SIAM REVIEW
Vol. 41, No. 2, pp. 236–266

The Riemann Zeros and Eigenvalue Asymptotics*

M. V. Berry[†]
J. P. Keating[‡]

Abstract. Comparison between formulae for the counting functions of the heights t_n of the Riemann zeros and of semiclassical quantum eigenvalues E_n suggests that the t_n are eigenvalues of an (unknown) hermitean operator H, obtained by quantizing a classical dynamical system with hamiltonian H_{cl}. Many features of H_{cl} are provided by the analogy; for example, the "Riemann dynamics" should be chaotic and have periodic orbits whose periods are multiples of logarithms of prime numbers. Statistics of the t_n have a similar structure to those of the semiclassical E_n; in particular, they display random-matrix universality at short range, and nonuniversal behaviour over longer ranges. Very refined features of the statistics of the t_n can be computed accurately from formulae with quantum analogues. The Riemann-Siegel formula for the zeta function is described in detail. Its interpretation as a relation between long and short periodic orbits gives further insights into the quantum spectral fluctuations. We speculate that the Riemann dynamics is related to the trajectories generated by the classical hamiltonian $H_{\mathrm{cl}} = XP$.

Key words. spectral asymptotics, number theory

AMS subject classifications. 11M26, 11M06, 35P20, 35Q40, 41A60, 81Q10, 81Q50

PII. S0036144598347497

1. Introduction.

Our purpose is to report on the development of an analogy, in which three areas of mathematics and physics, usually regarded as separate, are intimately connected. The analogy is tentative and tantalizing, but nevertheless fruitful. The three areas are eigenvalue asymptotics in wave (and particularly quantum) physics, dynamical chaos, and prime number theory. At the heart of the analogy is a speculation concerning the zeros of the Riemann zeta function (an infinite sequence of numbers encoding the primes): the Riemann zeros are related to the eigenvalues (vibration frequencies, or quantum energies) of some wave system, underlying which is a dynamical system whose rays or trajectories are chaotic.

Identification of this dynamical system would lead directly to a proof of the celebrated Riemann hypothesis. We do not know what the system is, but we do know many of its properties, and this knowledge has brought insights in both directions: from mathematics to physics, by stimulating the development of new spectral asymptotics, and from physics to mathematics, by indicating previously unsuspected correlations between the Riemann zeros. We have reviewed some of this material before

*Received by the editors October 16, 1998; accepted for publication (in revised form) November 21, 1998; published electronically April 23, 1999.
http://www.siam.org/journals/sirev/41-2/34749.html
[†]H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, United Kingdom.
[‡]School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom and Basic Research Institute in the Mathematical Sciences, Hewlett-Packard, Laboratories Bristol, Filton Road, Stoke Gifford, Bristol BS12 6QZ, United Kingdom (J.P.Keating@bristol.ac.uk).

[1, 2, 3, 4, 5, 6], but these accounts do not include several recent developments to be described here, especially those in the last part of section 4 and all of sections 5 and 6.

To motivate the approach from physics, we begin with the counting function for the primes, $\pi(x)$, defined as the number of primes less than x (thus $\pi(3.5) = 2$); this is a staircase function, with unit steps at the primes p. The density of primes is the distribution

$$\pi'(x) \equiv \sum_p \delta(x - p). \tag{1.1}$$

At the roughest level of description, and with the distribution appropriately smoothed,

$$\pi'(x) \sim \frac{1}{\log x} \tag{1.2}$$

(as implied by the prime number theorem: $\pi(x) \sim x/\log x$).

One of Riemann's great achievements [7, 8] was to give an exact formula for $\pi'(x)$, constructed as follows. First, $\pi'(x)$ is expressed in terms of a function $J(x)$ [7, Chap. 1] that has jumps at prime powers:

$$\pi'(x) = \frac{1}{x}\sum_{k=1}^{\infty}\frac{\mu_k x^{1/k}}{k^2}J'\left(x^{1/k}\right). \tag{1.3}$$

In this formula, μ_k are the Möbius numbers $(1, -1, -1, 0, -1, 1, \ldots)$ [7]. Each of the partial densities J' is the sum of a smooth part (dominated by (1.2)) and an infinite series of oscillations:

$$J'(x) = \frac{1}{\log x}\left(1 - \frac{1}{x(x^2-1)}\right) - \frac{2}{\sqrt{x}\log x}\sum_{\mathrm{Re}t_n > 0}\frac{\cos\{\mathrm{Re}(t_n)\log x\}}{x^{\mathrm{Im}t_n}} \tag{1.4}$$

(see section 1.18 of [7]). Here the numbers t_n in the oscillatory contributions are related to the complex Riemann zeros, defined as follows.

Riemann's zeta function, depending on the complex variable s, is defined as

$$\zeta(s) \equiv \prod_p \left(1 - p^{-s}\right)^{-1} = \sum_{n=1}^{\infty} n^{-s} \quad (\mathrm{Re}s > 1) \tag{1.5}$$

and by analytic continuation elsewhere in the s plane. It is known that the complex zeros (i.e., those with nonzero imaginary part) of $\zeta(s)$ lie in the "critical strip" $0 < \mathrm{Re}s < 1$, and the Riemann hypothesis states that in fact all these zeros lie on the "critical line" $\mathrm{Re}s = 1/2$ (see Figure 1). The numbers t_n in (1.4) are defined by

$$\zeta\left(\tfrac{1}{2} + it_n\right) = 0 \quad (\mathrm{Re}t_n \neq 0). \tag{1.6}$$

If the Riemann hypothesis is true, all the (infinitely many) t_n are real, and are the heights of the zeros above the real s axis. It is known by computation that the first 1,500,000,001 complex zeros lie on the line [9], as do more than one-third of all of them [10].

Each term in the sum in (1.4) describes an oscillatory contribution to the fluctuations of the density of primes, with larger $\mathrm{Re}t_n$ corresponding to higher frequencies.

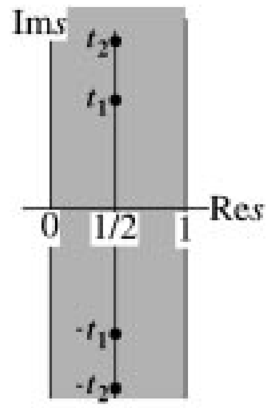

Fig. 1 *Complex s plane, showing the critical strip (shaded) and the complex Riemann zeros (there are trivial zeros at $s = -2, -4, \ldots$).*

Because of the logarithmic dependence, each oscillation gets slower as x increases. This slowing-down can be eliminated by the change of variable $u = \log x$; thus

$$f(u) \equiv \tfrac{1}{2} \exp\left(\tfrac{1}{2}u\right) \left[u\pi'(\exp u) - 1\right] + \tfrac{1}{4}$$

$$(1.7) \qquad = -\sum_{\mathrm{Re}t_n > 0} \cos\{\mathrm{Re}(t_n)\,u\} \exp\{-\mathrm{Im}(t_n)\,u\} + O\left(\exp\left(-\tfrac{1}{6}u\right)\right).$$

If the Riemann hypothesis is true, $\mathrm{Im}t_n = 0$ for all n, and the function $f(u)$, constructed from the primes, has a discrete spectrum; that is, the support of its Fourier transform is discrete. If the Riemann hypothesis is false, this is not the case. The frequencies t_n are reminiscent of the decomposition of a musical sound into its constituent harmonics. Therefore there is a sense in which we can give a one-line nontechnical statement of the Riemann hypothesis: "The primes have music in them."

However, readers are cautioned against thinking that it would be easy to hear this prime music by constructing $f(u)$ as defined in (1.7) and then converting it into an audio signal. In order for the human ear to hear the lowest Riemann zero, with $t_1 = 14.13\ldots$, it would be necessary to play $N \approx 100$ periods of $\cos(t_1 u)$, requiring primes in the range $0 < x < \exp(2\pi N/t_1) \approx \exp(45) \approx 10^{19}$.

On this acoustic analogy, the heights t_n (hereinafter referred to simply as "the zeros") are frequencies. This raises the compelling question: frequencies of what? A natural answer would be: frequencies of some vibrating system. Mathematically, such frequencies—real numbers—are discrete eigenvalues of a self-adjoint (hermitean) operator. That the search for such an operator might be a fruitful route to proving the Riemann hypothesis is an old idea, going back at least to Hilbert and Polya [7]; what is new is the physical interpretation of this operator and the detailed information now available about it.

The mathematics of almost all eigenvalue problems encountered in wave physics is essentially the same, but the richest source of such problems is quantum mechanics, where the eigenvalues are the energies of stationary states ("levels"), rather than frequencies as in acoustics or optics, and the operator is the hamiltonian. Reflecting this catholicity of context, we will refer to the t_n interchangeably as energies or frequencies, and the operator as H (Hilbert, Hermite, Hamilton...).

To help readers navigate through this review, here is a brief description of the sections. In section 2 we describe the basis of the Riemann-quantum analogy, which is an identification of the periodic orbits in the conjectured dynamics underlying the Riemann zeros, made by comparing formulae for the counting functions of the t_n and of asymptotic quantum eigenvalues. Section 3 explains the significance of the long periodic orbits in giving rise to universal (that is, system-independent) behaviour in classical and semiclassical mechanics and, by analogy, the Riemann zeros. The application of these ideas to the statistics of the zeros and quantum eigenvalues is taken up in section 4. Section 5 is a description of a powerful method for calculating the t_n (the Riemann-Siegel formula), with a physical interpretation in terms of resurgence of long periodic orbits that implies new interpretations of the periodic-orbit sum for quantum spectra. The properties of the conjectured dynamical system are listed in section 6, where it is speculated that the zeros are eigenvalues of some quantization of the dynamics generated by the hamiltonian $H_{\mathrm{cl}} = XP$.

2. The Analogy. The basis of the analogy is a formal similarity between representations for the fluctuations of the counting functions for the Riemann zeros t_n and for vibration frequencies associated with a system whose rays are chaotic. For the t_n (assumed real), the counting function is defined for $t > 0$ as

$$\mathcal{N}(t) \equiv \sum_{n=1}^{\infty} \Theta(t - t_n), \tag{2.1}$$

where Θ denotes the unit step. Central to our arguments is the fact that $\mathcal{N}(t)$ can be decomposed as follows [11]:

$$\mathcal{N}(t) = \langle \mathcal{N}(t) \rangle + \mathcal{N}_{\mathrm{fl}}(t), \tag{2.2}$$

where

$$\begin{aligned} \langle \mathcal{N}(t) \rangle &\equiv \frac{\theta(t)}{\pi} + 1 = \frac{1}{\pi}\left[\arg\Gamma\left(\frac{1}{4} + \frac{1}{2}it\right) - \frac{1}{2}t\log\pi\right] + 1 \\ &= \frac{t}{2\pi}\log\left(\frac{t}{2\pi e}\right) + \frac{7}{8} + O\left(\frac{1}{t}\right) \end{aligned} \tag{2.3}$$

and

$$\mathcal{N}_{\mathrm{fl}}(t) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\operatorname{Im}\log\zeta\left(\frac{1}{2} + it + \varepsilon\right). \tag{2.4}$$

(The branch of the logarithm is chosen to be continuous, with $\mathcal{N}_{\mathrm{fl}}(0) = 0$.)

These two components can be interpreted as the smooth and fluctuating parts of the counting function. Here and hereinafter the notation $\langle \cdots \rangle$ denotes a local average of a fluctuating quantity, over a range large compared with the length scales of the fluctuations but small compared with any secular variation. Implicit in such averaging is an asymptotic parameter; in the present case this is t, and the averaging range is large compared with the mean spacing of the zeros but small compared with t itself.

The formula for $\langle \mathcal{N} \rangle$ can be obtained from the functional equation for $\zeta(s)$ [7]. It follows by differentiating the last member of (2.3) that the asymptotic density of the zeros is

$$\langle d(t) \rangle = \frac{1}{2\pi}\log\left(\frac{t}{2\pi}\right) + O\left(\frac{1}{t^2}\right) \tag{2.5}$$

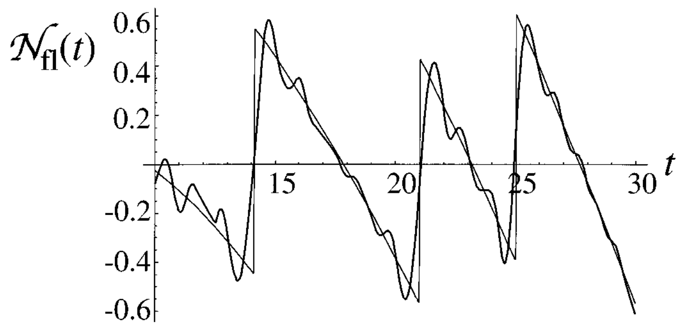

Fig. 2 *Thick line: Divergent series* (2.6) *for the counting function fluctuations* $\mathcal{N}_{\text{fl}}$ *of the Riemann zeros, including all values of* m *and the first* 50 *primes* p. *Thin line: Exact calculation of* $\mathcal{N}_{\text{fl}}$ *from* (2.4).

and therefore that the mean spacing between the zeros decreases logarithmically with increasing t. Underlying the formula for $\mathcal{N}_{\text{fl}}$ are the observations that the phase of a function jumps by π on passing close to a zero, and that $\zeta(s) \to 1$ as $\text{Re}s \to \infty$, so that between the jumps in $\mathcal{N}_{\text{fl}}$ this function varies smoothly, implying that its average value is zero.

Now we substitute into (2.4) the Euler product (1.5), disregarding the fact that this does not converge in the critical strip, and obtain the divergent but formally exact expression

$$\begin{aligned} \mathcal{N}_{\text{fl}}(t) &= -\frac{1}{\pi}\text{Im}\sum_p \log\left\{1 - \frac{\exp(-it\log p)}{\sqrt{p}}\right\} \\ &= -\frac{1}{\pi}\sum_p \sum_{m=1}^{\infty} \frac{\exp\left(-\frac{1}{2}m\log p\right)}{m} \sin\{tm\log p\}. \end{aligned} \tag{2.6}$$

This formula gives the fluctuations as a series of oscillatory contributions, each labelled by a prime p and an integer m, corresponding to the prime power p^m. Terms with $m > 1$ are exponentially smaller than those with $m = 1$. The oscillation corresponding to p has a "wavelength" (that is, t-period)

$$\tau_p = \frac{2\pi}{\log p}. \tag{2.7}$$

In order to discriminate individual zeros, sufficiently many terms must be included in the sum for this wavelength to be less than the mean spacing; from (2.5), this gives $p < t/2\pi$. When truncated in this way, the sum (2.6) can reproduce the jumps quite accurately for low-lying zeros, as Figure 2 shows, even though the complete sum diverges.

Consider now a classical dynamical system [12] in a configuration space with D freedoms, coordinates $\mathbf{q} = \{q_1, \ldots, q_D\}$, and momenta $\mathbf{p} = \{p_1, \ldots, p_D\}$. Trajectories are generated by a hamiltonian function $H(\mathbf{q}, \mathbf{p})$ on the two-dimensional phase space $\{\mathbf{q}, \mathbf{p}\}$, whose conserved value is the energy E. In quantum physics, $\mathbf{q}$ and $\mathbf{p}$ are operators, with commutation relation $[\mathbf{q}, \mathbf{p}] = i\hbar$, where $\hbar \equiv h/2\pi$ is Planck's constant. Then $H(\mathbf{q}, \mathbf{p})$, augmented by boundary conditions, becomes a hermitean wave operator, whose eigenvalues, discrete if the system is bound, are the quantum energy levels E_n. More generally, this formalism applies to any wave system (e.g., water waves [13]) with coordinates $\mathbf{q}$ and wavenumber $\mathbf{k}$, defined by a dispersion relation $\omega(\mathbf{q}, \mathbf{k})$, the connection between the quantum and wave formalisms being

$$\mathbf{p} = \hbar\mathbf{k}, \quad H(\mathbf{q}, \mathbf{p}) = \hbar\omega(\mathbf{q}, \mathbf{p}/\hbar). \tag{2.8}$$

Familiar wave equations appear when the commutation relations are implemented with $\mathbf{k} = -i\nabla$, and Hamilton's equations are the corresponding ray equations (in optics these are the rays generated by Snell's law or Fermat's principle). For example, a locally uniform medium (H independent of $\mathbf{q}$) with impenetrable walls corresponds to "quantum billiards," where waves are governed by the Helmholtz equation with Dirichlet boundary conditions, and the (straight) rays are reflected specularly at the walls [14]. Of special interest to us is the asymptotics of the eigenvalues E_n in the semiclassical limit $\hbar \to 0$, which from (2.8) is equivalent to the short-wavelength or high-frequency limit.

Waves, in particular the eigenfunctions of H, usually depend not on individual trajectories but on families of trajectories, whose global structure is an important determinant of the energy-level asymptotics. Of interest here is the case where the trajectories are chaotic [15, 16, 17], that is, where E is the only globally conserved quantity and neighbouring trajectories diverge exponentially. Then on a given energy shell (that is, for given E), the usual structure—and the one we will consider here—is that all initial conditions generate trajectories that explore the $(2D-1)$-dimensional energy surface ergodically, except for a set, dense but of zero measure, of (one-dimensional) isolated unstable periodic orbits.

An important result of modern mathematical physics, central to the Riemann-quantum analogy, is that these isolated periodic trajectories determine the fluctuations in the counting function $\mathcal{N}(E)$ of the energy levels [18, 19, 20, 21]. Using the notation (2.2), with E replacing t, we can separate $\mathcal{N}(E)$ into its smooth and fluctuating parts $\langle\mathcal{N}(E)\rangle$ and $\mathcal{N}_{\text{fl}}(E)$. The averaging is over an energy interval large compared with the mean level spacing but classically small, that is, vanishing with $\hbar$. We state the formula for $\mathcal{N}_{\text{fl}}(E)$ and then explain it:

$$N_{\text{fl}}(E) \sim \frac{1}{\pi}\sum_p \sum_{m=1}^{\infty} \frac{\sin\left\{mS_p(E)/\hbar - \frac{1}{2}\pi m\mu_p\right\}}{m\sqrt{\left|\det\left(\mathsf{M}_p^m - I\right)\right|}}. \tag{2.9}$$

The symbol $\sim$ indicates that the formula applies asymptotically, that is, for small $\hbar$. (In the special case of the Selberg trace formula [21], corresponding to waves on a compact surface of constant negative curvature, the formula is exact.) The index p labels primitive periodic orbits, that is, orbits traversed once. The index m labels their repetitions. Therefore, the two sums together include all periodic orbits. $S_p(E)$ is the action of the primitive orbit p, that is,

$$S_p(E) = \oint_p \mathbf{p} \cdot \mathbf{dq}. \tag{2.10}$$

In terms of S_p, the period of the orbit is

$$T_p = \frac{\partial S_p}{\partial E}. \tag{2.11}$$

The hyperbolic symplectic matrix M_p (the monodromy matrix) describes the exponential growth of deviations from p of nearby (linearized) trajectories, between successive crossings of a Poincaré surface of section transverse to p. μ_p is the Maslov phase, determined [22] by the winding round p of the stable and unstable manifolds containing the orbit.

Physically, the appearance of periodic orbits is not surprising. The levels E_n, counted by $\mathcal{N}$, are associated with stationary states, that is, states or modes that are time-independent. By the correspondence principle, their asymptotics should depend on phase space structures unchanged by evolution along rays, that is, the invariant manifolds with energy E. In the type of chaotic dynamics we are considering, there are two types of invariant manifold: the whole energy surface, which determines $\langle\mathcal{N}(E)\rangle$ as we will see, and, decorating this, the tracery of periodic orbits, which determines the finer details of the spectrum as embodied in the fluctuations $\mathcal{N}_{\mathrm{fl}}(E)$.

For long orbits, the determinant is dominated by its expanding eigenvalues, and, for large T_p,

$$\det\left(\mathsf{M}_p^m - 1\right) \sim \exp\left(m\lambda_p T_p\right), \tag{2.12}$$

where λ_p is the Liapunov (instability) exponent of the orbit p. Thus, approximately,

$$\mathcal{N}_{\mathrm{fl}}\left(E\right) \sim \frac{1}{\pi}\sum_p \sum_{m=1}^{\infty} \frac{\exp\left(-\frac{1}{2}m\lambda_p T_p\right)}{m} \sin\left\{mS_p\left(E\right)/\hbar - \frac{1}{2}\pi m\mu_p\right\}. \tag{2.13}$$

Now we can make the formal analogy with the corresponding formula (2.6) for the counting function fluctuations of the Riemann zeros:

(2.14)

	Quantum	Riemann
Dimensionless actions	$\frac{mS_p}{\hbar}$	$mt\log p$
Periods	mT_p	$m\log p$
Stabilities	$\frac{1}{2}\lambda_p T_p$	$\frac{1}{2}\log p \Rightarrow \lambda_p = 1$
Asymptotics	$\hbar \to 0$	$t \to \infty$

.

The nonappearance of $\hbar$ on the "Riemann" side indicates that the dynamical system underlying the zeros is scaling, in the sense that the trajectories are the same for all "energies" t, as in the most familiar scaling system, namely, quantum billiards, where, for a particle of mass m, energy scales according to the combination $k = \sqrt{(2mE)}/\hbar$, and, for an orbit of length L_p, $S_p/\hbar = kL_p$. With the analogy, primes acquire a new significance, as primitive periodic orbits, whose periods are log$p$. The index m in (2.6) then labels their repetitions.

The fact that all orbits have the same instability exponent (unity) indicates that the Riemann dynamics is homogeneously unstable, that is, uniformly chaotic. Moreover, the dynamics does not possess time-reversal symmetry. If it did, degeneracy of actions between each orbit and its time-reversed partner would lead to their contributing coherently to $\mathcal{N}(t)$, so that for most orbits (those that are not self-retracing) the prefactor in (2.6) would be $2/\pi$ rather than $1/\pi$.

An alternative form of the periodic-orbit sum (2.9), which will be useful later, is in terms of the level density

$$d\left(E\right) = \frac{d\,\mathcal{N}\left(E\right)}{dE}. \tag{2.15}$$

Denoting primitive and repeated periodic orbits by the common index j $(=\{p, m\})$, we can write

$$d_{\text{fl}}(E) = \frac{1}{\pi\hbar}\sum_j A_j \cos\{S_j(E)/\hbar\}, \tag{2.16}$$

where for convenience we have absorbed the Maslov indices into the actions, and the amplitude A_j is

$$A_j \sim \frac{T_j}{m\sqrt{\det|(\mathsf{M}_j - \mathsf{I})|}} \tag{2.17}$$

as $\hbar \to 0$. For the Riemann zeros, the corresponding formula, from (2.6) and (2.14), has $\mu_j = 0$,

$$A_j = -\frac{\log p}{p^{m/2}} = -\frac{T_j}{m}\exp\left\{-\tfrac{1}{2}T_j\right\}, \tag{2.18}$$

and is an identity rather than an asymptotic approximation.

There are two discordant features of the analogy [1], to which we will return. First, the exponential decay of long orbits in the quantum formula (2.13) is an approximation to the determinant in (2.9), whereas for the Riemann zeros the exponential in (2.6) is exact. Second, the negative sign in (2.6) indicates that when the Maslov phases $\pi m\mu_p/2$ are reinstated in (2.13) their value should be π for all orbits, but this is hard to understand because if the index is π for a given orbit it should be 2π for the same orbit traversed twice.

The smooth part $\langle\mathcal{N}(E)\rangle$ of the counting function is, to leading order in $\hbar$, the number of phase space quantum cells (volume h^D) in the volume $\Omega(E)$ of the energy surface $H = E$; thus $\langle\mathcal{N}(E)\rangle \approx \Omega(E)/h^D$. For billiards, Ω is proportional to the spatial volume confining the system (this is Weyl's asymptotics [23]). The mean level density is thus

$$\langle d(E)\rangle \sim \frac{\Omega'(E)}{h^D}. \tag{2.19}$$

In the quantum formula (2.13), each orbit contributes an oscillation to $\mathcal{N}_{\text{fl}}(E)$, with energy "wavelength" (cf. (2.7))

$$\varepsilon_p = \frac{h}{T_p(E)}. \tag{2.20}$$

This should be compared with the mean spacing of the eigenvalues, which is the reciprocal of the mean level density and so (from (2.19)) of order $\hbar^D$. An important implication is that the oscillation contributed by a given orbit has, asymptotically, a wavelength much larger than the mean level spacing. Thus in order to have a chance of resolving individual levels it is necessary to include at least all those orbits with periods up to

$$T_H(E) = 2\pi\hbar\langle d\rangle = O\left(\frac{1}{\hbar^{D-1}}\right). \tag{2.21}$$

This evokes the time-energy uncertainty relation, so T_H is called the Heisenberg time. Asymptotically, T_H corresponds to very long orbits, or, in the Riemann case, large primes $p_H(t) = t/2\pi$ (cf. the discussion following (2.7)). In what follows, this emphasis on long orbits will play a key role.

3. Long Orbits and Universality. In a classically chaotic system, the periodic orbits proliferate exponentially as their period increases [24], with density

$$\rho(T) \equiv \frac{\text{number of orbits with periods between } T \text{ and } T+dT}{dT} \sim \frac{\exp(\lambda T)}{T} \quad \text{as } T \to \infty. \tag{3.1}$$

Here, λ is the topological entropy of the system. In the cases we are interested in, λ can be identified with a suitable average of the instability exponents of long periodic orbits (cf. (2.12)). In the Riemann case, where according to (2.14) the periodic orbits correspond to primes, (3.1) nicely reproduces the prime number theorem (1.2) and thereby reinforces the analogy (the repetitions, labelled by m, give exponentially smaller corrections).

From (2.18), the proliferation in (3.1) cancels the decay of the intensities A_j^2 for long orbits. One way to write this is

$$\lim_{T\to\infty} \frac{1}{T} \sum_j A_j^2 \delta(T - T_j) = 1. \tag{3.2}$$

This is the sum rule of Hannay and Ozorio de Almeida [25]. Its importance is threefold: first, it does not contain $\hbar$ and so is a *classical* sum rule. Second, the amplitudes A_j nevertheless have significance in *quantum* (i.e., wave) asymptotics, because they give the strengths of the contributions to spectral density fluctuations. Third, the rule is universal: (3.2) contains no specific feature of the dynamics—it holds for all systems that are ergodic. One way to appreciate the naturalness of this universality is to imagine that a long orbit with energy E, inscribed on the constant-energy surface $H = E$, forms an intricate tracery that, with the slightest smoothing, could cover the surface uniformly with respect to the microcanonical (Liouville) measure. This "phase-space democracy" is the basis of Hannay and Ozorio de Almeida's derivation.

Expressed mathematically, this ergodicity-related sum rule corresponds to an eigenvalue $\nu_0 = 1$ (associated with the invariant measure) of the Perron–Frobenius operator that generates the classical flow in phase space. Equivalently [26, 27], it corresponds to a simple pole at $s = 0$ of the dynamical zeta function $\zeta_D(s)$, defined (for two-dimensional systems, for example) by

$$\frac{1}{\zeta_D(s)} \equiv \prod_p \prod_{m=0}^{\infty} \left(1 - \frac{\exp(sT_p)}{|\Lambda_p| \Lambda_p^m}\right)^{m+1}, \tag{3.3}$$

where Λ_p is the larger eigenvalue ($|\Lambda_p| > 1$) of the monodromy matrix M. The rest of the spectrum of the Perron–Frobenius operator, or equivalently the analytic structure of $\zeta_D(s)$ away from $s = 0$, determines the rate of approach to ergodicity—that is, it is related to the system-specific short-time dynamics.

Now recall that according to (2.21) the long orbits determine spectral fluctuations on the scale of the mean level separation. The universality of the classical sum rule suggests that the spectral fluctuations should also show universality on this scale. And by the Riemann-quantum analogy, we expect this spectral universality to extend to the Riemann zeros t_n.

It is in the *statistics* of the levels and Riemann zeros that the universality appears. This is to be expected, since ergodicity is a statistical property of long orbits.

It is important to note that we are here considering individual systems and not ensembles, so statistics cannot be defined in the usual way, as ensemble averages. Instead, we rely on the presence of an asymptotic parameter (see the remarks after (2.4), and before (2.9)): high in the spectrum (or for large t in the Riemann case), there are many levels (or zeros) in a range where there is no secular variation, and it is this large number that enables averages to be performed. Universality then emerges in the limit $\hbar \to 0$ (or $t \to \infty$) for correlations between fixed numbers of levels or zeros.

A mathematical theory of universal spectral fluctuations already exists in the more conventional context where statistics are defined by averaging over an ensemble. This is *random-matrix theory* [28, 29, 30, 31, 32], where the correlations between matrix eigenvalues are calculated by averaging over ensembles of matrices whose elements are randomly distributed, in the limit where the dimension of the matrices tends to infinity. Here the relevant ensemble is that of complex hermitean matrices: the "Gaussian unitary ensemble" (GUE). As will be discussed in the next section, it is precisely these statistics that apply to high eigenvalues of individual chaotic systems without time-reversal symmetry, and also to high Riemann zeros, in the sense that the spectral or Riemann-zero averages described in the previous paragraph coincide with GUE averages.

First, however, we give a very simple argument [33] showing that the approach to universality must be nonuniform. The classical sum rule (3.2) applies to long orbits but not to short ones, because these will reflect the specific dynamics of the system whose spectrum is being considered. Therefore, spectral features that depend on short orbits can be expected to be nonuniversal. From (2.20), these are fluctuations on the energy scale $\varepsilon_0 = h/T_0$, where T_0 is the period of the shortest orbit. This scale is asymptotically small but still large compared with the separation of order h^D between neighbouring eigenvalues. On this basis, we expect universality to be a good approximation for correlations between eigenvalues separated by up to $O(1/h^{(D-1)})$ mean spacings, but not for larger separations. For the Riemann zeros, $T_0 = \log 2$ (equation (2.14)), whereas the mean separation between zeros is $2\pi/\log(t/2\pi)$. Therefore universality for zeros near t should break down beyond about $\log(t/2\pi)/\log 2$ mean spacings. We regard the observation of the breakdown of random-matrix universality for the Riemann zeros [34], in accordance with this prediction, as giving powerful support to the analogy with quantum or wave eigenvalues.

4. Periodic-Orbit Theory for Spectral Statistics. In discussing statistics, it will be simplest to measure intervals between eigenvalues or Riemann zeros in units of the local mean spacing. We denote such intervals by x, and the corresponding levels or zeros, referred to a local origin, by x_n; in these units, $\langle d(x)\rangle = 1$. We will mainly be concerned with statistics that are bilinear in the level density, the simplest being the *pair correlation* of the density fluctuations, defined in [31], in the sense of a distribution, as

$$
\begin{aligned}
&R(x;y) \equiv \text{probability density of separations } x \text{ of levels or zeros} \\
&\quad \text{close to a scaled position } y \\
&= \frac{1}{N}\sum_{m\neq n} \delta\left(x_m - x_n - x\right) \\
&= \left\langle d\left(y-\tfrac{1}{2}x\right) d\left(y+\tfrac{1}{2}x\right)\right\rangle - \delta(x) \\
&= 1 + \left\langle d_{\text{fl}}\left(y-\tfrac{1}{2}x\right) d_{\text{fl}}\left(y+\tfrac{1}{2}x\right)\right\rangle - \delta(x)\,.
\end{aligned}
\tag{4.1}
$$

(In the second member, the sum is over a stretch of N levels near y, with $N >> 1$.) R gives the correlation between levels near E, or, correspondingly, Riemann zeros near

t; for simplicity of notation, we will henceforth not indicate these base levels (denoted y in (4.1)).

Closely related to R is the *form factor* $K(\tau)$ (the name comes from crystallography), defined as

$$K(\tau) = 1 + \int_{-\infty}^{\infty} dx \exp\{2\pi i x\tau\}(R(x) - 1) \\ = -\delta(\tau) + \frac{1}{N}\sum_m \sum_n \exp\{2\pi i\tau(x_m - x_n)\}, \tag{4.2}$$

where the sum is as in (4.1). Here the variable τ (conjugate to x) is the scaled time

$$\tau = \frac{T}{T_H}, \tag{4.3}$$

where T_H is the Heisenberg time (equation (2.21)). With the definitions given, both R and K tend to 1 at long range; the term $\delta(x)$ in R ensures that this requirement is compatible with (4.2).

Other statistics that are bilinear in d can be expressed in terms of K or R. A useful one is the *number variance*:

$$\begin{aligned} &\Sigma^2(x) \equiv \text{variance of number of levels or zeros in} \\ &\text{an interval where the mean number is } x \\ &= \left\langle \left[\mathcal{N}\left(y + \tfrac{1}{2}x\right) - \mathcal{N}\left(y - \tfrac{1}{2}x\right) - x\right]^2 \right\rangle \\ &= \frac{2}{\pi^2}\int_0^\infty d\tau \frac{K(\tau)}{\tau^2}\sin^2(\pi x\tau) \\ &= x + 2\int_0^x dy\,(x-y)\,[R(y) - 1]. \end{aligned} \tag{4.4}$$

The correlation function (4.1) is determined by the spectral density fluctuations, for which there is the semiclassical formula (2.16). Our aim in this section is to explain how to employ this observation to calculate these bilinear statistics, obtaining not only the universal random-matrix limit but also the corrections to this corresponding to large eigenvalue or zero separations, or short times. The argument is subtle and has several levels of refinement, of which we start with the simplest [3, 5, 33].

We will calculate $K(\tau)$. The first step is to substitute (2.16) into (4.1), thereby obtaining a double sum over periodic orbits. Since all the actions are positive, we can simplify the averages (over a small interval of eigenvalues or along the critical line) using

$$\langle\cos\{S_j/\hbar\}\cos\{S_k/\hbar\}\rangle = \tfrac{1}{2}\langle\cos\{(S_j - S_k)/\hbar\}\rangle. \tag{4.5}$$

The dimensionless intervals x that we will be considering may be large but must correspond to classically small energy ranges, so we can approximate the actions using

$$S_j\left(E_0 \pm \frac{x}{2\langle d\rangle}\right) = S_j \pm \frac{xT_j}{2\langle d\rangle} + O\left(\frac{x^2}{\langle d\rangle^2}\right), \tag{4.6}$$

where S_j, T_j, and d are evaluated at E_0. Elementary manipulations, and evaluating the integral in (4.2), give the asymptotic (that is, small-$\hbar$) form factor as the double sum

$$(4.7)\; K(\tau) = \frac{1}{4(\pi\langle d\rangle\hbar)^2}\left\langle \sum_j \sum_k A_j A_k \cos\{(S_j - S_k)/\hbar\}\,\delta\left(|\tau| - \frac{T_j + T_k}{4\pi\langle d\rangle\hbar}\right)\right\rangle.$$

It is convenient now to consider separately the diagonal part $K_{\rm diag}$ of the sum (terms with $j = k$) and the off-diagonal part $K_{\rm off}$ (terms with $j \neq k$). For $K_{\rm diag}$, we have

$$K_{\rm diag}(\tau) = \frac{1}{4(\pi\langle d\rangle\hbar)^2}\sum_j A_j^2 \delta\left(|\tau| - \frac{T_j}{2\pi\langle d\rangle\hbar}\right). \tag{4.8}$$

In the limit $\hbar \to 0$, τ fixed, the sum over orbits can be evaluated using the Hannay-Ozorio sum rule (3.2), giving

$$\lim_{\hbar\to 0} K_{\rm diag}(\tau) = |\tau|\,. \tag{4.9}$$

This is universal: all details of the specific dynamics have disappeared. Because of the Riemann-quantum analogy, the same behaviour should hold for the pair correlation of the Riemann zeros. Here we make contact with the seminal work of Montgomery [35], who indeed proved (4.9) in that case.

Now we observe that in random-matrix theory the exact form factor of the GUE is

$$K_{\rm GUE}(\tau) = |\tau|\,\Theta(1 - |\tau|) + \Theta(|\tau| - 1)\,. \tag{4.10}$$

(Θ is the unit step.) For later reference, the GUE pair distribution function, obtained from (4.2), is

$$R_{\rm GUE}(x) = 1 - \left(\frac{\sin(\pi x)}{\pi x}\right)^2. \tag{4.11}$$

Evidently the approximation (4.9), based on periodic orbits, captures exactly the random-matrix behaviour for $|\tau| < 1$, without invoking any random matrices. This led Montgomery [35] to conjecture (following a suggestion of Dyson and independently of any semiclassical argument) that for the Riemann zeros $K(\tau) = K_{\rm GUE}(\tau)$ in the limit $t \to \infty$.

Clearly, (4.9) does not give the random-matrix result when $|\tau| > 1$. Indeed it fails drastically by not satisfying the requirement, necessary for any form factor representing a discrete set of points (eigenvalues or zeros), that $K(\tau) \to 1$ as $\tau \to \infty$. This failure reflects the importance of $K_{\rm off}$, and implies that for large τ (long orbits) the off-diagonal terms in the double sum (4.7) cannot vanish through incoherence, as might naively be thought, but must conspire by destructive coherent interference to cancel the term τ from $K_{\rm diag}$ and replace it by 1. This is consistent with the Montgomery conjecture, which implies

$$K_{\rm off}(\tau) = \Theta(|\tau| - 1)(1 - |\tau|) \tag{4.12}$$

in the limit $t \to \infty$.

One reason why $K_{\rm diag}$ alone is inadequate is the proliferation of orbits: for sufficiently long times, there will be many pairs of orbits whose actions differ by less than $\hbar$, so that they cannot be regarded as incoherent in (4.7). This phenomenon, that in some appropriate sense the large-τ limit of the double sum must be 1, is the *semiclassical sum rule*. Originally [33] the rule was obtained by a different argument, and was mysterious. Now there is a better understanding of the mechanism by which the cancellation occurs [36, 37]; we will discuss it later.

Indeed, for the Riemann zeros, (4.12) can be derived [4] using a conjecture of Hardy and Littlewood [38] concerning the pair distribution of the prime numbers. These correlations are important because if the logarithms of the primes (primitive orbit periods) were pairwise uncorrelated, K_{off}, being the average of a sum of random phases, would be zero. The Hardy-Littlewood conjecture is that $\pi_2(k;X)$, defined as the number of primes $p \le X$ such that $p+k$ is also a prime, has the following asymptotic form for large X:

$$\pi_2(x) \sim \frac{X}{\log^2 X} C(k) \tag{4.13}$$

with

$$C(k) = \begin{cases} 0 \quad \text{if } k \text{ is odd} \\ 2\prod_{q>2}\left(1-\frac{1}{(q-1)^2}\right)\prod_{\substack{p>2\\ p|k}}\left(\frac{p-1}{p-2}\right) \quad \text{if k is even,} \end{cases} \tag{4.14}$$

where the q-product includes all odd primes, and the p-product includes all odd prime divisors of k. Pairwise randomness would correspond to $C(k)=1$. It can be demonstrated [4] that as $K\to\infty$

$$\sum_{k=1}^{K} C(k) \sim K - \tfrac{1}{2}\log K \tag{4.15}$$

and so on average

$$C(k) \sim 1 - \frac{1}{2|k|} \tag{4.16}$$

for large k. This in turn was shown to imply (4.12) in the limit $t\to\infty$ [4].

We have seen that K_{diag} is universal in the limit $\hbar\to 0$; that is, it is independent of the specific features of the dynamics. These reappear—in a dramatically nonuniform way—in the approach to the limit. To see this, note first that it is only for short orbits, that is, when $\tau << 1$, that universality breaks down. Next, choose a τ^* corresponding to a time much longer than the shortest period T_0 and shorter than the Heisenberg time T_H, that is,

$$\frac{T_0}{2\pi\langle d\rangle\hbar} << \tau^* < 1. \tag{4.17}$$

We continue to use the Hannay-Ozorio sum rule for $\tau > \tau^*$, the limit (4.12) for K_{off} ensuring the correct GUE formula (4.10) for $\tau > 1$, but take the contributions from orbits with period $T_j < 2\pi\langle d\rangle\hbar\tau^*$ directly from (4.8). Thus

$$\begin{aligned} K(\tau) &\approx K_{\text{GUE}}(\tau) \\ &\quad + \frac{1}{4(\pi\langle d\rangle\hbar)^2}\sum_{T_j<2\pi\langle d\rangle\hbar\tau^*} A_j^2\delta\left(|\tau| - \frac{T_j}{2\pi\langle d\rangle\hbar}\right) - |\tau|\,\Theta(\tau^*-|\tau|) \end{aligned} \tag{4.18}$$

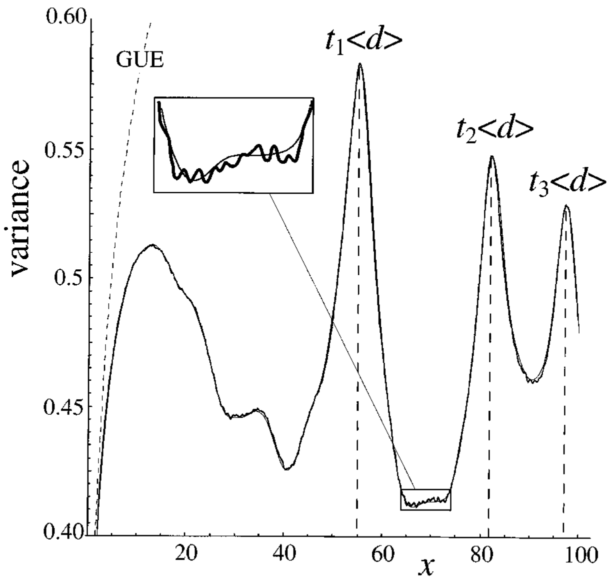

Fig. 3 *Number variance* $\sum^2(x)$ (4.4) *of the Riemann zeros* t_n *near* $n = 10^{12}$*, calculated from* (4.18) *(with* $\tau^* = 1/4$*),* (2.14)*, and* (2.18) *(thin line), compared with* $\sum^2(x)$ *computed from numerically calculated zeros by Odlyzko* [39, 40] *(thick line); all the zeros are close to* $t = 2.677 \times 10^{11}$*, and their smoothed density is* $\langle d \rangle = 3.895\ldots$*. Note the resurgence resonances (cf.* (4.23)*) associated with the lowest zeros* t_1*,* t_2*, and* t_3*, and that the theory fails to capture small, fast oscillations in the data.*

is a candidate for a semiclassical formula for the form factor. Later we will see that this is not quite correct: the proper incorporation of the off-diagonal terms in the double sum introduces a small but important modification near $\tau = 1$. For the moment, we continue to discuss (4.18).

This formula for $K(\tau)$, applied to the Riemann zeros, is extremely accurate. When employed in conjunction with (4.4) to calculate the number variance of the zeros [34], it reproduces almost perfectly this statistic as computed from numerical values of high zeros [39, 40]. Figure 3 shows that the agreement extends from the random-matrix regime (small x) to the far nonuniversal regime. Note however the tiny oscillatory deviations; we will return to these later.

For the pair correlation, we have

$$R(x) = R_{\mathrm{GUE}}(x) + R_c(x). \tag{4.19}$$

Remarkably, it is possible to calculate the correction R_c explicitly and in closed form at this level of approximation. The formula was obtained for both the Riemann zeros and for general systems in [41], and independently in [42] for the Riemann zeros. From (4.18), (4.2), and (2.14), we get

$$\begin{aligned} R_c(x) \approx R_c^1(x) = {} & \frac{1}{2(\pi\langle d\rangle)^2} \sum_{\substack{m,p \\ p^m < \exp(2\pi\langle d\rangle\tau^*)}} \frac{\log^2 p}{p^m}\cos\left\{\frac{xm\log p}{\langle d\rangle}\right\} \\ & -2\int_0^{\tau^*} d\tau \cos\{2\pi x\tau\}, \end{aligned} \tag{4.20}$$

where $\langle d\rangle$ is given by (2.15). The sum is insensitive to the value of τ^* provided this is not too small, so we set $\tau^* = \infty$. Next, we write

$$\log^2 p = \frac{\log^2(p^m)}{m} + (1-m)\log^2 p. \tag{4.21}$$

This corresponds to separating R_c^1 into contributions from primitive orbits (first term) and repetitions (second term). In the repetitions, the sum over m can be evaluated explicitly. For the first term, we use [7]

$$J'(w) = \sum_{p,m} \frac{\delta(w-p^m)}{m} = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} ds w^{s-1}\log\zeta(s) \quad (a>1). \tag{4.22}$$

Some tricky but elementary manipulations now give

$$R_c^1(x) = \frac{1}{2(\pi\langle d\rangle)^2}\left[\frac{1}{\xi^2} - \partial_\xi^2 \mathrm{Re}\log\zeta(1-i\xi) - \mathrm{Re}\sum_{p=2}^{\infty}\frac{\log^2 p}{(p\exp\{i\xi\log p\}-1)^2}\right], \tag{4.23}$$

where

$$\xi \equiv \frac{x}{\langle d\rangle}. \tag{4.24}$$

This formula has a very interesting structure, worth discussing in detail. First, $\xi \to 0$ in the limit $t \to \infty$ for any fixed x, and so the pole in the zeta function cancels the singularity $1/\xi^2$. Second, the prefactor $1/\langle d\rangle^2$ ensures that the correction R_c^1 is asymptotically small in comparison with R_{GUE} (equation (4.11)). Third, the dependence on ξ shows that R_c^1 involves the separation between zeros in the original variable $t = \mathrm{Im}s$ (heights of zeros along the critical line), rather than the scaled separation x; this means that structural features of R_c^1 appear asymptotically at larger x than the oscillations in R_{GUE}, as expected for nonuniversal features of correlations. Fourth, the contributions from repetitions (the sum over p in (4.23)) are less significant than those from primitive orbits (first two terms), as Figure 4 shows. Fifth, and most important, the appearance of $\zeta(1-i\xi)$ indicates an astonishing resurgence property of the zeros: in the pair correlation of high Riemann zeros, the low Riemann zeros appear as resonances. This is illustrated in Figure 5. The resonances also appear as peaks in the nonuniversal part of the number variance (Figure 3).

For generic dynamical systems without time-reversal symmetry, it can be verified directly that the analogue of (4.23) is [41, 43]

$$R_c^1(x) = \frac{1}{2(\pi\hbar\langle d\rangle)^2} \times \left[\frac{1}{\xi^2} - \partial_\xi^2\mathrm{Re}\log\zeta_D(i\xi) - \mathrm{Re}\sum_p\sum_{m=0}^{\infty}\frac{(m+1)T_p^2}{\left(|\Lambda_p|\Lambda_p^m\exp\{-i\xi T_p\}-1\right)^2}\right], \tag{4.25}$$

where ζ_{D} is the dynamical zeta function defined in (3.3), and now $\xi = x/\hbar\langle d\rangle$. Again, the pole in the zeta function (now at $s=0$) cancels the singularity $1/\xi^2$. In this case, the resonances discussed above are caused by singularities of $\log\zeta_{\mathrm{D}}(s)$ away from $s=0$, that is, by subdominant eigenvalues of the Perron–Frobenius operator.

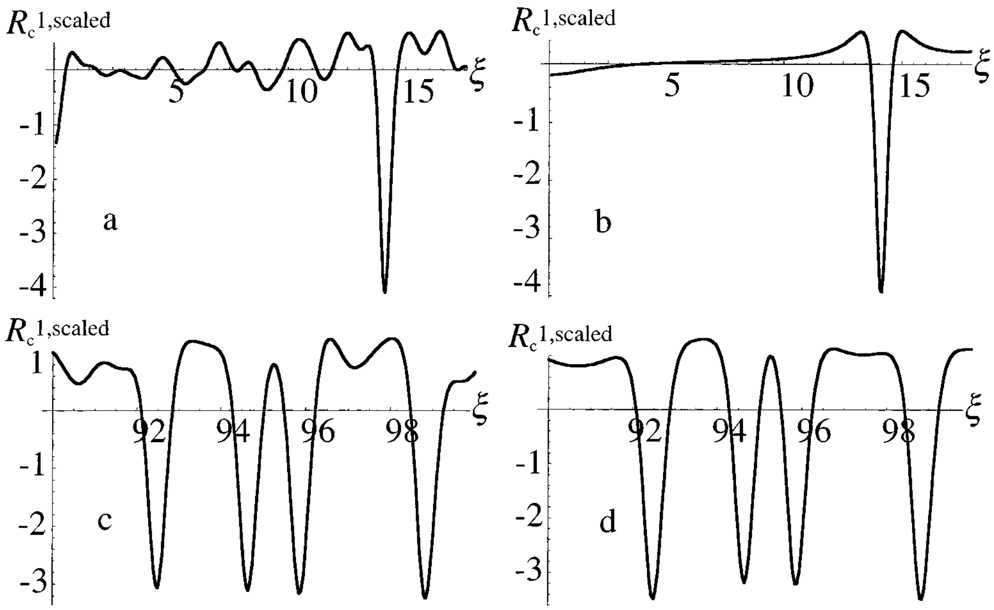

Fig. 4 *Nonuniversal correction to the pair correlation of the Riemann zeros, calculated from* (4.23) *as* $R_c^{1\ scaled}(\xi) \equiv 2(\pi\langle d\rangle)^2 R_c(x)$. *Parts* (a) *and* (c) *include repetitions;* (b) *and* (d) *omit repetitions.*

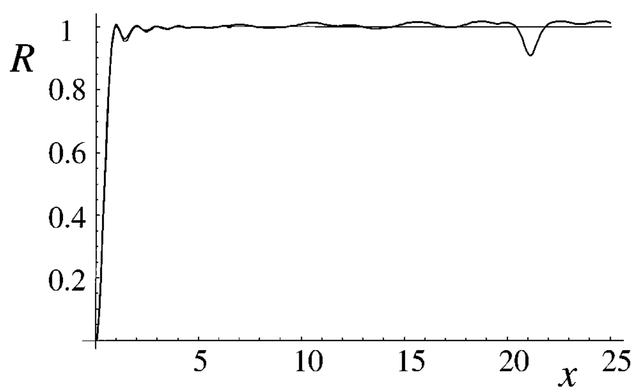

Fig. 5 *Pair correlation* $R(x)$ *of the Riemann zeros, calculated "semiclassically" (thick line) from* (4.19) *and* (4.23), *for zeros near* $n = 10^5$, *and random-matrix behaviour* $R_{GUE}(x)$ *(thin line); note the first nonuniversal resurgence resonance near* $x = 21$.

Now we return to the tiny oscillatory deviations noticeable in Figure 3, reflecting small errors in (4.19) and (4.23). These are again associated with the approach to the $t \to \infty$ limit of the form factor, rather than the limit itself: whereas (4.23) captures the appropriate large-t asymptotics of K_{diag}, the GUE-motivated replacement (4.12) incorporates only the $t \to \infty$ limit of K_{off}.

For the Riemann zeta function, this can be corrected as follows. We have already noted above that the formula (4.12) for K_{off} can be derived using the smoothed expression (4.16) for the Hardy-Littlewood conjecture. The large-t asymptotics we

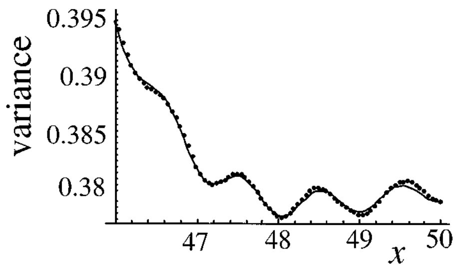

Fig. 6 *Number variance* $\sum^2(x)$ (4.4) *of the Riemann zeros* t_n *near* $n = 10^9$*, calculated from* (4.19) *and* (4.26)*, including the off-diagonal correction* (4.27)–(4.28) [70] *(full line), compared with* $\sum^2(x)$ *computed from numerically calculated zeros by Odlyzko* [39, 40] *(dots); all the zeros are close to* $t = 3.719 \times 10^8$*, and their smoothed density is* $\langle d \rangle = 2.848\ldots$.

seek comes from using the original unsmoothed form (4.14) [5, 41]. The result is that

$$R_c(x) = R_c^1(x) + R_c^2(x), \tag{4.26}$$

in which R_c^1 is given by (4.23), and

$$\begin{aligned} R_c^2(x) \approx \frac{1}{2(\pi\langle d\rangle)^2}\Bigg[&-\frac{\cos(2\pi x)}{\xi^2} \\ &+ |\zeta(1+i\xi)|^2 \operatorname{Re}\{\exp(2\pi i x)\, b(\xi)\}\Bigg], \end{aligned} \tag{4.27}$$

where

$$b(\xi) = \prod_p \left(1 - \frac{(p^{i\xi}-1)^2}{(p-1)^2}\right) \tag{4.28}$$

is a convergent product over the primes. As with the diagonal term (cf. the discussion after (4.24)), convergence as $\xi \to 0$ is ensured by the pole of the zeta function.

This second correction, although small, does incorporate the small oscillations, through the trigonometric functions with argument $2\pi x$. Asymptotically (that is, as $t \to \infty$), these oscillations are fast (cf. (4.23)–(4.24)) in comparison with the variations from the resonances of the zeta function. When employed in conjunction with (4.4), the correction accurately reproduces the oscillatory deviation (Figure 3) in the number variance of the zeros; this is illustrated in Figure 6.

Unfortunately, this derivation of (4.27) for the Riemann zeros cannot be imitated for general chaotic dynamical systems because we have no a priori knowledge of the correlations between the actions of different periodic orbits, analogous to the Hardy-Littlewood conjecture for the primes. It is possible to get some information by working backwards and, assuming that the GUE expression (4.10) or (4.11) describes the pair correlation of eigenvalues in generic chaotic systems without time-reversal symmetry, deriving the universal limiting form of the implied action correlations [37]. (This procedure essentially follows an analogous derivation for the primes themselves, assuming the Montgomery conjecture [44]). An interesting feature of this approach is that it leads to predictions about *classical* trajectories based on the distribution of *quantum* energy levels. However, it gives no information about the deviations from random-matrix universality that are the focus of our concern here.

Recently a theory has been developed that overcomes these difficulties [5, 41]. It is based on two observations.

First, as already noted above, quantum eigenvalues (or Riemann zeros) are resolved by the trace formula if the sum (2.9) over periodic orbits is truncated near the Heisenberg time T_H (this will be made more precise in the next section). Hence, if the trace formula thus truncated generates the approximation

$$\tilde{\mathcal{N}}(E) = \langle \mathcal{N}(E) \rangle + \tilde{\mathcal{N}}_{\mathrm{fl}}(E) \tag{4.29}$$

to the counting function, the quantities $\tilde{E}_n$ defined by

$$\tilde{\mathcal{N}}\left(\tilde{E}_n\right) = n + \tfrac{1}{2} \tag{4.30}$$

should be good semiclassical approximations to the exact eigenvalues. The theory is based on calculating the correlations in this approximate spectrum.

Second, the diagonal terms $K_{\mathrm{diag}}(\tau)$ are asymptotically dominant in the form factor for $\tau < 1$, corresponding to times less than T_H. This implies that orbits with periods less than T_H make contributions that are effectively uncorrelated; treating them in this way allows the correlations in the $\tilde{E}_n$ spectrum to be computed exactly.

For chaotic systems without time-reversal symmetry, the result is that when $x >> 1$ the deviations from the GUE formula can also be represented in the form (4.26), (4.25), and (4.27), where $\zeta(1 + i\xi)$ is replaced by $\zeta_{\mathrm{D}}(i\xi)$ (defined by (3.3)) and, in (4.28), $b(\xi)$ is replaced by

$$\begin{aligned} b(\xi) = \frac{1}{\gamma^2} \prod_p {}_2\phi_1 \{ &\exp(-iT_p\xi), \exp(-iT_p\xi); \\ &\Lambda_p^{-1}; \Lambda_p^{-1}, \left|\Lambda_p^{-1}\right| \exp(iT_p\xi) \} \frac{\left|\zeta_D^{(p)}(0)\right|^2}{\left|\zeta_D^{(p)}(i\xi)\right|^2} \end{aligned} \tag{4.31}$$

(again $\xi = x/\hbar\langle d \rangle$). Here γ is the residue of the pole at $s = 0$ of $\zeta_D(s)$, ${}_2\phi_1$ is the q-hypergeometric function [45], and ζ_D^p is the pth element of the product over primitive orbits in (3.3).

The formal similarity between the results for the Riemann zeros and for the semiclassical eigenvalues is striking, and reinforced by the fact that the derivation of (4.31) just outlined leads precisely to (4.28) when applied to the zeros. Indeed, by Fourier-transforming (4.26) with respect to t, this can be regarded as a heuristic derivation of the Hardy-Littlewood conjecture. In the same way, Fourier-transforming the corresponding result for dynamical systems with respect to $1/\hbar$ leads to a classical periodic orbit correlation function corresponding directly to the Hardy-Littlewood conjecture and reducing to the universal form conjectured in [37] in the long-time limit. It is a challenge to derive these correlations within classical mechanics.

We finish this section on connections between statistics of the Riemann zeros and quantum eigenvalues by remarking that the results for pair correlations extend to correlations of higher order. Thus Montgomery's conjecture for the two-point correlation of the Riemann zeros generalizes to all n-point correlations. Specifically, the irreducible n-point correlation function

$$\tilde{R}_n(x_1, x_2, \ldots, x_n) \equiv \frac{1}{\langle d \rangle^n} \left\langle \prod_{i=1}^{n} d_{fl}\left(t + \frac{x_i}{\langle d \rangle}\right) \right\rangle \tag{4.32}$$

tends asymptotically to the corresponding GUE expression:

$$\lim_{t\to\infty} \tilde{R}_n\left(x_1, x_2 \ldots, x_n\right) = \det \mathbf{S}, \tag{4.33}$$

where the elements S_{ij} of the $n \times n$ matrix $\mathbf{S}$ are given by

$$S_{ij} = s\left(x_i, x_j\right) = \frac{\sin\left\{\pi\left(x_i - x_j\right)\right\}}{\pi\left(x_i - x_j\right)}\left(1 - \delta_{ij}\right). \tag{4.34}$$

The analogue of Montgomery's theorem for the diagonal contributions to $\tilde{R}_n$ was proved for $n = 3$ [46] and then for all $n \geq 2$ [47]. The off-diagonal contributions were calculated using a generalization of the Hardy-Littlewood conjecture for $n = 3$ and $n = 4$ [48] and then for all $n \geq 2$ [49]. In all cases the results confirm the conjecture (4.33) and (4.34). The nonuniversal deviations from the GUE formulae (4.33)–(4.34) were calculated for $n = 3$ and $n = 4$ [41] using the method outlined above, and take a form (related to the structure of $\zeta(s)$ as $s \to 1$) directly analogous to that already discussed. As expected, this extends to the higher order correlations of quantum eigenvalues.

5. Riemann-Siegel Formulae. A powerful stimulus to the development of analogies between quantum eigenvalues and the Riemann zeros has been the Riemann-Siegel formula for $\zeta(s)$. As explained in [7], this very effective way of computing the zeros (especially high ones)—employed in most numerical computations nowadays—was discovered by Siegel in the 1920s among papers left by Riemann after his death 60 years earlier. We present the formula in an elementary way, chosen to facilitate our subsequent exploration of its intricate interplay with quantum mechanics. Riemann's derivation [11, 50] was different, and a remarkable achievement, because although it was one of the first applications of his method of steepest descent for integrals it was more sophisticated than most applications today, in that the saddle about which the integrand is expanded is accompanied by an infinite string of poles.

It is a consequence of the functional equation satisfied by $\zeta(s)$ [11] that the following function $Z(t)$ is even, and real for real t:

$$Z\left(t\right) \equiv \exp\left\{i\theta\left(t\right)\right\}\zeta\left(\tfrac{1}{2} + it\right). \tag{5.1}$$

Here $\theta(t)$ is the function appearing in the smoothed counting function (2.3) for the zeros. Naive substitution of the Dirichlet series (1.5) gives the formal expression

$$Z\left(t\right) = \exp\left\{i\theta\left(t\right)\right\}\sum_{n=1}^{\infty}\frac{\exp\left\{-it\log n\right\}}{n^{1/2}}. \tag{5.2}$$

This is doubly unsatisfactory. First, it does not converge—a defect shared with its relative (2.6) for $\mathcal{N}_{\mathrm{fl}}(t)$ (cf. (2.4)) and similarly originating in the inadmissibility of (1.5) in the critical strip. Second, it is not manifestly real as $Z(t)$ must be.

Both defects can be eliminated by truncating the series (5.2) at a finite $n = n^*(t)$ and resumming the tail. The truncation $n^*(t)$ is chosen to be the term whose phase $\theta(t) - t\log n$ is stationary with respect to t; the asymptotic formula for θ (last member of (2.3)) gives

$$n^*\left(t\right) = \mathrm{Int}\left(\sqrt{\frac{t}{2\pi}}\right). \tag{5.3}$$

A crude resummation [1] using the Poisson summation formula leads to a result equivalent to the "approximate functional equation" [11]:

$$Z(t) = 2\sum_{n=1}^{n^*(t)} \frac{\cos\{\theta(t) - t\log n\}}{n^{1/2}} + \cdots. \tag{5.4}$$

This is a remarkable example of resurgence: the resummed terms in the tail $n > n^*(t)$ are the complex conjugates of the early terms $1 \le n \le n^*(t)$, so that the series in (5.4)—called the "main sum" of the Riemann-Siegel expansion—is real, like the exact $Z(t)$. The zeros generated by the first term alone ($n = 1$), that is, $\cos\theta(t) = 0$, have the correct mean density (cf. (2.3)). Higher terms shift the zeros closer to their true positions, and introduce the random-matrix fluctuations. It is worth mentioning that the zeros obtained by including successive terms in (5.4) cannot be regarded as the eigenvalues of hermitean operators that approximate the still-unknown Riemann operator, because these partial sums of the main sum each have zeros for complex t [6].

Unfortunately, the truncation (5.3) introduces another defect: the sum is a discontinuous function of t, unlike $Z(t)$, which is analytic. The discontinuities can be eliminated by formally expanding the difference between (5.2) and the sum in (5.4) about the truncation limit $N(t)$, to obtain the correction terms in (5.4). This will depend on the fractional part of $\surd(t/2\pi)$ as well as its integer part n^*, so it is convenient to define

$$a(t) \equiv \sqrt{\frac{t}{2\pi}} \equiv n^*(t) + \frac{1}{2}(1 - z(t)). \tag{5.5}$$

The expansion is in powers of $1/a$ (henceforth we do not write the t-dependences explicitly), and gives

$$Z(t) = 2\sum_{n=1}^{n^*} \frac{\cos\{\theta(t) - t\log n\}}{n^{1/2}} + \frac{(-1)^{n^*+1}}{a^{1/2}} \sum_{r=0}^{\infty} \frac{C_r(z)}{a^r}. \tag{5.6}$$

This procedure was devised in [4], where it was used to calculate the first correction term $C_0(z)$, and elaborated in [51] in a study of the higher corrections.

The sum over r is the Riemann-Siegel expansion. Its terms $C_r(z)$ are constructed from derivatives (up to the $3r$th) of

$$C_0(z) = \frac{\cos\left\{\frac{1}{2}\pi\left(z^2 + \frac{3}{4}\right)\right\}}{\cos\{\pi z\}} \tag{5.7}$$

with coefficients determined by an explicit recurrence relation involving the coefficients (Bernoulli numbers) in the Stirling expansion of $\theta(t)$ for large t. The next few coefficients are

$$\begin{aligned}
C_1(z) &= \frac{C_0^{(3)}(z)}{12\pi^2}, \\
C_2(z) &= \frac{C_0^{(2)}(z)}{16\pi^2} + \frac{C_0^{(6)}(z)}{288\pi^4}, \\
C_3(z) &= \frac{C_0^{(1)}(z)}{32\pi^2} + \frac{C_0^{(5)}(z)}{120\pi^4} + \frac{C_0^{(9)}(z)}{10368\pi^6}, \\
C_4(z) &= \frac{C_0(z)}{128\pi^2} + \frac{19C_0^{(4)}(z)}{1536\pi^4} + \frac{11C_0^{(8)}(z)}{23040\pi^6} + \frac{C_0^{(12)}(z)}{497664\pi^8}
\end{aligned} \tag{5.8}$$

(superscripts in brackets denote derivatives). Gabcke [50] calculated $C_r(z)$ for $r \leq 12$. Later terms get very complicated; for example,

$$
\begin{aligned}
C_{20}(z) = {} & \frac{332727711 C_0(z)}{274877906944\ \pi^{10}} + \frac{117753804989 C_0^{(4)}(z)}{3298534883328\ \pi^{12}} \\
& + \frac{13899745416281 C_0^{(8)}(z)}{692692325498880 \pi^{14}} + \frac{311274631265011 C_0^{(12)}(z)}{164583696538533888\ \pi^{16}} \\
& + \frac{24311037030485304 17 C_0^{(16)}(z)}{449313491550197514 24000 \pi^{18}} \\
& + \frac{2325442687388622149 41 C_0^{(20)}(z)}{3731869485532640496844 80000 \pi^{20}} \\
& + \frac{361888761444289010497 C_0^{(24)}(z)}{1064899933783461120599654 40000 \pi^{22}} \\
& + \frac{665406310453227159231 77 C_0^{(28)}(z)}{684304697449252116097337917 4400000 \pi^{24}} \\
& + \frac{391261681973226653 C_0^{(32)}(z)}{2505753945351719007289344 0000000 \pi^{26}} \\
& + \frac{1259995823308801 C_0^{(36)}(z)}{8571719378669228821315584 0000000 \pi^{28}} \\
& + \frac{713214794639 C_0^{(40)}(z)}{8571719378669228821315584 0000000 \pi^{30}} \\
& + \frac{50407933481 C_0^{(44)}(z)}{17650884544555675988853050572 800000 \pi^{32}} \\
& + \frac{1039499 C_0^{(48)}(z)}{1768363201124316332834999500 800000 \pi^{34}} \\
& + \frac{22411 C_0^{(52)}(z)}{32139197378979392841852100018176 0000 \pi^{36}} \\
& + \frac{59 C_0^{(56)}(z)}{13636202316509828105757248150568960000 \pi^{38}} \\
& + \frac{C_0^{(60)}(z)}{9327162384492722424337957734989168640000 \pi^{40}}.
\end{aligned}
\tag{5.9}
$$

An elaborate asymptotic analysis [51] shows that the high orders ("asymptotics of the asymptotics") can be represented compactly as a "decorated factorial series" whose terms are

$$
C_r(z) = \frac{\Gamma\left(\frac{1}{2}r\right)}{\left(\pi\sqrt{2}\right)^{r+1}} f(r,z), \tag{5.10}
$$

where for large r

$$
\begin{aligned}
f(r,z) \sim {} & \sum_{m=0}^{\infty} (-1)^{m(m-1)/2} \exp\left\{-\left(m+\frac{1}{2}\right)^2\right\} \\
& \times \left\{ \begin{array}{ll} \sin\left\{(2m+1)\sqrt{r}\right\} \cos\left\{\left(m+\frac{1}{2}\right)\pi z\right\} & (r \text{ even}) \\ \cos\left\{(2m+1)\sqrt{r}\right\} \sin\left\{\left(m+\frac{1}{2}\right)\pi z\right\} & (r \text{ odd}) \end{array} \right\}.
\end{aligned}
\tag{5.11}
$$

Comparison with numerically computed $C_r(z)$ (up to $r = 50$, using special techniques to evaluate the derivatives of $C_0(z)$) shows that these formulae capture the fine details of the Riemann-Siegel coefficients, even for small r.

The factorial in (5.10) means that the sum over r in (5.6) is divergent in the manner familiar in asymptotics: the terms get smaller and then diverge. Asymptotics folklore suggests, and Borel summation (implemented analytically and checked numerically) confirms, that optimal accuracy obtainable from the Riemann-Siegel formula (without further resummation) corresponds to truncating the sum at the least term. This has

$$
r^* = \mathrm{Int}(2\pi t) \tag{5.12}
$$

and the resulting error is of order

$$(5.13)\quad Z(t) - 2\sum_{n=1}^{n^*} \frac{\cos\{\theta(t) - t\log n\}}{n^{1/2}} - \frac{(-1)^{n^*+1}}{a^{1/2}} \sum_{r=0}^{r^*} \frac{C_r(z)}{a^r} = O(\exp\{-\pi t\}).$$

The accuracy is very high: even for the lowest Riemann zero, $r^*(t_1) = 89$ and $\exp\{-\pi t_1\} \sim 10^{-20}$. Nevertheless, it is possible to do better, as we shall see later.

Now we turn to the quantum analogues of the Riemann-Siegel formula for classically chaotic systems with $D > 1$, as envisaged in [1], explored in detail in [52], and derived in [53]. These studies are motivated by the hope that such an effective method of computing Riemann zeros might lead to a useful way to calculate quantum eigenvalues.

First, the counterpart of $Z(t)$ in (5.1) is a function with zeros at the quantum energy levels E_n; this is the quantum spectral determinant

$$(5.14)\quad \begin{aligned} \Delta(E) &= \prod_n A(E, E_n)(E - E_n) = \det\{A(E,H)(E-H)\} \\ &= \det A \exp\{\operatorname{trlog}(E-H)\}, \end{aligned}$$

where H is the hermitean wave operator (section 2) and the real factor A is introduced to make the product converge. Hermiticity implies that Δ is real for real E; this "quantum functional equation" is analogous to the functional equation for $\zeta(s)$, which implies that $Z(t)$ is real for real t.

To find the counterpart of the Dirichlet series (5.2), we note that the quantum eigenvalue counting function can be written (cf. (2.4)) as

$$(5.15)\quad \mathcal{N}(E) = -\frac{1}{\pi}\lim_{\varepsilon\to 0} \operatorname{Im}\operatorname{Trlog}\{1 - (E + i\varepsilon)/H\}.$$

Now the decomposition into smooth and fluctuating parts, together with the periodic-orbit formula (2.9), leads to

$$(5.16)\ \Delta(E) \sim B(E)\exp\{-i\pi\langle\mathcal{N}(E)\rangle\} \prod_p \exp\left\{-\sum_{m=1}^{\infty} \frac{\exp\{imS_p(E)/\hbar\}}{m\sqrt{\left|\det\left(\mathsf{M}_p^m - \mathsf{I}\right)\right|}}\right\},$$

where $B(E)$ is real and nonzero for real E and where we have absorbed the Maslov indices into S.

Expanding the product over primitive orbits p and the exponential of the sum over repetitions m, we obtain a series of terms that can be labelled by

$$(5.17)\quad n = \{0, 1, 2\ldots\} \Leftrightarrow \{m_p\} = \{m_1, m_2 \ldots\}.$$

Here m_p represents the number of repetitions of the orbit p. Each term corresponds to a sum over actions:

$$(5.18)\quad S_n(E) = \sum_p m_p S_p(E).$$

The expansions lead to

$$(5.19)\quad \Delta(E) \sim B(E)\exp\{-i\pi\langle\mathcal{N}(E)\rangle\} \sum_{n=0}^{\infty} D_n(E)\exp\{iS_n(E)/\hbar\}$$

with an explicit form for the coefficients D_n that we do not give here [52]. As (5.18) indicates, the terms n correspond to composite orbits, or pseudo-orbits, consisting of combinations of repetitions of different periodic orbits. We label the composite orbits so that increasing n corresponds to increasing period

$$\mathcal{T}_n(E) = \frac{\partial \mathcal{S}_n(E)}{\partial E} \tag{5.20}$$

with $n = 0$ representing no orbit at all, that is, $m_p = 0$ (for which the coefficient $D_0 = 1$).

The sum (5.19) is the counterpart of the Dirichlet series (5.2) for $Z(t)$, with composite orbits n related to primitive orbits p in the same way that the integers n are related to the primes p (cf. (1.5)). Moreover (5.19) diverges, like the sum (2.9) from which it was obtained, and it is not manifestly real as the exact $\Delta(E)$ must be. Our interpretation of the Riemann-Siegel formula suggests a similar resummation of the tail of the series (5.19) after truncation at the term whose phase is stationary with respect to E. This term—the counterpart of $n^*(t)$ in (5.3)—represents the composite orbit defined by

$$\frac{d}{dE}\left[\mathcal{S}_n(E)/\hbar - \pi\langle\mathcal{N}(E)\rangle\right] = 0. \tag{5.21}$$

The corresponding period $\mathcal{T}^*(E)$ is

$$\mathcal{T}^*(E) = \pi\hbar\langle d(E)\rangle = \tfrac{1}{2}T_H(E), \tag{5.22}$$

where $T_H(E)$ is the Heisenberg time (2.20).

Comparison with the Riemann-Siegel main sum in (5.4) suggests that the sum of the composite orbits with $\mathcal{T}_n > \mathcal{T}^*$ is, approximately, the complex conjugate of the sum of the orbits with $\mathcal{T}_n < \mathcal{T}^*$. In fact, this relation can be derived using arguments based on analytic continuation with respect to E [53]. These arguments also indicate a more detailed correspondence: between the sums of groups of terms with periods $\mathcal{T}^* + X$ and $\mathcal{T}^* - X$. The resulting "Riemann-Siegel lookalike" formula is

$$\Delta(E) \sim 2B(E)\sum_{\mathcal{T}_n < \mathcal{T}^*(E)} D_n(E)\cos\{S_n(E)/\hbar - \pi\langle\mathcal{N}(E)\rangle\} + \cdots. \tag{5.23}$$

(For a different derivation, see [54].)

With (5.23) it is possible to reproduce some low-lying quantum eigenvalues, and of course the fact that the sum is finite is a major advantage over the infinite divergent series (2.9) and (5.19). However, for a chaotic system with $D > 1$ the number of terms with $\mathcal{T}_n < \mathcal{T}^*$ is exponentially large in $1/\hbar$, so the Riemann-Siegel lookalike is not as useful for calculating high quantum eigenvalues as (5.4) is for calculating Riemann zeros. The origin of the difference is the exponential proliferation of periodic orbits (and composite orbits), together with the fact that $\langle d\rangle$ increases as $1/\hbar^D$, whereas for the Riemann zeros, whose classical counterpart appears to be quasi-one-dimensional, $\langle d\rangle$ increases as $\log t$. Moreover, (5.23) is discontinuous at the energies of composite orbits with period $\mathcal{T}^*$.

No way has yet been found to implement the obvious suggestion of cancelling the discontinuities in the quantum formula (5.23) by a series of corrections analogous to the terms involving $C_r(z)$ in the Riemann-Siegel expansion (5.6). However, a different completion of the Riemann-Siegel main sum was discovered ([55], generalizing an idea in [4]), that does have a quantum analogue.

In this alternative approach to the resummed Dirichlet series, the abrupt truncation is replaced by a smoothed cutoff involving the complementary error function and an optimization parameter K. An argument involving analytic continuation in t leads to

$$(5.24)\qquad \begin{aligned} Z(t) = 2\mathrm{Re}\sum_{n=1}^{\infty}\Bigg[&\frac{\exp\{i[\theta(t) - t\log n]\}}{n^{1/2}} \\ &\times \frac{1}{2}\mathrm{Erfc}\left\{(\log n - \theta'(t))\sqrt{\frac{t}{2(K^2 - i\theta''(t))}}\right\}\Bigg] + \cdots \end{aligned}$$

with an explicit expression for the correction terms. With K chosen appropriately, this smoothed sum can reproduce $Z(t)$ to an accuracy equivalent to that of the Riemann-Siegel main sum together with several correction terms. The corrections in (5.24) form an explicit asymptotic series enabling $Z(t)$ to be calculated with an accuracy of order $\exp(-t^2)$; this improvement over the Riemann-Siegel $\exp(-\pi t)$ is possible because (5.24) involves the higher transcendental function Erfc, whereas the Riemann-Siegel expansion involves only elementary functions. Several related representations of $Z(t)$ are now known [56, 57, 58].

The improved representation (5.24), together with the explicit correction terms, can readily be adapted to the quantum spectral determinant. The smoothed version of the Riemann-Siegel lookalike (5.23) is obtained by an argument involving analytic continuation with respect to $1/\hbar$, leading to

$$(5.25)\qquad \begin{aligned} \Delta(E) = 2B(E)\,\mathrm{Re}\sum_{n=0}^{\infty}\Bigg[&D_n(E)\exp\{i[\pi\langle\mathcal{N}(E)\rangle - S_n(E)/\hbar]\} \\ &\times\frac{1}{2}\mathrm{Erfc}\left\{\frac{S_n(E) - \pi\langle\mathcal{N}_1(E)\rangle}{2(K^2\hbar - i\pi\langle\mathcal{N}_2(E)\rangle)}\right\}\Bigg] + \cdots, \end{aligned}$$

where $\mathcal{N}_i$ denotes the ith derivative of $\mathcal{N}$ with respect to $1/\hbar$. A numerical test of this formula for the hyperbola billiard (a classically chaotic system with $D = 2$) shows that it can reproduce quantum eigenvalues with high accuracy, even resolving near-degenerate pairs of levels [59].

Finally, we note an important clue to the Riemann dynamics, hidden in the asymptotics (5.10), (5.11) of the Riemann-Siegel expansion (5.6). It concerns the implied small exponential $\exp\{-\pi t\}$ (cf. the error in (5.13)). The same exponential appears in the asymptotics of the gamma functions in $\theta(t)$ (equation (2.3)). Quantum mechanics suggests this is the "phase factor" corresponding to a periodic orbit with imaginary action (an "instanton" in physics jargon). If we write

$$(5.26)\qquad \exp\{iS\} = \exp\{-\pi t\}$$

(remembering $\hbar = 1$ for the Riemann zeros), the implied period is

$$(5.27)\qquad T = \frac{\partial S}{\partial'\mathrm{energy}'} = \frac{\partial S}{\partial t} = i\pi.$$

So, it seems that as well as the real periodic orbits in (2.14), with periods $m\log p$, there are complex periodic orbits, with periods that are multiples of $i\pi$.

6. Spectral Speculations. Although we do not know the conjectured Riemann operator H whose eigenvalues (all real) are the heights t_n of the Riemann zeros, the analogies presented so far suggest a great deal about it. To summarize:

a. H has a classical counterpart (the "Riemann dynamics"), corresponding to a hamiltonian flow, or a symplectic transformation, in a phase space.

b. The Riemann dynamics is chaotic, that is, unstable and bounded.

c. The Riemann dynamics does not have time-reversal symmetry. In addition, we note the recent discovery [60, 61] of modified statistics of the low zeros for the ensemble of Dirichlet L-functions, associated with a symplectic structure.

d. The Riemann dynamics is homogeneously unstable.

e. The classical periodic orbits of the Riemann dynamics have periods that are independent of "energy" t, and given by multiples of logarithms of prime numbers. In terms of symbolic dynamics, the Riemann dynamics is peculiar, and resembles Chinese: each primitive orbit is labelled by its own symbol (the prime p) in contrast to the usual situation where periodic orbits can be represented as words made of letters in a finite alphabet.

f. The Maslov phases associated with the orbits are also peculiar: they are all π. The result appears paradoxical in view of the relation between these phases and the winding numbers of the stable and unstable manifolds associated with periodic orbits [22], but finds an explanation in a scheme of Connes [62].

g. The Riemann dynamics possesses complex periodic orbits (instantons) whose periods are multiples of $i\pi$.

h. For the Riemann operator, leading-order semiclassical mechanics is exact: as in the case of the Selberg trace formula [21], $\zeta(1/2+it)$ is a product over classical periodic orbits, without corrections.

i. The Riemann dynamics is quasi-one-dimensional. There are two indications of this. First, the number of zeros less than t increases as $t\log t$; for a D-dimensional scaling system, with energy parameter $\alpha(E)$ proportional to $1/\hbar$, the number of energy levels increases as $\alpha(E)^D$. Second, the presence of the factor $p^{-m/2}$ in the counting function fluctuation formula (2.6), rather than the determinant in the more general Gutzwiller formula (2.9), suggests that there is a single expanding direction and no contracting direction.

j. The functional equation for $\zeta(s)$ resembles the corresponding relation—a consequence of hermiticity—for the quantum spectral determinant.

We have speculated [6] that the conjectured Riemann operator H might be some quantization of the following extraordinarily simple classical hamiltonian function $H_{\mathrm{cl}}(X,P)$ of a single coordinate X and its conjugate momentum P:

$$H_{\mathrm{cl}}(X,P) = XP. \tag{6.1}$$

Now we outline the reasons for this tentative association of XP with $\zeta(s)$.

At the *classical* level, (6.1) has a hyperbolic point at the origin in the infinite-phase (X,P) plane, and generates the following equations of motion and trajectories:

$$\dot{X} = X, \quad \text{i.e.,} \quad X(t) = X(0)\exp(t)\,; \qquad \dot{P} = -P, \quad \text{i.e.,} \quad P(t) = P(0)\exp(-t)\,. \tag{6.2}$$

Thus classical evolution is uniformly unstable, with stretching in X and contraction in P. Furthermore, the motion has the desired lack of time-reversal symmetry: velocity cannot be reversed ($\dot{X}$ is tied to X in (6.2)) and so the orbit cannot be retraced.

At the *semiclassical* level, we can try to estimate the smoothed counting function $\langle\mathcal{N}(E)\rangle$ of energy levels E_n generated by the quantum version of (6.1). For this it is necessary to specify a value of Planck's constant $\hbar$. We choose $\hbar=1$; other choices simply rescale the energies. $\langle\mathcal{N}(E)\rangle$ is the area $\mathcal{A}$ under the constant-energy hyperbola $E=XP$, measured in units of the "Planck cell" area $2\pi\hbar=2\pi$, with a

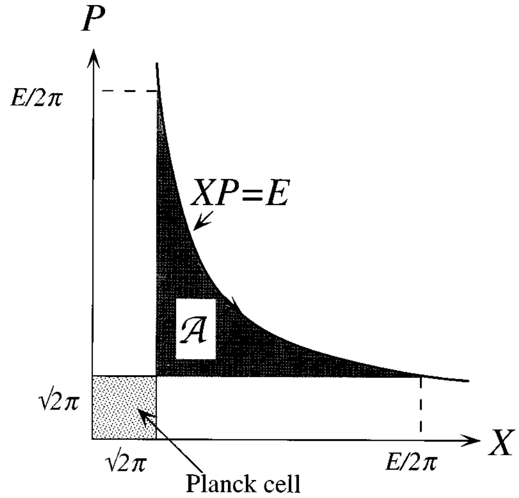

Fig. 7 *Phase space for* $H_{cl} = XP$*, with cutoffs for semiclassical regularization.*

Maslov index correction given by $\alpha/4\pi$, where α is the angle turned through along the orbit in phase space (this correction gives the "1/2" in the quantization of the harmonic oscillator). We encounter the immediate difficulty that $\mathcal{A}$ is infinite: motion generated by $H = XP$ is unbounded, and so does not give discrete quantum energies. As will be clear later, closing the phase space to make the motion bounded is a central unsolved problem. In the interim, a simple (perhaps the simplest) expedient is to regularize by truncating in X and P as indicated in Figure 7. The result (unaltered by representing the Planck cell by a rectangle instead of a square) is that $\langle \mathcal{N}(E) \rangle$ is precisely the asymptotics of the smoothed counting function for the Riemann zeros (last member of (2.3)), including the term 7/8, with t replaced by the energy E.

At the *quantum* level, the simplest formally hermitean operator corresponding to (6.1) is

$$H = \frac{1}{2}(XP + PX) = -i\left(X\frac{d}{dX} + \frac{1}{2}\right). \tag{6.3}$$

The formal eigenfunctions, satisfying

$$H\psi_E(X) = E\psi_E(X) \tag{6.4}$$

are

$$\psi_E(X) = \frac{A}{X^{1/2-iE}}. \tag{6.5}$$

We note the appearance of the power X^{-s} appearing in the Dirichlet series for $\zeta(s)$ (as integer^{-s}) and the Euler product (as prime^{-s}), with the symmetrization (6.3) placing s on the critical line.

It is evident that XP is simply a canonically rotated version of the inverted harmonic oscillator $P^2 - X^2$, which in turn is a complexified version of the usual harmonic oscillator $P^2 + X^2$. Some of these connections have been noted before [63, 64, 65, 66, 67]. The first-order operator XP is the simplest representative of this class, with the monomials (6.5) avoiding the complications of the parabolic cylinder eigenfunctions of $P^2 - X^2$.

To evaluate the corresponding momentum eigenfunction $\phi_E(P)$ (Fourier transform of (6.5)), it is necessary to specify a continuation across $X = 0$. The simplest choice, for a reason to be given later, is to make the wavefunction even in X, that is, to replace X by $|X|$. Then

$$\begin{aligned} \phi_E(P) &\equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dX \psi_E(X) \exp(-iPX) \\ &= \frac{A}{|P|^{1/2+iE}} 2^{iE} \frac{\Gamma\left(\frac{1}{4}+\frac{1}{2}iE\right)}{\Gamma\left(\frac{1}{4}-\frac{1}{2}iE\right)} \\ &= \frac{A}{\sqrt{2\pi}\,|P/2\pi|^{1/2+iE}} \exp\{2i\theta(E)\}. \end{aligned} \tag{6.6}$$

It follows that, up to factors that can easily be made symmetrical, the position and momentum eigenfunctions are each other's time-reverses. Thus we find a physical interpretation of the function $\theta(t)$ (defined in (2.3)) at the heart of the functional equation (cf. (5.1)) for $\zeta(s)$.

The major problem remaining is to find boundary conditions that would convert XP into a well-defined hermitean operator with discrete eigenvalues. This is equivalent to specifying the way in which parts of the (X, P) plane are connected so as to compactify the (quantum and classical) motion. Some hints in this direction follow.

Our observations about the complex periodic orbits of the Riemann dynamics (see the last paragraph of section 5) suggest that X and $-X$ should be identified. The reason is that the complex orbits of X, obtained by replacing t by $i\tau$ in (6.2), have period $2\pi i$, which becomes the desired $i\pi$ (equation (5.27)) on identifying $\pm X$.

To proceed further, we consider the symmetries of XP, in the hope (so far unrealized) of superposing solutions of (6.4) acted on by operations in the symmetry group, with each solution multiplied by the appropriate group character. An obvious symmetry is dilation: XP is invariant under

$$X \to KX, \quad P \to P/K. \tag{6.7}$$

From (6.2), K corresponds to evolution after time $\log K$. This implies that the operator (6.3) generates dilations, in the same way that the momentum operator generates translations, and the following series of transformations makes this obvious:

$$\begin{aligned} f(KX) &= f(\exp\{\log K + \log X\}) = \exp\left\{(\log K)\frac{d}{d\log X}\right\} f(X) \\ &= \exp\left\{(\log K)\, X\frac{d}{dX}\right\} f(X) = K^{X\frac{d}{dX}} f(X) = \frac{1}{K^{1/2-iH}} f(X). \end{aligned} \tag{6.8}$$

One possibility is to choose the integer dilations $K = m$, and the characters unity. Then the superposition of solutions (6.5) does contain $\zeta(1/2-iE)$ as a factor, but there seems no reason to impose the condition that this must vanish. Moreover, the set of integer dilations does not form a group (the inverse multiplications $1/m$ are missing).

Another possibility, closely related to the ideas of [62], is to use not all integers but the group of integers under multiplication (mod k) [68]. This would have two advantages. First, it involves only integer dilations. Second, including the characters $\chi(n)$ of this group (sets of k complex numbers with unit modulus) opens the possibility of widening the interpretation as eigenvalues of XP, to include the zeros of Dirichlet

L-functions. These are defined by the series

$$L_\chi(s) \equiv \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}. \tag{6.9}$$

(The special case $\chi = 1$ corresponds to $\zeta(s)$.) It is conjectured that for all these L-functions the complex zeros lie on the line $\mathrm{Re}s = 1/2$. On this interpretation, each L-function corresponds to a different self-adjoint extension of XP under identification of positions X that are related by dilations in the group of integers under multiplication (mod k). An analogy is with the quantum mechanics of a particle in a periodic potential (e.g., an electron in a crystal): from the Bloch-Floquet theorem, solutions of the underlying differential equation are all periodic up to a phase factor $\exp(i\alpha)$; each choice of α is a different self-adjoint extension, and generates a discrete spectrum. The analogy is imperfect, because α is continuous, whereas the L-functions cannot be continuously parameterized. A closer analogy is with quantization on a torus phase space [69], where for topological reasons the permited phases are discrete.

The dynamics (6.2) suggests that the system might be closed by connecting the asymptotic positions with the asymptotic momenta. Then particles flowing out at $X = \pm\infty$ would be reinjected at $P = \pm\infty$. Related to this is a class of dilations where K is H-dependent (of course these are still symmetries of H). Specifically, the choice $K = 2\pi/(XP)$ yields the canonical transformation

$$X \to X_1 = \frac{2\pi}{P}, \quad P \to P_1 = \frac{XP^2}{2\pi}, \tag{6.10}$$

corresponding to exchange of X and P (the more familiar $X \to P$, $P \to -X$ does not leave XP invariant). A short calculation gives the transformed quantum wavefunction $\psi_1(X_1)$ in terms of the untransformed momentum wavefunction ϕ as

$$\psi_1(X_1) = \frac{(2\pi)^{1/4}}{|X_1|}\phi\left(\frac{\sqrt{2\pi}}{|X_1|}\right). \tag{6.11}$$

We do not know how to convert this “quantum exchange” into an effective boundary condition, but note its connection with the following intriguing identity, obtained from the momentum wavefunction formula (6.6) and the functional equation for $\zeta(s)$:

$$\begin{gathered} X^{1/2}\zeta\left(\tfrac{1}{2} - iE\right)\psi_E(X) - P^{1/2}\zeta\left(\tfrac{1}{2} + iE\right)\phi_E(P) = 0, \\ \text{where } PX = 2\pi \ (= h). \end{gathered} \tag{6.12}$$

If (only) the minus were a plus, this would be a condition generating the Riemann zeros.

We can sum up these scattered remarks about XP by returning to the properties listed at the beginning of this section. XP is consistent with point a, part of b (XP dynamics is unstable but not bounded), and c, d, g, h, i, and j. Concerning point e, the appearance of times that are logarithms of integers begins to be plausible in view of the association between dilation and evolution, but primes do not appear in any obvious way. We have no explanation of property f.

REFERENCES

[1] M. V. Berry, *Riemann's zeta function: A model for quantum chaos?*, in Quantum Chaos and Statistical Nuclear Physics, T. H. Seligman and H. Nishioka, eds., Lecture Notes in Phys. 263, Springer-Verlag, New York, 1986, pp. 1–17.

[2] M. V. Berry, *Quantum chaology (the Bakerian lecture)*, Proc. Roy. Soc. Lond. Ser. A, 413 (1987), pp. 183–198.

[3] M. V. Berry, *Some quantum-to-classical asymptotics*, in Chaos and Quantum Physics, Les Houches Lecture Series 52, M.-J. Giannoni, A. Voros, and J. Zinn-Justin, eds., North-Holland, Amsterdam, 1991, pp. 251–304.

[4] J. P. Keating, *The Riemann zeta-function and quantum chaology*, in Quantum Chaos, G. Casati, I. Guarneri, and V. Smilansky, eds., North-Holland, Amsterdam, 1993, pp. 145–185.

[5] J. P. Keating, *Periodic orbits, spectral statistics, and the Riemann zeros*, in Supersymmetry and Trace Formulae: Chaos and Disorder, J. P. Keating, D. E. Khmelnitskii, and I. V. Lerner, eds., Plenum, New York, 1998, pp. 1–15.

[6] M. V. Berry and J. P. Keating, *$H = xp$ and the Riemann zeros*, in Supersymmetry and Trace Formulae: Chaos and Disorder, J. P. Keating, D. E. Khmelnitskii, and I. V. Lerner, eds., Plenum, New York, 1998, pp. 355–367.

[7] H. M. Edwards, *Riemann's Zeta Function*, Academic Press, New York, London, 1974.

[8] D. Zagier, *The first 50 million prime numbers*, Math. Intelligencer, 0 (1977), pp. 7–19.

[9] J. van de Lune, H. J. J. te Riele, and D. T. Winter, *On the zeros of the Riemann zeta function in the critical strip*. IV, Math. Comp., 46 (1986), pp. 667–681.

[10] N. Levinson, *More than one third of the zeros of Riemann's zeta-function are on $\sigma = 1/2$*, Adv. Math., 13 (1974), pp. 383–436.

[11] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, Clarendon Press, Oxford, UK, 1986.

[12] V. I. Arnold, 1978, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, 1978.

[13] J. L. Synge, *The Hamiltonian method and its application to water waves*, Proc. Roy. Irish Acad. A, 63 (1963), pp. 1–34.

[14] M. V. Berry, *Regularity and chaos in classical mechanics, illustrated by three deformations of a circular billiard*, European J. Phys., 2 (1981), pp. 91–102.

[15] M. V. Berry, *Regular and irregular motion*, in Topics in Nonlinear Mechanics, S. Jorna, ed., AIP Conf. Proc. 46, 1978, pp. 16–120.

[16] H. G. Schuster, *Deterministic Chaos. An Introduction*, VCH Verlagsgesellschaft, Weinheim, Germany, 1988.

[17] A. M. Ozorio de Almeida, 1988, *Hamiltonian Systems: Chaos and Quantization*, Cambridge, University Press, Cambridge, UK, 1988.

[18] M. C. Gutzwiller, *Periodic orbits and classical quantization conditions*, J. Math. Phys., 12 (1971), pp. 343–358.

[19] M. C. Gutzwiller, *Chaos in classical and quantum mechanics*, Springer-Verlag, New York, 1990.

[20] R. Balian and C. Bloch, *Distribution of eigenfrequencies for the wave equation in a finite domain:* III. *Eigenfrequency density oscillations*, Ann. Physics, 69 (1972), pp. 76–160.

[21] N. L. Balazs and A. Voros, *Chaos on the pseudosphere*, Phys. Rep., 143 (1986), pp. 109–240.

[22] J. M. Robbins, *Maslov indices in the Gutzwiller trace formula*, Nonlinearity, 4 (1991), pp. 343–363.

[23] H. P. Baltes and E. R. Hilf, *Spectra of Finite Systems*, B-I Wissenschaftsverlag, Mannheim, Germany, 1976.

[24] Y. G. Sinai, *Introduction to Ergodic Theory*, Princeton University Press, Princeton, NJ, 1976.

[25] J. H. Hannay and A. M. Ozorio de Almeida, *Periodic orbits and a correlation function for the semiclassical density of states*, J. Phys. A., 17 (1984), pp. 3429–3440.

[26] W. Parry and M. Pollicott, *Zeta-functions and the periodic orbit structure of hyperbolic dynamics*, Asterisque, 187/188 (1990), pp. 9 et seq.

[27] P. Cvitanovic and B. Eckhardt, J. Phys. A., 24 (1991), pp. L237–L241.

[28] M. L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels*, Academic Press, New York, London, 1967.

[29] F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Math. Phys., 3 (1962), pp. 1191–1198.

[30] F. J. Dyson and M. L. Mehta, *Statistical theory of the energy levels of complex systems* IV, J. Math. Phys., 4 (1963), pp. 701–712.

[31] C. E. Porter, *Statistical Theories of Spectra Fluctuations*, Academic Press, New York, 1965.

[32] O. Bohigas and M. J. Giannoni, *Chaotic Motion and Random-Matrix Theories*, in Mathematical and Computational Methods in Nuclear Physics, J. S. Dehesa, J. M. G. Gomez, and A. Polls, eds., Lecture Notes in Phys. 209, Springer-Verlag, 1984, pp. 1–99.

[33] M. V. Berry, *Semiclassical theory of spectral rigidity*, Proc. Roy. Soc. Lond. Ser. A, New York, 400 (1985), pp. 229–251.

[34] M. V. Berry, *Semiclassical formula for the number variance of the Riemann zeros*, Nonlinearity, 1 (1988), pp. 399–407.

[35] H. L. Montgomery, *The pair correlation of zeros of the zeta function*, Proc. Sympos. Pure Math., 24 (1973), pp. 181–193.

[36] J. P. Keating, *The semiclassical sum rule and Riemann's zeta function*, in Adriatico Research Conference on Quantum Chaos, H. Cerdeira, R. Ramaswamy, M. Gutzwiller, and G. Casati, eds., World Scientific, River Edge, NJ, 1991, pp. 280–294.

[37] N. Argaman, F. M. Dittes, E. Doron, J. P. Keating, A. Kitaev, M. Sieber, and U. Smilansky, *Correlations in the actions of periodic orbits derived from quantum chaos*, Phys. Rev. Lett., 71 (1992), pp. 4326–4329.

[38] G. H. Hardy and J. E. Littlewood, *Some problems in "Partitio Numerorum"* III: *On the expression of a number as a sum of primes*, Acta Math., 44 (1923), pp. 1–70.

[39] A. M. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Math. Comp., 48 (1987), pp. 273–308.

[40] A. M. Odlyzko, *The* 10^{20}*th zero of the Riemann zeta function and* 175 *million of its neighbours*, AT&T Bell Laboratory preprint, 1990.

[41] E. B. Bogomolny and J. P. Keating, *Gutzwiller's trace formula and spectral statistics: Beyond the diagonal approximation*, Phys. Rev. Lett., 77 (1996), pp. 1472–1475.

[42] A. Fujii, *On the Berry Conjecture*, preprint, 1997.

[43] A. V. Andreev, O. Agam, B. D. Simons, and B. L. Altshuler, *Quantum chaos, irreversible classical dynamics, and random matrix theory*, Phys. Rev. Lett., 76 (1996), pp. 3947–3950.

[44] D. Goldston and H. L. Montgomery, *Pair correlation of zeros and primes in short intervals*, in Analytic Number Theory and Diophantine Problems, Proc. 1984 Stillwater Conference at Oklahoma State University, pp. 183–203.

[45] G. Gasper and M. Rahman, *Basic Hypergeometric Series*, Cambridge University Press, Cambridge, UK, 1990.

[46] D. A. Hejhal, *On the Triple Correlation of Zeros of the Zeta Function*, IMRN, 1994, pp. 293–302.

[47] Z. Rudnick and P. Sarnak, *Zeros of principal L-functions and random-matrix theory*, Duke Math. J., 81 (1996), pp. 269–322.

[48] E. B. Bogomolny and J. P. Keating, *Random matrix theory and the Riemann zeros* I: *Three- and four-point correlations*, Nonlinearity, 8 (1995), pp. 1115–1131.

[49] E. B. Bogomolny and J. P. Keating, *Random-matrix theory and the Riemann zeros* II: *n-point correlations*, Nonlinearity, 9 (1996), pp. 911–935.

[50] W. Gabcke, *Neue Herleitung und Explizite Restabschätzung der Riemann-Siegel-Formel*, Ph.D. Thesis, University of Göttingen, Germany, 1979.

[51] M. V. Berry, *The Riemann-Siegel formula for the zeta function: High orders and remainders*, Proc. Roy. Soc. Lond. Ser. A, 450, 1995, pp. 439–462.

[52] M. V. Berry and J. P. Keating, *A rule for quantizing chaos?*, J. Phys. A, 23 (1990), pp. 4839–4849.

[53] J. P. Keating, *Periodic-orbit resummation and the quantization of chaos*, Proc. Roy. Soc. Lond. Ser. A, 436 (1992), pp. 99–108.

[54] E. B. Bogomolny, *Semiclassical quantization of multidimensional systems*, Nonlinearity, 5 (1992), pp. 805–866.

[55] M. V. Berry and J. P. Keating, *A new approximation for* $\zeta(1/2+ it)$ *and quantum spectral determinants*, Proc. Roy. Soc. Lond. Ser. A, 437 (1992), pp. 151–173.

[56] R. B. Paris, *An asymptotic representation for the Riemann zeta function on the critical line*, Proc. Roy. Soc. Lond. Ser. A, 446 (1994), pp. 565–587.

[57] R. B. Paris and S. Cang, *An exponentially-improved Gram-type formula for the Riemann zeta function*, Methods Appl. Anal., 4 (1997), pp. 326–338.

[58] R. B. Paris and S. Cang, *An asymptotic representation for* $\zeta(1/2+ it)$, Methods. Appl. Anal., 4 (1997), pp. 449–470.

[59] J. P. Keating and M. Sieber, *Calculation of spectral determinants*, Proc. Roy. Soc. Lond. Ser. A, 447 (1994), pp. 413–437.

[60] N. Katz and P. Sarnak, *Zeros of Zeta Functions, Their Spacings and Their Spectral Nature*, preprint, 1997.

[61] P. Sarnak, *Quantum chaos, symmetry and zeta functions*, Curr. Dev. Math. (1997), pp. 84–115.

[62] A. Connes, *Formule de trace en géométrie non-commutative et hypothèse de Riemann*, C.R. Acad. Sci. Paris, 323 (1996), pp. 1231–1236.

[63] S. Nonnemacher and A. Voros, *Eigenstate structures around a hyperbolic point*, J. Phys. A., 30 (1997), pp. 295–315.

[64] R. K. Bhaduri, A. Khare, and J. Law, *Phase of the Riemann zeta function and the inverted harmonic oscillator*, Phys. Rev. E, 52 (1995), pp. 486–491.
[65] A. Khare, *The phase of the Riemann zeta function*, Pramana, 48 (1997), pp. 537–553.
[66] J. V. Armitage, *The Riemann hypothesis and the Hamiltonian of a quantum mechanical system*, in Number Theory and Dynamical Systems, M. M. Dodson and J. A. G. Vickers, eds., London Math. Soc. Lecture Note Ser. 134, Cambridge University Press, Cambridge, UK, 1989, pp. 153–172.
[67] S. Okubo, *Lorentz-invariant hamiltonian and Riemann hypothesis*, J. Phys. A., 31 (1998), pp. 1049–1057.
[68] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, 1976.
[69] J. P. Keating, F. Mezzadri, and J. M. Robbins, *Quantum boundary conditions for torus maps*, Nonlinearity, 1999, in press.
[70] E. Bogomolny and J. P. Keating, 1999, *Asymptotics of the Pair Correlation of Riemann Zeros*, in preparation.