistics are in excellent agreement with those of the GUE.

A theorem of Montgomery [5] supports the conclusion suggested by Odlyzko's results, that the statistics of Riemann zeros are precisely those of the GUE. The theorem concerns the form factor $K(\tau)$ of the zeros; this is the Fourier transform of the pair correlation function, and is defined as

$$K(\tau) = \lim_{M\to\infty}\left\{\frac{1}{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\exp\left\{2\pi i\tau(x_j - x_k)\right\} - \frac{\sin M\pi\tau}{\pi\tau}\right\} \tag{9}$$

where $\{x_j\}$ are the Riemann zeros, scaled so as to have unit mean spacing, that is

$$x_j \equiv \langle N_R(E_j)\rangle . \tag{10}$$

Montgomery proves that for $|\tau|<1$, $K(\tau)$ coincides with the GUE form factor [6]

$$\left.\begin{aligned} K_{GUE}(\tau) &= |\tau| \qquad (|\tau|<1) \\ &= 1 \qquad (|\tau|>1) \end{aligned}\right\}, \tag{11}$$

and conjectures that this agreement continues to hold when $|\tau|>1$.

Now, GUE statistics are distinctive [4]: they differ sharply, for example, from the Poisson statistics of numbers generated sequentially by a random process, and from the Gaussian orthogonal ensemble (GOE) statistics of eigenvalues of random real symmetric (as opposed to complex Hermitian) matrices. It would seem difficult to simulate GUE statistics by a process not involving eigenvalues of complex Hermitian matrices, so from now on I will regard the experimental observation that the Riemann zeros obey GUE statistics as evidence that there is indeed a nontrivial complex infinite Hermitian matrix $\hat{H}$ with eigenvalues E_j. ('Nontrivial' means that $\hat{H}$ must not contain the $\{E_j\}$ explicitly - as in the diagonal matrix $E_i \delta_{ij}$ or any simple unitary transform thereof.)

Among the class of operators $\hat{H}$ which are quantum Hamiltonians with classical limits, those with discrete energy spectra obeying GUE statistics have classical orbits which are chaotic and without time-reversal symmetry. This has been demonstrated numerically by Seligman and Verbaarschot [7], Berry and Robnik [8], and Robnik and Berry [9], and explained theoretically by Berry [10]. Both of the above conditions are necessary for GUE statistics: for systems whose orbits are not chaotic but integrable (with or without time-reversal symmetry), Berry and Tabor [11] showed the energies to be Poisson-distributed; and for systems which are chaotic but which do have time-reversal symmetry, Bohigas, Giannoni and Schmit [12] and others [13,4,10] have shown the energies to be GOE-distributed.

Of course the observation that GUE statistics are shared by both the Riemann zeros E_j and the eigenvalues of classically chaotic systems without time-reversal symmetry does not imply that the $\{E_j\}$ come from a classically chaotic $\hat{H}$, but we now turn to analytical evidence which strongly suggests that they do.

3 OUROBOROLOGY

Ouroboros [14] was the mythical snake that swallowed its tail, and serves to symbolize the constructive interference of quantum waves associated with classical orbits (fig.3). Such interference forms the basis of a technique for generating the energy spectrum, developed by Gutzwiller [15-17] and Balian and Bloch [18]; for an elementary review, see [19]. The technique has been used to explain why classically integrable systems have Poisson-distributed energy levels

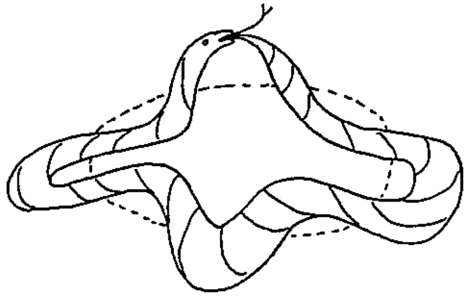

Fig.3 A quantum wave (Ouroboros) interfering constructively round a classical closed orbit.

[11] and why chaotic systems have GOE- or GUE- distributed levels [10].

Ouroborology is based on representing the spectral density as the imaginary part of the trace of the Fourier transform of the propagator (Green function of the time-dependent Schrödinger equation), which is expressed semiclassically (i.e. for small Planck's constant $\hbar$) as a sum over classical paths. The spectral staircase N(E), defined by (4) with $\{E_j\}$ now being the energy levels, is just the integral of the spectral density, so N(E) can similarly be expressed in terms of closed classical paths with energy E, the relation being that the fluctuating part of the staircase is

$$N_{osc}(E) \equiv N(E) - \langle N(E) \rangle = \mathrm{Im} \sum_p \sum_{m=1}^{\infty} B_{pm} \exp\{i S_{pm}(E)/\hbar + \phi_{pm}\} . \tag{12}$$

In this formula, $\langle N(E) \rangle$ is the average staircase (cf (5)) and will be discussed later. The double sum is over all closed orbits, each being an m-fold traversal of a primitive orbit labelled p. S_{pm} is the action of the orbit m, given in terms of the canonical phase-space variables q_μ, p_μ by

$$S_{pm}(E) = \oint p_\mu dq_\mu \quad \left(= m\, S_{p1}(E)\right). \tag{13}$$

The phases ϕ_{pm} and amplitudes B_{pm} depend on the focusing and stability of a bundle of (non-closed) orbits centred on the closed one. We require only the formulae for systems with two freedoms whose closed orbits are all isolated and unstable (making the dynamics chaotic) and without focal points. Then $\phi_{pm}=0$, and [15,16]

$$B_{pm} = \left[2\pi m \sinh\{m\lambda_p(E)/2\}\right]^{-1}, \tag{14}$$

where $\lambda_p(E)(>0)$ is the instability exponent of the primitive orbit p at energy E (i.e. $\exp\{\pm\lambda_p\}$ are the eigenvalues of the 2x2 matrix M_p of linearized phase-space deviations transverse to p, and

$$2\sinh\{m\lambda_p/2\} = \left[-\det\{M_p^m - 1\}\right]^{1/2}.$$

Thus for such systems the semiclassical asymptotic formula for the spectral fluctuations is

$$N_{osc}(E) = \frac{1}{2\pi}\sum_p \sum_{m=1}^{\infty} \frac{\sin\{m S_{p1}(E)/\hbar\}}{m \sinh\{m\lambda_p(E)/2\}}. \tag{15}$$

Next, we note that $N_{osc}(E)$, as defined by (4) and the first member of (12), has unit discontinuities at each eigenvalue. Therefore the closed-orbit sum (15) can at best be conditionally convergent, with the discontinuities determined by the very long orbits. For these, $m\lambda_p/2$ is large, and so we can replace the series by

$$N_{osc}(E) \approx \frac{1}{\pi}\sum_p \sum_{m=1}^{\infty} \frac{1}{m}\exp\{-m\lambda_p(E)/2\}\sin\{m S_{p1}(E)/\hbar\}. \tag{16}$$

This replacement will be reconsidered later.

Let us now turn to $\zeta(z)$ with $z=\frac{1}{2}-iE$. Just above the real E axis, the phase of decreases by as ReE passes each Riemann E_j. Moreover $\zeta(z)\to 1$ as $\mathrm{Im}(E)\to+\infty$ (i.e. as $\mathrm{Re}z\to+\infty$). Therefore the fluctuating part of the Riemann staircase is

$$N_{R,osc}(E) \equiv N_R(E) - \langle N_R(E)\rangle = -\frac{1}{\pi}\lim_{\eta\to 0} \mathrm{Im}\ln\zeta\left(\tfrac{1}{2} - i(E+i\eta)\right). \tag{17}$$

Now pretend that the product formula (2a) can be used when Rez=1/2 (more about this later), substitute into (17) and expand the logarithms. This gives

$$N_{R,osc}(E) \approx -\frac{1}{\pi}\sum_p \sum_{m=1}^{\infty} \frac{\sin\{mE\ln p\}}{m\,p^{m/2}} \tag{18}$$

which apart from a sign, to be discussed later, has the same form as the semiclassical expression (16) if the following identifications are made: the label p for primitive closed orbits denotes prime numbers; the actions of the closed orbits are

$$S_{pm} = mE\ln p\ ; \tag{19}$$

Planck's constant $\hbar$ is unity, so that the semiclassical limit is $E\to\infty$; and the instability exponents (independent of E) are

$$\lambda_p = \ln p\,. \tag{20}$$

It is worth remarking that it follows from (19) that the periods of the closed orbits would be

$$T_{p^m} = \frac{dS_{p^m}}{dE} = m \ln p = \ln p^m. \tag{21}$$

There is thus a formal analogy between fluctuations of the Riemann staircase and fluctuations of the spectral staircase of a classically chaotic system. Because of the distinctive GUE statistics (and for other reasons [10]) the classical orbits must lack time-reversal symmetry. Mathematicians have noticed essentially the same analogy as that between (15) or (16) and (18), in the context of a special case for which a certain transform of (15) is exact rather than asymptotic. This is the Selberg trace formula, which equates a sum over eigenvalues of the Laplace-Beltrami operator (playing the role of $\hat{H}$)on a manifold of constant negative curvature to a sum over closed geodesics on this manifold (all unstable, i.e. chaotic). But despite extensive study (see McKean [20] and Hejhal [21]) this analogy has not led to the identification of the mysterious 'Riemann' classical system with the properties (19-21). The 'closest approach' has been the discovery by Pavlov and Faddeev [22] and Gutzwiller [23] of a scattering (rather than bound) system whose phaseshifts (rather than energy levels) are given by $\zeta(z)$ with Rez=1 (rather than 1/2).

An apparently anomalous outcome [3] of Odlyzko's computation of the form factor $K(\tau)$(eq.9) of the Riemann zeros gives further support to the semiclassical analogy. Although he finds good overall agreement with the GUE formula (11), close examination of the difference $K-K_{GUE}$ reveals a series of spikes for small τ . Such spikes are predicted by the semiclassical theory, because as I have shown elsewhere [10] the universality of the GUE statistics ceases to hold for large energy scales, that is short time scales. For $K(\tau)$ the semiclassicl formula for $|\tau|<1$, expressed in terms of the closed-orbit amplitudes and periods, is [10]

$$K(\tau) = \pi^2 \tau^2 \sum_p \sum_{m=1}^{\infty} B_{pm}^2 \, \delta(\tau - T_{pm}/2\pi\hbar\langle d\rangle). \tag{22}$$

Now as $\hbar\to 0$, $\hbar\langle d\rangle\sim\hbar^{-(D-1)}$ for a system with D freedoms (D>1), so that the spike associated with a given orbit slides towards $\tau=0$ as $\hbar\to 0$. Near any finite τ , then, spikes are semiclassically thickly clustered and it is their average [10] which gives $K\approx|\tau|$ as in (11). But an accurate evaluation of $K(\tau)$ should reveal at least the first few spikes. In the Riemann case, (21) shows that the spikes should occur at τ-values proportional to logarithms of powers of primes. If the first of Odlyzko's spikes occurs at $\tau = K\ln 2$, the others occur at positions which on his picture are indistinguishable from $K\ln 3$, $K\ln 4$, $K\ln 5$ and $K\ln 7$, but as expected there is no spike at $K\ln 6$ because 6 is not a power of a prime.

Four objections may be raised against the chaos analogy for the Riemann zeros.

Objection 1: the Riemann closed-orbit formula (18) depends on the product (2a) which does not converge when $\mathrm{Re}z=1/2$. But the analogous semiclassical formulae (15) and (16) almost certainly do not converge either. The physical reason for this is that the number of closed orbits of a chaotic system proliferates exponentially as their period (or action) increases, overwhelming the effect of the instability exponents λ_p in making the amplitudes decay exponentially. The mathematical reason was explained to me by Dr.A.Voros in the context of the Selberg trace formula: the fundamental quantum object for which semiclassical techniques give an expression in terms of closed orbits is not $N_{osc}(E)$ (or its derivative which is the fluctuating part of the spectral density) but the trace of the resolvent

$$g(E) \equiv \sum_j \frac{1}{E - E_j} \tag{23}$$

for which the closed-orbit formula makes sense only when $\eta \equiv \mathrm{Im}E$ exceeds some finite value. This means that in the formula

$$N(E) = -\frac{1}{\pi} \lim_{\eta \to 0} \mathrm{Im} \int_0^{E+i\eta} dE' g(E') \tag{24}$$

the limit $\eta \to 0$ cannot be taken when using ourolorology for g(E). Retaining finite η has the effect of introducing further factors $\exp[-i\eta T_{pm}(E)/\hbar]$ into (15) and (16), making the sums converge but at the price of giving a staircase whose steps are smoothed by η, thereby frustrating attempts to discriminate individual eigenvalues. Objection 1 therefore disappears because both the Riemann and semiclassical closed-orbit formulae share the disadvantage of not converging for real E. One could argue that this shared disadvantage strengthens the analogy.

It is interesting to look numerically at the divergence of the product (2a), especially in view of earlier computations [10] of the spectral density (derivative of (18) which with small numbers of primes showed pronounced peaks at the lowest few zeros, nicely simulating the delta-functions that the exact spectral density must possess. It is simplest to calculate the truncated product

$$\left|\zeta_M\left(\tfrac{1}{2}-iE\right)\right| \equiv \prod_{p<M}\left|1-\frac{e^{iE\ln p}}{p^{1/2}}\right|^{-1}. \tag{25}$$

As fig.4a shows, very few factors suffice to discriminate the lowest zeros. As M increases, however, $|\zeta_M|$ oscillates increasingly fast between the zeros, in contrast to the exact $|\zeta|$ which has only one maximum between each pair of zeros (cf fig.6 later); and when M=10000 (fig.4b) the oscillations are threatening to obscure the first zero. That such

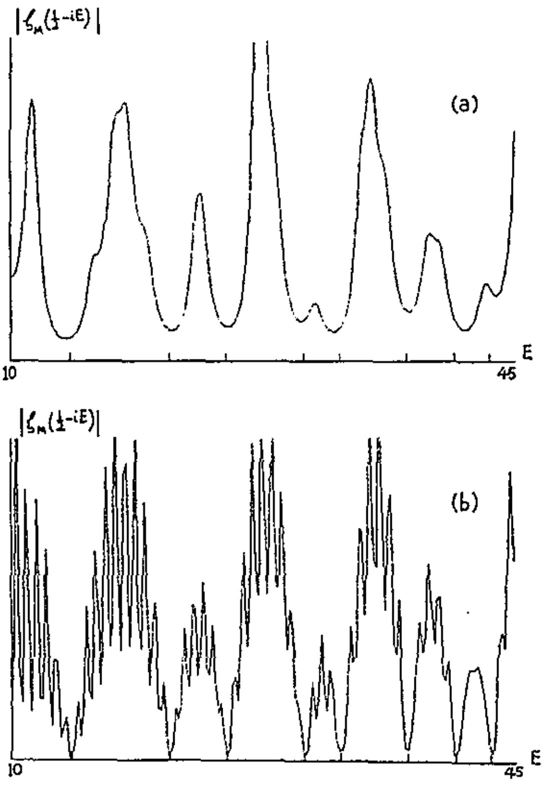

Fig.4a Truncated Riemann product $\zeta_M(\frac{1}{2}-iE)$ as a function of E for M=5 (three factors in (25)), with ticks marking the exact Riemann zeros E_j; b) as a) but with M=10000

obscuration will eventually occur is illustrated in fig.5, which shows $|\zeta_M|$ as a function of M evaluated at the exact positon of the first Riemann zero E_1=14.135... : at first $|\zeta_M|$ decreases, apparently indicating convergence onto E_1, but, when M exceeds about 2000, $|\zeta_M|$ begins to oscilalte with increasing amplitude. Rough asymptotics shows

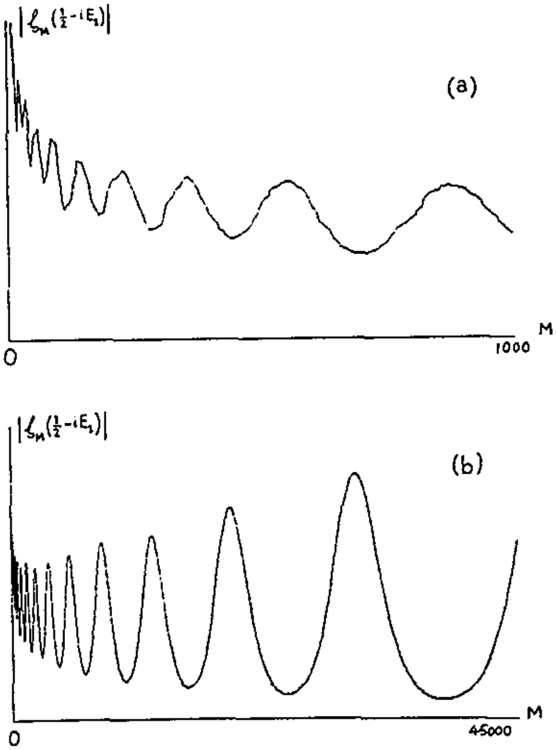

Fig.5 Truncated Riemann product $|\zeta_M(\frac{1}{2}-E_1)|$ as a function of M, evaluated at the lowest Riemann zero E_1, for a) M<1000; b) M<45000.

that $|\zeta_M|$ eventually diverges as

$$|\zeta_M| \sim \exp\left\{M^{1/2}\sin\{E\ln M\}/E\ln M\right\}. \tag{26}$$

The fact that (18) relies on evaluating the Riemann product on the line Rez=1/2, which is displaced by 1/2 from the nearest line on which it converges (z=1), suggests that E=1/2 is the smallest interval over which the Riemann zeros can be discriminated in this way, and hence that ouroborology might fail altogether when the mean separation of zeros is about 1/2. From (7), this occurs when $E_j \sim 2\pi\exp(4\pi) \sim 2\times10^6$, i.e. $j \sim 3\times10^6$, and preliminary numerical exploration in this region indeed suggests the beginning of a failure of (18) to discriminate individual zeros.

Objection 2: the passage from (15) to (16), in which $\sinh(m\lambda/2)$ was replaced by $\exp(m\lambda/2)/2$, was a swindle, implying that for a proper analogy the Riemann formula (18) ought to involve not $p^{m/2}=\exp\{m\ln p/2\}$

but $2\sinh\{m\ln p/2\}$. One answer lies in considering not ζ itself but the product

$$P(E) \equiv \prod_{k=0}^{\infty} \zeta\left(\tfrac{1}{2}+k-iE\right). \tag{27}$$

For real E this converges and has the same zeros as $\zeta(\frac{1}{2}-iE)$, and so can be employed instead of $\zeta(\frac{1}{2}-iE)$ to approximate the fluctuations of the Riemann staircase. An easy calculation (cf.(17-18)) gives

$$N_{P,osc}(E) = -\frac{1}{\pi}\lim_{\eta\to 0} \operatorname{Im}\ln P(E+i\eta) \approx -\frac{1}{2\pi}\sum_{p}\sum_{m=1}^{\infty}\frac{\sin\{mE\ln p\}}{m\sinh\{m\ln p/2\}} \tag{28}$$

which obviously is analogous to (15).

Objection 3: the average density of Riemann zeros (7) has a logarithmic form which is not easy to interpret as the average density $\langle d\rangle$ of eigenvalues of an operator $\hat{H}$ with a classical limit $H(q_\mu, p_\mu)$. For example, in a finite D-dimensional 'billiard' enclosure, $\langle d\rangle \sim E^{(D/2-1)}$; replacing the enclosure by a binding potential similarly fails to give a logarithm. However, Simon [24,25] draws attention to a class of Hamiltonians which are classically unbound but have discrete quantal energy spectra. Of these, a planar billiard with a channel reaching to infinity whilst narrowing hyperbolically does have a logarithmic average level density (and moreover displays intermittent chaos). This can be seen in an elementary way using the Weyl formula

$$\langle d(E)\rangle \sim \frac{\text{area}}{4\pi} \tag{29}$$

(defining energy as $E \equiv k^2$ where k is the de Broglie wavenumber), together with the idea that waves do not penetrate (except with exponential evanescence) where the channel is narrower than a wavelength, that is narrower than about k^{-1}. If the channel boundary has equation $y=B/x$, the area for waves with energy E is thus

$$\text{area} \sim B\int^{y\sim 1/k} y\,dx = B\int^{Bk} dx/x \sim B\ln B\sqrt{E} \sim \ln E \tag{30}$$

giving $\langle d\rangle \sim \ln E$ as claimed. This disposes of objection 3, although there is of course no suggestion that the Riemann $\hat{H}$ really is a billiard of this type (even with magnetic field, to break time-reversal symmetry).

Objection 4: the semiclassical and Riemann formulae (16) and (18) have opposite signs. I have no clear answer to this. It is possible to get negative signs for some of the B_{pm} in (14), when the unstable orbits have focal points [15,16]: for a primitive orbit with n_p focal points, the sign is $(-1)^{[mn_p+1)/2]}$, where [ ] denotes integer part. But no choice of n_p gives negative signs for all the B_{pm}. However, this determination of the sign implicitly assumes an $\hat{H}$ of the form 'Kinetic energy + potential energy', acting on scalar states; without these restrictions it might be possible to get all-negative amplitudes.

4 THE RIEMANN-SIEGEL FORMULA: A RULE FOR QUANTIZING CHAOS?

It follows from the functional equation for $\zeta(z)$ [1] that the following function Z(E) is real and even for real E:

$$Z(E) \equiv \exp\{-i\theta(E)\}\,\zeta(\tfrac{1}{2}-iE) \tag{31}$$

where

$$\begin{aligned}\Theta(E) &= \arg\Gamma(\tfrac{1}{4}+\tfrac{E}{2})-\tfrac{E}{2}\ln\pi \approx \tfrac{E}{2}\ln\tfrac{E}{2\pi}-\tfrac{E}{2}-\tfrac{\pi}{8}\\ &= \pi\left(\langle N_R(E)\rangle+1\right)\end{aligned} \tag{32}$$

(cf.5). The Riemann-Siegel formula is an asymptotic representation of Z(E) for large E:

$$Z(E) = -2\sum_{n=1}^{Q(E)} \frac{\cos\{\pi\langle N_R(E)\rangle - E\ln n\}}{n^{1/2}} + R(E) \tag{33}$$

where

$$Q(E) \equiv \left[\sqrt{(E/2\pi)}\right] \tag{34}$$

and R(E) is a series of remainder terms [1] whose main effect is to cancel the discontinuities of the main sum arising from the E-dependence of the limit Q. Fig.6 shows how accurate the formula is, even for small E.

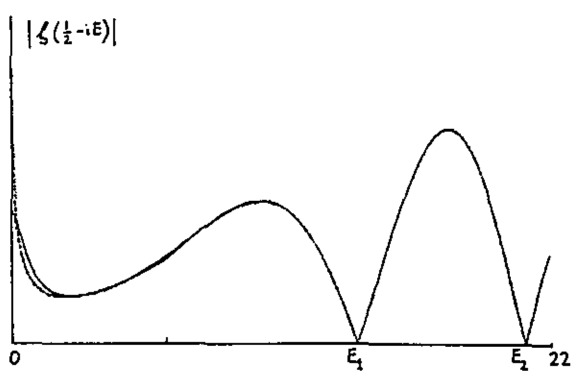

Fig.6. Comparison of $|\zeta(\frac{1}{2}-iE)|$ (full line) with Riemann-Siegel formula plus one correction term (dotted line). (The number Q(E) of terms in the sum (33) changes from zero to one at $E=2\pi$, indicated by a tick)

Now I will outline how the Riemann-siegel formula can be obtained from the series (2b) by an argument with a semiclassical interpretation suggesting a generalization. First note that in obtaining (2b) by expanding the products in (2a), use is made of the factorization theorem that for any integer n we can write

$$\ln n = \sum_p m_p \ln p \tag{35}$$

with a unique choice of the set of integers $\{m_p = 0, 1, 2 \ldots\}$. On the semiclassical analogy, the sum index n can thus be interpreted as runing over <u>all possible combinations of orbit periods</u> (cf.21). Each such combination $\ln n$ will be called a <u>pseudoperiod</u> (pseudoperiods where all m_p except one are zero are periods of actual orbits).

Next, split the sum (2b) into two, with pseudoperiods with n less than and greater than some initially arbitrary value Q, and apply Poisson's summation formula to the second sum:

$$\begin{aligned}\zeta(\tfrac{1}{2} - iE) &= \sum_{n=1}^{Q} \frac{\exp\{iE\ln n\}}{n^{1/2}} + \sum_{n=Q+1}^{\infty} \frac{\exp\{iE\ln n\}}{n^{1/2}} \\ &= \sum_{n=1}^{Q} \frac{\exp\{iE\ln n\}}{n^{1/2}} + \sum_{m=-\infty}^{\infty} \int_{Q+\delta}^{\infty} dn \, \frac{\exp\{i(E\ln n - 2\pi mn)\}}{n^{1/2}}\end{aligned} \tag{36}$$

where $0<\delta<1$. For large E the integrals may be approximated by the method of stationary phase. The stationary point of the m'th integral is at $n=E/2\pi m$, which lies in the integration range only if $1 \leqslant m \leqslant [E/2\pi(Q+\delta)]$. The choice $Q=Q(E)$ (equation 34) and $\delta \to 0$ gives the same limits for the sums over n and m, and then stationary phase leads to

$$\zeta(\tfrac{1}{2} - iE) \approx \sum_{n=1}^{Q(E)} \left(\frac{\exp\{iE\ln n\}}{n^{1/2}} + \frac{\exp\{i(-\frac{\pi}{4} + E\ln\{E/2\pi m\} - E)\}}{n^{1/2}} \right) \tag{37}$$

which gives the Riemann-Siegel sum (33) when combined with the definitions (31) and (32).

What the Poisson technique has achieved in (37) is a <u>resummation of the orbits with long pseudoperiods,</u> to give a series of the same form as the sum over the short pseudoperiods, with precisely the correct phase relation to make $\zeta(\frac{1}{2}-iE)$ have the required analytic structure (31). To see what immense advantage this resummation has produced as compared with the naive ouroborology result (18), consider the oscillations of the terms in (33). The first term, with n=1 (corresponding to the zero pseudoperiod with all m_p equal to zero in 35), oscillates fastest. This term alone gives zeros

$$\langle N_R(E_j) \rangle = j - \tfrac{1}{2} \quad (j = 1, 2, 3 \ldots), \tag{38}$$

with the correct density but which are (asymptotically) uniformly distributed instead of GUE-distributed. The terms with n>1 (nonzero pseudoperiods) oscillate more slowly, and the highest term n=Q(E) is almost constant:

$$\frac{d}{dE}\left(\pi\langle N_R(E)\rangle - E\ln Q\right) = \pi\langle d_R(E)\rangle - \ln Q(E) = \ln\{\sqrt{(E/2\pi)}\} - \ln\{[\sqrt{(E/2\pi)}]\} \approx 0 \tag{39}$$

By contrast, the terms given by unresummed ouroborology oscillate ever faster to give the noise and divergence visible in figs. 4b and 5b.

It is natural to speculate that a similar resummation might be possible for semiclassical quantum chaos. For the simplest nontrivial case, embodied in equations (14-16), this speculation generates, by an argument to be outlined in a moment, a real function W(E) whose zeros would be the eigenvalues. The pseudoorbits are simpler than in the Riemann case, in that each primitive orbit contributes at most one traversal. Thus the actions, periods and instability exponents may be written

$$\left.\begin{array}{l} S^{(k)}(E) = \sum_p \ell_p S_{p1}\;;\; T^{(k)}(E) = \sum_p \ell_p\, dS_{p1}/dE\;;\; \lambda^{(k)}(E) = \sum_p \ell_p \lambda_p \\ \text{with } \{\ell_p = 0 \text{ or } 1\} \text{ labelled } k \text{ in order of increasing } T^{(k)} \end{array}\right\} \tag{40}$$

Then the analogue of the Riemann-Siegel formula (33) would be

$$W(E) = -2\sum_{k=1}^{k_{max}} (-1)^{g_k} \exp\{-\lambda^{(k)}(E)/2\}\cos\{\pi\langle N(E)\rangle - S^{(k)}(E)/\hbar\} \tag{41}$$

where g_k is the number of terms (nonzero ℓ_p's) in the sums (40), and k_{max} is given by the condition that the highest term is non-oscillatory:

$$T^{(k_{max})}(E) = \pi\hbar\,\langle d(E)\rangle \tag{42}$$

Because $\langle d\rangle \sim \hbar^{-D}$ for a system with D freedoms, the longest pseudoperiod is of order $\hbar^{-(D-1)}$, so the number of terms in (41) grows exponentially as ħ decreases or E increases; nevertheless (41) should at least give sensible results for individual eigenvalues in the semiclassical limit, unlike the ouroborology formula (16).

The conjectured Riemann-Siegel analogue (41) arises from considering the function

$$w(E) = \exp\left\{\int_0^E dE'\left(g(E') - \langle g(E')\rangle\right)\right\}, \tag{43}$$

where g(E) is the resolvent (23) and <g(E)> its average. For real E approaching the real axis from above, use of the average of (24) gives

$$\lim_{\eta \to 0} \omega(E+i\eta) = \exp\left\{-\int_{-\infty}^{\infty} dE' \ln\left|1-\frac{E}{E'}\right| \langle d(E')\rangle\right\} \exp\left\{i\pi\langle N(E)\rangle\right\} \prod_j \left(1-\frac{E}{E_j}\right) \tag{44}$$

Therefore $\omega(E)$ has zeros at the eigenvalues E_j. On the other hand, for sufficiently large Im E the fluctuations g-<g> can be approximated semiclassically by the analogue of (16) (which replaces the analogue of (15) by an argument similar to that centred on (27) and (28)). Thus

$$\omega(E) \approx \exp\left\{-\sum_p \sum_{m=1}^{\infty} \frac{\exp\{-m\lambda_p/2\}}{m} \exp\{imS_{p1}/\hbar\}\right\}$$

$$= \prod_p \left(1-\exp\{-\lambda_p/2\}\exp\{iS_{p1}(E)/\hbar\}\right), \tag{45}$$

giving $\omega(E)$ as a product over primitive orbits, analogous to the product (2a) for $\zeta(z)$ with the difference that the factors do not appear as reciprocals (this originates in the mysterious sign difference described in objection 4 at the end of section 3).

Expanding the product in (45) we obtain an analogue of the sum (2a), involving the pseudoorbits (40). Then the conjectured Riemann-Siegel analogue (41), with W(E) identified from (44) as proportional to the product $\prod_j(1 - E/E_j)$, would follow, if only (!) one could resum the high pseudoorbits in a way similar to that which gave the true Riemann-Siegel formula (33).

Exactly the dynamical zeta function $\omega(E)$, in the form (45), has been written down by Gutzwiller [17]. Ruelle [26,27] introduced analogous zeta functions to study chaotic <u>dissipative</u> dynamical systems and these have been further investigated by Parry [28,29] (who introduced what I here call pseudoorbits) and Parry and Pollicott [30]; in these dissipative zeta functions, orbit actions and instability exponents are (by implication) considered as proportional to periods and entropy. None of these authors appear to have considered the possibility that a Riemann-Siegel formula might exist for dynamical zeta functions.

<u>ACKNOWLEDGMENT</u> No military agency supported this research.

<u>REFERENCES</u>

[1] H.M.Edwards, <u>Riemann's Zeta Function</u> (Academic Press: New York and London) (1974)

[2] C.E.Porter, Statistical theories of Spectra: Fluctuations (Academic Press: New York (1965)
[3] F.J.Dyson, in 'Symmetries in particle physics' (eds: I.Bars, A.Chodos and C-H Tze; Plenum: New York and London) pp 279-281 (1984)
[4] O.Bohigas and M.J.Giannoni, Chaotic motion and Random-matrix Theories in Mathematical and Computational Methods in Nuclear Physics eds.J.S.Dehesa, J.M.G.Gomez and A.Polls, Lecture Notes in Physics 209 (Springer-Verlag, N.Y.) pp.1-99 (1984)
[5] H.L.Montgomery, Proc.Symp.Pure Math. 24 181-193 (1973)
[6] M.L.Mehta, Random Matrices and the Statistical Theory of Energy Levels (Academic Press: New York and London) (1967)
[7] T.H.Seligman, and J.J.M.Verbaarschot, Phys.Lett. 108A 183-187 (1985)
[8] M.V.Berry and M.Robnik, J.Phys.A. (1986) In press
[9] M.Robnik and M.V.Berry, J.Phys.A. (1986) In press
[10] M.V.Berry, Proc.Roy.Soc.Lond. A400 229-251 (1985)
[11] M.V.Berry and M.Tabor, Proc.Roy.Soc.Lond. A356 375-394 (1977)
[12] O.Bohigas, M.J.Giannoni and C.Schmit, Phys.Rev.Lett. 52 1-4 (1984)
[13] M.V.Berry, Ann.Phys. (N.Y) 131 163-216 (1981)
[14] Ouroboros: see Encyclopaedia Brittanica (15th ed.) vol.9 p13 (1985)
[15] M.C.Gutzwiller, J.Math.Phys. 12 343-358 (1971)
[16] M.C.Gutzwiller, in 'Path Integrals and their Applications in Quantum, Statistical and Solid-State Physics' (Eds. G.J.Papadopoulos and J.T.Devreese) Plenum, N.Y. 163-200 (1978)
[17] M.C.Gutzwiller, in 'Stochastic Behaviour in Classical and Quantum Hamiltonian Systems' (eds. G.Casati and J.Ford) Lecture Notes in Physics 93 (Springer: Berlin)(1979)
[18] R.Balian and C Bloch, Ann.Phys.(N.Y) 69 76-160 (1972)
[19] M.V.Berry, Semiclassical Mechanics of Regular and Irregular Motion in Chaotic Behavior of Deterministic Systems (Les Houches Lectures XXXVI, eds.G.Iooss, R.H.G.Helleman and R.Stora (North-Holland: Amsterdam) pp171-271 (1983)
[20] H.P.McKean, Comm.Pure App. Math. 25 225-246 (1972)
[21] D.A.Hejhal, Duke Math.J. 43 441-482 (1976)
[22] B.S.Pavlov and L.D.Faddeev, J.Soviet Math 3 522-548 (1975)
[23] M.C.Gutzwiller, Physica 7D 341-355 (1983)
[24] B.Simon, Ann.Phys. 146 209-220 (1983)
[25] B.Simon, J.Funct.Anal. 53 84-98 (1983)
[26] D.Ruelle, Inventiones Math 34 231-242 (1976)
[27] D.Ruelle, Ergod.Th.and Dynam.Syst. 2 99-107 (1982)
[28] W.Parry Israel J.Math. 45 41-52 (1983)
[29] W.Parry Ergod Th.and Dynam.Syst. 4 117-134 (1984)
[30] W.Parry and M.Pollicott, Annals of Math. 118 573-591 (1983)