The Riemann Zeta-Function and Quantum Chaology.

J. KEATING

H. H. Wills Physics Laboratory - Tyndall Avenue, Bristol BS18TL (*)

Quantum chaology is concerned with the study of how the chaotic nature of classical mechanics influences the semi-classical ($\hbar \to 0$) behaviour of quantum phenomena. One of the central aims of this study is the development of a theory, asymptotic in the limit of small $\hbar$, relating properties of the quantum energy levels of bound systems directly to classical trajectories when the classical dynamics is completely chaotic. It might be hoped that such a theory would provide, using only classical information, both a method for the calculation of individual high-energy levels and an explanation of their general distribution.

Formally, the problem of calculating individual levels was solved, to leading order as $\hbar \to 0$, by Gutzwiller's trace formula [1, 2]. This gives rise to semi-classical approximations for spectral functions, such as the density of states, in terms of sums or products of contributions from the classical periodic orbits. The analytical properties of these formulae are, however, still far from clear. In general, this is due to difficulties in performing the necessary classical calculations. First, there is the considerable practical difficulty of computing the periodic orbits of chaotic systems—they proliferate exponentially with period and are, by definition, unstable. Second, there is the problem of finding suitable ways of describing their contributions so that an analytical study of the sums can be made. This has led people to study particular model systems where the required periodic-orbit information can be characterized by simple equations [3-5] and/or the periodic-orbit formulae are exact rather than semi-classical approximations [4, 6-8]. Unfortunately, none of these models has managed to combine the desirable features of both behaving «typically» and being mathematically simple. It is for this reason that the Riemann zeta-function is of such interest in quantum chaology.

The zeta-function is certainly not a «model system», because there is no actual system yet known to which it corresponds. Nevertheless it does provide a

(*) Address after 1 October 1991: Department of Mathematics, University of Manchester, Manchester M13 9PL, U.K.

remarkably simple and useful mathematical model of the equations of quantum chaology. Specifically, there is a strong similarity between the number-theoretical relationship which exists between the nontrivial zeros of the zeta-function and the prime numbers, and the semi-classical relationship which exists between the quantum energy levels and classical periodic orbits of typical chaotic systems. This connection is particularly attractive because many of the number-theoretical formulae are exact, rather than just asymptotic approximations, and the analogues of the periodic orbits—the prime numbers—are very easy to compute and to deal with analytically.

There is also another compelling reason for studying the Riemann zeta-function (apart from its inherent mathematical interest). There now appears to be overwhelming evidence that the energy levels of typical systems have statistical distributions which, in the semi-classical limit, exhibit universality. Specifically, it has been shown numerically that in many systems the energy level statistics are closely related to the statistics of the eigenvalues of random matrices (see, for example, [9]). It is one of the goals of quantum chaology to develop a theory which can explain this remarkable fact in terms of the properties of the periodic orbits. In this direction, BERRY [10] developed a semi-classical theory of certain particular statistics which goes a good way towards showing precisely which classical quantities this universality is related to, and also when and in which way it breaks down. However, this theory still leaves some important gaps to be filled. One of the difficulties is that in general little is known about the statistical distribution of the actions of periodic orbits.

The zeta-function plays an important role in this story because it has been found that its nontrivial zeros have a statistical distribution which is remarkably close to the GUE form of random-matrix theory [11]. Moreover, this agreement eventually breaks down in precisely the manner predicted by Berry's theory [10, 12]. For the zeta-function, however, many aspects of the semi-classical theory which are not in general accessible can be worked out in detail [13, 14]. Consequently, it provides an important guide as to how progress might be made in the general case.

Of course, one of the main concerns in the study of the zeta-function is, at least for physicists, to disentangle properties which are essentially number-theoretical from those which are related to the basic form of the equations involved, and hence which might cast new light upon our semi-classical theories. It is my aim in this lecture to review some aspects of the theory of the zeta-function with this in mind. It is hoped that the style will be informal; certainly no pretence at rigour will be made, since this can be found in abundance elsewhere. I will not give detailed references for the mathematics because almost everything that could be needed may be found in the excellent books by TITCHMARSH [15] and EDWARDS [16].

Finally, I have been forced to omit two interesting topics that certainly deserve some mention. The first is the appearance of the Riemann zeta-function in an important chaotic-scattering problem and thence its use in the characterization of chaotic features in the quantum descriptions of general scattering systems [17]. Secondly, because we do not even know whether there is a real dynamical system lurking behind the zeta-function, it can give no help to the study of how dynamical properties affect the analytical behaviour of the periodic-orbit formulae. Thus it was not possible to discuss the important role played by shadowing and the curvature expansion [18].

1. – The Riemann zeta-function.

Riemann's zeta-function is a function of a complex variable s, and may be defined, for $\mathrm{Re}\, s > 1$, by the Euler product over primes p

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

Expanding out the terms in the product and then the product itself gives, using the theorem of unique prime factorization, a Dirichlet series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{2}$$

which again converges if $\mathrm{Re}\, s > 1$. It may be helpful to view the set of integers n in (2) as the set of all possible multiplicative combinations of prime powers p^α, $\alpha \geqslant 0$.

There also exists a representation of $\zeta(s)$ which converges everywhere in the complex plane, except for the simple pole at $s = 1$. Although we shall not make explicit use of this result, it will be referred to later on and so for completeness we now outline its derivation. Using the definition of $\Gamma(s)$,

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \sum_{n=1}^{\infty} \frac{1}{\Gamma(s)} \int_0^{\infty} x^{s-1} \exp[-nx]\, \mathrm{d}x . \tag{3}$$

If $\mathrm{Re}\, s > 1$, the order of summation and integration can be interchanged to give

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{x^{s-1}}{\exp[x] - 1}\, \mathrm{d}x . \tag{4}$$

Now it may be checked directly that, if $\mathrm{Re}\, s > 1$, this integral is the limit as $R \to 0$ in the contour C (shown in fig. 1) of

$$\zeta(s) = \frac{1}{\Gamma(s)(\exp[2\pi i s] - 1)} \int_C \frac{z^{s-1}}{\exp[z] - 1}\, \mathrm{d}z , \tag{5}$$

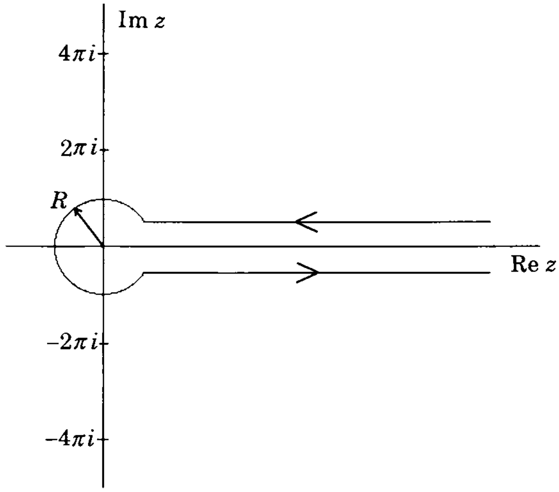

Fig. 1. – The contour of the integral (5). The circle has radius $R < 2\pi$. There is a cut along the positive real axis.

where we have taken the cut in the complex plane to lie along the positive real axis. If, however, the limit is not taken, then the integral converges for all values of s and so represents an analytic continuation of (2) to the whole complex plane.

Riemann's main interest in the zeta-function was as a tool for studying the primes, the representation (1) giving an explicit connection. He gave an explicit formula for the prime counting function $\pi(x)$—the number of primes less than or equal to x—in terms of the zeros of $\zeta(s)$ and hence showed that the distribution of the primes is related to the distribution of these zeros.

The zeta-function has an infinite set of zeros, called the trivial zeros, at the negative even integers. These are due to the poles of $\Gamma(s)$ in (5) and do not play any part in the following story. It also has infinitely many complex, or nontrivial zeros and it is these that we shall be concerned with here.

It is known that these nontrivial zeros all have real parts greater than 0 and less than 1. Thus they lie in the strip in the complex plane bordered by these lines. The Riemann hypothesis is that in fact they all lie on the line with real part 1/2; that is, that the nontrivial solutions of

$$\zeta\left(\frac{1}{2} - iE\right) = 0 \tag{6}$$

are all real.

The proof, or disproof, of the Riemann hypothesis is undoubtedly one of the most important problems in pure mathematics. The best that has been achieved so far is to prove that at least 1/3 of the infinitely many nontrivial solutions of (6) are indeed real. Computations have shown that these include the first $1.5 \cdot 10^9$ nontrivial zeros.

It is an old suggestion, going back at least to POLYA [19], that one way of proving the hypothesis would be to show that the nontrivial solutions of (6) are the eigenvalues of an Hermitian operator. Some quantum chaologists like to consider the possibility that they are, in fact, the energy levels of a real dynamical system; that is, that the operator has a Hamiltonian classical limit. Evidence that this might indeed be the case is provided by the explicit formulae connecting the zeros to the primes. Before discussing these results, however, I first need to present one of the most important equations describing the analytic structure of the zeta-function. This will play a central role in the subsequent connection to the formulae of quantum chaology.

2. – The functional equation.

The functional equation is so important in the theory of the zeta-function that TITCHMARSH [15] gives seven different methods of proof. Lying at the heart of many of these methods is the Poisson summation formula:

$$\sum_{m=-\infty}^{\infty} f(m) = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \exp[2\pi i k x]\, dx. \tag{7}$$

As discussed on p. 17 of Titchmarsh's book, simply applying (7) to $f(m) = |m|^{-s}$ gives, ignoring all questions of convergence,

$$\sum_{m=1}^{\infty} \frac{1}{m^s} = 2^s \pi^{s-1} \sin\left(\frac{1}{2} s\pi\right) \Gamma(1-s) \sum_{k=1}^{\infty} \frac{1}{k^{1-s}}, \tag{8}$$

which may be interpreted as

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{1}{2} s\pi\right) \Gamma(1-s)\, \zeta(1-s), \tag{9}$$

which is the functional equation. However, in reality, (8) has only formal meaning because the two sums share no common domain of convergence. Nevertheless, the above method does give the correct answer.

A mathematically sound way of obtaining the functional equation (9) is to take the contour of the integral (5) and expand the radius of the circle, *i.e.* let $R \to \infty$. If $\operatorname{Re} s < 0$, the integral around the expanded contour may easily be seen to tend to zero in this limit and so the only contributions come from the poles of the integrand through which the contour passes. This then leads, using the theorem of residues, directly to (9).

It is worth noting that both the above methods of deriving the functional equation are inherently number-theoretical; the first using the Poisson summation formula for integer sums, and the second using the integral representation (5), which we were able to obtain because the sum over the integers on the right-hand side of (3) can be performed explicitly.

In the light of the functional equation (9) it is convenient to introduce a new function $\Delta_R(E)$, which is defined by

$$\Delta_R(E) = -\frac{1}{2}\left(\frac{1}{4}+E^2\right)\pi^{-1/4+iE/2}\Gamma\left(\frac{1}{4}-i\frac{E}{2}\right)\zeta\left(\frac{1}{2}-iE\right) \tag{10}$$

and in terms of which (9) becomes

$$\Delta_R(E) = \Delta_R(-E). \tag{11}$$

Then, using the definition (10), the functional equation means that $\Delta_R(E)$ is real when E is real; *i.e.* that, when $\operatorname{Im} E = 0$,

$$\Delta_R(E) = \Delta_R^*(E). \tag{12}$$

Involved in the definition (10) is the implicit transformation of the complex variable s to $E = i(s-1/2)$. Hence the line $\operatorname{Im} E = 0$ is just the *critical line*, where the Riemann hypothesis states that the nontrivial zeros of the zeta-function lie. In fact, it may be seen from (10) that the zeros of Δ_R correspond precisely to these nontrivial zeros, the trivial ones being cancelled by the poles of the Γ-function. Hence the Riemann hypothesis is equivalent to the statement that all of the zeros E_n of $\Delta_R(E)$ are real.

Importantly, $\Delta_R(E)$ may be expressed as a product over its zeros (see, for example, [16])

$$\Delta_R(E) = \Delta_R(0)\prod_n\left(1-\frac{E}{E_n}\right). \tag{13}$$

The product does not converge absolutely, however it follows from (11) that the zeros come in pairs, $\pm E_n$, and, if these are taken together in (13), then it converges conditionally (this will be shown in the next section). It can be made to converge absolutely using the standard Hadamard factorization methods (see, for example, [15]). For future convenience, we choose to write (13) in the following way:

$$\Delta_R(E) = \text{const}\cdot\prod_n (E-E_n)A(E,E_n), \tag{14}$$

where $A(E, E_n)$ is a real regularizing function chosen to make the product converge (see, for example, [20]). In the present case, taking $A(E, E_n) = -1/E_n$ will suffice for conditional convergence, as discussed above.

The functional equation implies that, if the Riemann hypothesis is not true, then all of the complex zeros must come in pairs with equal and opposite imaginary parts. Consider now the case if it were true. Then all of the zeros E_n in (13) would be real and so $\Delta_R(E)$ would clearly be real when E is real. Thus, if the Riemann hypothesis were true, this would automatically give the functional equation (12), which at present can only be obtained using special number-theoretical methods. This fact will be of considerable importance in making the connection with quantum chaology.

Finally, we note that, if $\operatorname{Im} E > 1/2$, then the representations (1) and (2) of the zeta-function can be substituted into the definition (10). For future convenience, we write the resulting equations in the following way: if $\operatorname{Im} E > 1/2$

$$\Delta_R(E) = B_R(E) \exp[-i\pi \overline{N}_R(E)] \prod_p \left(1 - \exp[-\frac{1}{2}\log p + iE \log p]\right)^{-1}, \tag{15}$$

$$\Delta_R(E) = B_R(E) \exp[-i\pi \overline{N}_R(E)] \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \exp[iE \log n], \tag{16}$$

where $B_R(E)$ is the modulus of the term multiplying the zeta-function in (10) (and is real and nonzero when E is real), and $-\overline{N}_R(E)$ is its argument divided by π. It will be shown in the next section that $\overline{N}_R(E)$ is, in fact, the mean number of zeros with real parts between 0 and E. It is worthwhile pointing out that (15) and (16), the two simplest representations of Δ_R, do not obviously satisfy the functional equation (12). Of course, neither of them converges when $\operatorname{Im} E = 0$.

3. – The staircase of zeros.

We are now in a position to go straight to the equations which provide the link between the theory of the Riemann zeta-function and the semi-classical formulae of quantum chaology. The results to be obtained in this section will form the basis for the remainder of this lecture.

We begin by considering the staircase function $N_R(E)$, defined to be the number of zeros of $\Delta_R(E)$ with real parts between 0 and E. It follows from the results stated in sect. **1** that these zeros must have imaginary parts between $-1/2$ and $1/2$, and so, using the fact that they come in pairs $\pm E_n$, we have that

$$N_R(E) = \frac{1}{2} \times \frac{1}{2\pi i} \int_C \frac{\Delta_R'}{\Delta_R}\, dz, \tag{17}$$

where the integral is taken anticlockwise around the rectangular contour with vertices $\pm E \pm (1/2 + \alpha) i$, $\alpha > 0$. Then, using the functional equation in the forms (11) and (12), this can be reduced to

$$N_R(E) = \frac{1}{\pi i} \int_{C'} \frac{\Delta_R'}{\Delta_R}\, dz, \tag{18}$$

where C' is the section of the rectangle connecting E and $(1/2 + \alpha) i$. The integral can obviously be evaluated as i times the change in the argument of $\Delta_R(z)$ along C'. Thus finally we have, using the definition (10), that

$$N_R(E) = 1 - \frac{1}{2\pi} E \log \pi - \frac{1}{\pi} \operatorname{Im} \log \Gamma\left(\frac{1}{4} - i\frac{E}{2}\right) - \frac{1}{\pi} \operatorname{Im} \log \zeta\left(\frac{1}{2} - iE\right), \tag{19}$$

where the fact that all of the factors in (10) are real at $(1/2 + \alpha)i$ has been used. In the notation of the last section this is

$$N_R(E) = \overline{N}_R(E) - \frac{1}{\pi} \operatorname{Im} \log \zeta\left(\frac{1}{2} - iE\right). \tag{20}$$

Thus far, everything is legitimate and can be found in the standard books on the subject (*e.g.*, [15, 16]). But now we depart from rigorous mathematics by substituting for $\zeta(1/2 - iE)$ in (20) the Euler product (1). Strictly speaking, this is, of course, not allowed since the product only converges if $\operatorname{Im} E > 1/2$ and we are interested at present in values of E on the real axis (see the above definition of the staircase function). Nevertheless, making this substitution we obtain

$$N_R(E) = \overline{N}_R(E) - \frac{1}{\pi} \sum_p \sum_{k=1}^{\infty} \frac{1}{k} \exp\left[-\frac{1}{2} k \log p\right] \sin(Ek \log p), \tag{21}$$

where the sum over the integers k comes from expanding the logarithms of the individual terms in the Euler product. It may be seen from (21) that $\overline{N}_R(E)$ is the mean of the staircase function.

Now differentiating with respect to E, we obtain a formula for the density of states [10]:

$$d_R(E) = \sum_n \delta(E - E_n), \tag{22}$$

$$d_R(E) = \overline{d}_R(E) - 2 \sum_p \sum_{k=1}^{\infty} \frac{\log p}{2\pi} \exp\left[-\frac{1}{2} k \log p\right] \cos(Ek \log p), \tag{23}$$

where $\overline{d}_R = \mathrm{d}\overline{N}_R/\mathrm{d}E$ is the mean density of states. This formula represents the main result of this section. It provides the promised explicit relationship between the zeros of the zeta-function and the prime numbers. In fact, the relationship is symmetrical. Fourier-transforming both sides with respect to E gives a formula for a weighted density of states for the logarithms of the primes in terms of a trigonometric sum involving the zeros.

Before going on to discuss the properties of (23) and the reason for its importance, I can now justify the statement made in the last section concerning the conditional convergence of the product (13). If the terms with $\pm E_n$ are taken together, we have

$$\Delta_R(E) = \Delta_R(0) \prod_{E_n>0} \left(1 - \frac{E^2}{E_n^2}\right), \tag{24}$$

and so we only need consider the convergence of

$$\sum_{E_n>0} \frac{1}{E_n^2}. \tag{25}$$

Using Stirling's asymptotic formula for the Γ-function in (19), we have

$$\overline{N}_R(E) = \frac{E}{2\pi}\left(\log\frac{E}{2\pi} - 1\right) + \frac{7}{8} + O\left(\frac{1}{E}\right), \tag{26}$$

and so the mean density of states is given approximately by

$$\overline{d}_R(E) \approx \frac{1}{2\pi}\log\frac{E}{2\pi}. \tag{27}$$

Thus the late terms in the sum (25) may be approximated by

$$\int \frac{\mathrm{d}E}{E^2}\,\frac{1}{2\pi}\log\frac{E}{2\pi}, \tag{28}$$

which converges as the upper limit of the integral tends to ∞. Hence, as claimed, the product (24) converges.

4. – The quantum chaology connection.

For a bound chaotic Hamiltonian system with energy levels E_n, Gutzwiller's trace formula gives for the density of states, defined as in (22), that

$$d(E) \approx \overline{d}(E) + \frac{2}{\hbar}\sum_p \sum_{k=1}^{\infty} A_{p,k}(E)\cos\left\{\frac{k}{\hbar}S_p(E) - k\frac{\pi}{2}\alpha_p\right\} \tag{29}$$

to leading semi-classical order [1]. Here $\overline{d}(E)$ is the mean density of states represented, for example, by Weyl's asymptotic series [21], the sum labelled p runs over all primitive periodic orbits of the classical motion, and the integer k denotes the k-th repetition of such an orbit. The phase of each periodic-orbit contribution contains the action around the orbit

$$S_p(E) = \oint \boldsymbol{p}\cdot\mathrm{d}\boldsymbol{q} \tag{30}$$

and a characteristic integer α_p, the Maslov index of its invariant manifolds [22, 23]. The amplitude contribution is given by

$$A_{p,k}(E) = \frac{T_p(E)}{2\pi\sqrt{|\det(\boldsymbol{M}_p^k(E) - \boldsymbol{I})|}}, \tag{31}$$

where T_p is the period of the orbit ($T_p = \mathrm{d}S_p/\mathrm{d}E$) and $\boldsymbol{M}_p$ is the Poincaré map linearized around it. This amplitude thus depends upon the stability of the orbit, and indeed for long orbits

$$A_{p,k}(E) \approx \frac{T_p(E)}{2\pi} \exp\left[-\frac{1}{2}\lambda_p(E)\, kT_p(E)\right], \tag{32}$$

where λ_p is the sum of the Liapunov exponents.

Now the point is that the formula for the density of states of the zeta-function zeros (23) bears a remarkable similarity to the semi-classical formula (29) [10]. Specifically, it is just like a periodic-orbit formula, scaled so that $\hbar = 1$, and with orbits which have indices

$$\alpha_p = 0, \tag{33}$$

actions

$$S_p(E) = E \log p \tag{34}$$

and hence periods

$$T_p = \log p. \tag{35}$$

The amplitude of each oscillatory contribution is then given by

$$A_{p,k} = \frac{T_p}{2\pi} \exp\left[-\frac{1}{2} kT_p\right]. \tag{36}$$

Finally, from (34), the semi-classical limit in this case corresponds to $E \to \infty$.

Thus, if there is a Hamiltonian system whose quantum energy levels are the zeros of $\Delta_R(E)$, then the above results suggest that it has primitive periodic orbits with periods equal to the logarithms of the primes. Moreover, comparing (36) and (32), each orbit has one stability exponent

$$\lambda_p = 1. \tag{37}$$

Of course, the existence of such a system would immediately prove the Riemann hypothesis and it is certainly a problem of immense interest to see whether the above evidence gives any clues as to what it might be. But this is not a problem that I wish to pursue here (even if I could!). Rather, I would like to view the above relationship as a tool for understanding the mathematical structure of the trace formulae. In fact it is worth pointing out at this stage that there are two features of (23) which cannot yet be related to what we know about the periodic formula (31). The first is the fact that the sum of the oscillatory terms comes with an overall minus sign, and the second is that, since (36) is exact, it is as if the «orbits» have only a single, unstable transverse dimension, *i.e.* as if the Poincaré section were one-dimensional. These difficulties will not, however, bear upon the use of the analogy between (23) and (29) as a guide to the understanding of the analytic properties of semi-classical formulae.

If (23) is to be used as a mathematical model to study the behaviour of the trace formula, the «orbit» properties (33)-(37) must be consistent with what is known about the periodic orbits of genuine chaotic systems. The first check is the proliferation of their number with period. For chaotic systems the number $n(T)$ of periodic orbits with periods less than or equal to T increases as $T \to \infty$ like

$$n(T) \sim \frac{\exp[h_t T]}{h_t T}, \tag{38}$$

where h_t is the topological entropy. For the Riemann case we need the number $n_R(T)$ of prime powers p^k such that $\log p^k \leq T$. From the prime-number theorem [15] this is given by

$$n_R(T) \sim \frac{\exp[T]}{T} \tag{39}$$

as $T \to \infty$. This is exactly the same as for the periodic orbits of a chaotic system with $h_t = 1$, from (38). Now, since all of the stability exponents are equal to 1, (23) is like the periodic-orbit formula for a uniformly hyperbolic system. But for such systems it is known that this exponent should, in fact, be equal to the topological entropy, and this is obviously also true for the Riemann case. It then follows, since these quantities are equal, that the classical sum rule of Hannay and Ozorio de Almeida [24],

$$\sum_p \sum_{k=1}^{\infty} A_{p,k}^2 \delta(T - kT_p) \sim \frac{T}{4\pi^2} \tag{40}$$

as $T \to \infty$, is also satisfied by the primes.

It is thus apparent that the number-theoretical relationship (23) really is remarkably similar to the general semi-classical formula (29). It has, moreover, the attractive feature of considerable mathematical simplicity. First, it is exact rather than just an asymptotic approximation. It shares this feature with certain special systems such as the geodesic motion on surfaces of constant negative curvature [6, 7] and the cat maps [4]. Second, the «periodic orbits»—the primes—are exceptionally simple to find, and there are explicit formulae giving their actions, periods and stabilities. Consequently, (23) is an almost ideal mathematical model equation to work with. Before proceeding to an analysis of its structure, however, I first wish to show that many of the results concerning the function Δ_R also have direct semi-classical analogues consistent with the associations made above.

The function Δ_R is, by definition, zero when $E = E_n$. For a quantum system with energy levels E_n, it is natural to define the analogous function by

$$\Delta(E) \equiv \det\{(E - \hat{H})A(E, \hat{H})\} = \prod_n (E - E_n)A(E, E_n), \tag{41}$$

where A is a regularizing function which makes the product converge (see, for example, [20]), and which can be chosen to be real on the real energy axis and everywhere nonzero. Δ is called the *spectral determinant*. Its representation (41) is the obvious analogue of (14).

It was emphasized in sect. **1** that one of the most important properties of Δ_R is the functional equation (12). This was derived using number-theoretical methods, but it was noted that the result would, in fact, follow directly from the Riemann hypothesis, if true. The energy levels E_n in the product (41) are, being the eigenvalues of an Hermitian operator, clearly real. Hence the spectral determinant automatically satisfies a functional equation the same as (12) [25, 26]: if $\operatorname{Im} E = 0$

$$\Delta(E) = \Delta^*(E). \tag{42}$$

Obviously the equality in (42) holds exactly. The equation has nontrivial content, however, only if there is a representation of Δ for which it is not manifestly satisfied. In the semi-classical limit there is such a representation (see, for example, [25]): to leading order semi-classically,

$$\Delta(E) \approx B(E) \exp[-i\pi\overline{N}(E)]\, Z(E), \tag{43}$$

where $B(E)$ is a function which will not be specified (for details see [20]) but which is real and nonzero when E is real, and $\overline{N}(E)$ is the mean of the spectral staircase (*i.e.* the mean number of levels less than E), given, for example, by Weyl's asymptotic series. $Z(E)$ is a dynamical zeta-function, defined for a 2-dimensional system as a product over the primitive periodic orbits p:

$$Z(E) = \prod_p \prod_{k=0}^{\infty} (1 - \exp[-(k + 1/2)\lambda_p T_p + iS_p/\hbar - i\pi\alpha_p/2]) \tag{44}$$

in the region where the product converges (see next section). Such zeta-functions were introduced by SELBERG [27], and they form the starting point for the method of cycle expansions [18]. In higher dimensions their definition may be found in, for example, [26]. The representation (43) is the obvious analogue of (15). It may be derived directly from Gutzwiller's trace formula, and, indeed, just as the product formula for Δ_R was used in the last section to obtain the density of states for the Riemann zeros d_R, so (43) can be used in exactly the same way to derive the density-of-states formula (29).

Substituting (43) into (42) gives the *semi-classical functional equation*: if $\operatorname{Im} E = 0$

$$\exp[-i\pi\overline{N}(E)]\,Z(E) \approx \exp[i\pi\overline{N}(E)]\,Z^*(E) \tag{45}$$

as $\hbar \to 0$ [25, 26, 28]. There is an approximate equality between the two sides because (43) is only an approximation to leading order in $\hbar$ and so, while it must be real to this order, it could have an imaginary part of some higher order. If the whole asymptotic expansion of Δ were known to all orders in $\hbar$ (and beyond), then the result must, by virtue of (42), be real on the real energy axis. Thus, in general, truncating the series at any given level, the resulting approximation to Δ may have an imaginary part of some higher order which must eventually be cancelled off if more orders are included in the expansion. Hence it is consistent to *impose* the condition that each level of the semi-classical expansion individually be real; that is, we may impose an exact equality in (45) [28]. It appears that the resulting equation is then useful in finding semi-classical approximations to energy levels [29, 30]. Its use in the study of the analytical properties of periodic-orbit formulae will be discussed in a later section.

Finally, we give a semi-classical analogue of (16), the Dirichlet series representation of Δ_R. It is obtained by expanding out the products in (44), just as (2) may be obtained by expanding out the Euler product (1). Indeed, it was noted in sect. 1 that the integers in (2) arise as the set of all combinations of prime powers. So the phases of the terms in the sum (16) run through the set of all linear combinations of the phases of the terms in the product (15), since

$$\log n = \sum_p m_p(n) \log p\,, \tag{46}$$

where the $m_p(n) \geqslant 0$ are integers and taking all possible combinations of these coefficients gives the set of all integers.

A similar result follows from the expansion of (44) [25]. The product over k can be performed using Euler's identity

$$\prod_{k=0}^{\infty}(1 - ax^k) = \sum_{m=0}^{\infty} \frac{a^m(-1)^m x^{m(m-3)/4}}{\prod_{j=1}^{m}(x^{-j/2} - x^{j/2})}\,, \tag{47}$$

and then the product over the primitive periodic orbits gives, as $\hbar \to 0$,

$$\Delta(E) \approx B(E)\exp[-i\pi\overline{N}(E)]\sum_n C_n(E)\exp\left[\frac{i}{\hbar}\mathcal{S}_n(E)\right], \tag{48}$$

where the set of values of $\mathcal{S}_n$ is given by

$$\mathcal{S}_n(E) = \sum_p m_p(n)\left(S_p(E) - \frac{\pi}{2}\alpha_p\right) \tag{49}$$

with each coefficient $m_p(n)$ running through all integer values between 0 and

∞. The amplitude C_n is given in terms of these coefficients by

(50) $$C_n = \prod_p \left[(-1)^{m_p(n)} \exp[-m_p(n)(m_p(n)-1)\lambda_p T_p/4] \left(\left| \prod_{j=1}^{m_p(n)} \det(\boldsymbol{M}_p^j - \boldsymbol{I}) \right| \right)^{-1/2} \right].$$

The linear combinations of periodic orbits (49) are known as *pseudo-orbits* [25, 31]. In (48) they may be taken to be ordered by increasing period

(51) $$\mathcal{T}_n = \frac{\mathrm{d}\mathcal{S}_n}{\mathrm{d}E}.$$

5. – Convergence properties of periodic-orbit formulae.

One of the central questions in quantum chaology is: do the periodic-orbit formulae converge, and, if not, how do they behave and how can the required information be extracted from them? It is this question which we now address for the appropriate formulae from the theory of the Riemann zeta-function.

The representations of $\Delta_R(E)$ involving the Euler product (15) and the Dirichlet series (16) converge only if $\operatorname{Im} E > 1/2$. In the second case, this can be seen most immediately by replacing the sum with an integral (*i.e.* the first term of an Euler-Maclaurin expansion), and, in the first case, by considering the logarithm of the product and in the same way replacing the resulting sum over primes by an integral with the appropriate weighting obtained from the prime-number theorem (39). The formula (23) for the density of the zeros, which was obtained from the Euler product representation of Δ_R, inherits the same domain of convergence. This may be seen directly from (23). It follows from its derivation that, in the complex energy plane at, say, $E + i\varepsilon$, the sum over primes becomes

(52) $$\frac{1}{\pi} \operatorname{Re} \left[\sum_p \sum_{k=1}^{\infty} \log p \exp\left[-\frac{1}{2} k \log p + iEk \log p - \varepsilon k \log p \right] \right].$$

Using the prime-number theorem in the form (39) then gives as an approximation to this sum

(53) $$\frac{1}{\pi} \operatorname{Re} \left[\int \mathrm{d}T \exp[T] \exp[-T/2 + iET - \varepsilon T] \right]$$

(which should certainly be a good estimate of the contribution from the large primes) and this obviously only converges if $\varepsilon > 1/2$.

A similar analysis can be applied to the general semi-classical formula (29).

Consider the absolute convergence of the sum (*i.e.* the convergence of the sum of the moduli of the terms). Very roughly, using (38) this behaves in the complex energy plane at $E + i\varepsilon$ like

$$\frac{1}{\pi\hbar}\mathrm{Re}\left[\int \mathrm{d}T \exp[h_t T]\langle \exp[-\lambda_j T/2]\rangle_{T_j \approx T} \exp[-\varepsilon T/\hbar]\right], \tag{54}$$

where $\langle\rangle$ denotes a local average of the orbits j with periods close to T. If the distribution of Liapunov exponents is sharp enough, then this average might be approximated by $\exp[-\bar{\lambda}T/2]$, where $\bar{\lambda}$ is the mean Liapunov exponent, and thus there would be absolute convergence only if $\varepsilon > \hbar(h_t - \bar{\lambda}/2)$ (for a more careful discussion see, for example, [32]).

Now consider the domain of conditional convergence of the sum. In systems where all of the orbits have $\alpha_p \equiv 0 \bmod 4$ then this will be the same as the domain of absolute convergence estimated above (the orbit actions vary too slowly to have any influence—see, for example, the Riemann case above). In systems where this is not the case, however, the alternating signs of the terms in the sum might extend the domain of conditional convergence beyond that of absolute convergence. Indeed, if it were the case that the alternating signs came at random (and in the hyperbola billiard there is some evidence for this [33]), then the sum would, in fact, grow in magnitude like the square root of the sum of the squares of its terms. Then from the classical sum rule (40), one might expect conditional convergence for all $\varepsilon > 0$. For systems not satisfying this sum rule (*e.g.*, systems which are not classically bound, such as the hyperbola billiard [33]), the above crude averaging estimate implies conditional convergence if $\varepsilon > (h_t - \bar{\lambda})\hbar/2$. For the spectral determinant the same results would be expected to hold for the product representation (44) and the pseudo-orbit sum (48). In fact, since in the sum each term has a sign which depends upon the number of orbits contributing to the particular pseudo-orbit, this is another mechanism for introducing effectively random sign changes. Hence the restriction on systems having $\alpha_p \equiv 0 \bmod 4$ might not apply. In the case of the hyperbola billiard, there is some numerical evidence that the pseudo-orbit sum does indeed converge on the real energy axis [30]; this is a special system which has $h_t < \bar{\lambda}$.

Now we return to the density-of-states formula for the Riemann zeros. This was shown above not to converge on the real E axis. The closest that this line can be approached while still using the formula where it does converge is just beyond $E + i/2$. Complex energies correspond to replacing the δ-functions in (22) with Lorentzians, and in this case they will have width $1/2$. However, the mean density of the zeros increases with E (27) and so eventually their mean separation will be less than $1/2$ and then the individual zeros will be unresolvable in this way. Roughly, this will occur when $E \sim 2\pi \exp[4\pi]$, *i.e.* from (26), at about 3316672nd zero [34].

So, what actually happens if the formula (23) is used on the real E line? The results are interesting—it works very well! As more primes are included in the sum, definite peaks form at the positions of the zeros [10]. For the staircase function N_R, given by (21), this is shown in fig. 2. In fig. 3, the step at the first zero is shown. What then is the effect of the divergence of the sum? It may be seen that as well as the actual step there are oscillations. As more primes are included in the sum the steps begin to form, but then these oscillations begin to grow (fig. 4); that is, the result does not converge to the true staircase. Probably they will keep on growing until eventually the form of the staircase is lost.

An interesting way to use the staircase as a quantization condition results from the following scheme [35]. Let us try, instead of finding peaks and steps, to find the zeros of Δ_R. We could use the Dirichlet series representation, and this approach will be considered in the next section, but here we consider the Euler product representation (15). By definition, the zeros of Δ_R are given by

$$\Delta_R(E) \equiv B_R(E) \exp[-i\pi\overline{N}_R(E)]\,\zeta\left(\frac{1}{2} - iE\right) = 0 . \tag{55}$$

From the functional equation (12) this must be real when E is real, and so for zeros on this line we have

$$\mathrm{Re}\left[\exp[-i\pi\overline{N}_R(E)]\,\zeta\left(\frac{1}{2} - iE\right)\right] = 0 \tag{56}$$

since B_R is already real, and nonzero there. Now we could look for the zeros of the modulus of $\zeta(1/2 - iE)$ [31], but it turns out to be much more efficient to con-

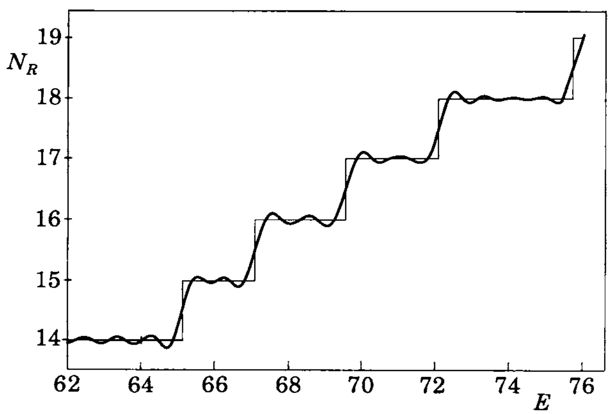

Fig. 2. – The staircase function N_R obtained from (21) using the first 200 primes and all repetitions. Also shown is the exact staircase.

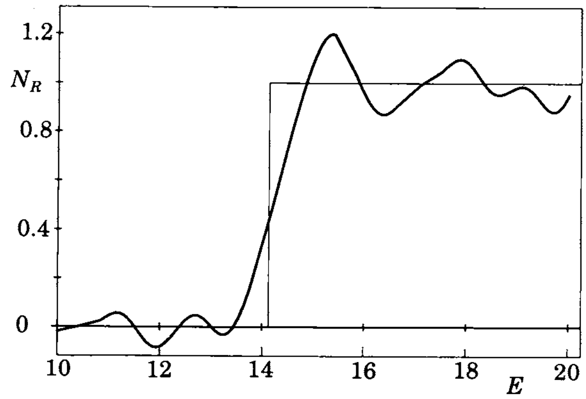

Fig. 3. – As for fig. 2, but using only the first 5 primes.

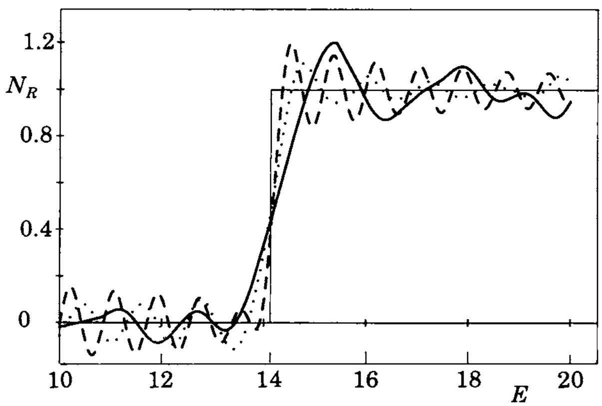

Fig. 4. – As in fig. 2, using the formula (21) with *a*) the first 5 primes (black line), *b*) the first 50 primes (dotted line) and *c*) the first 250 primes (dashed line). All repetitions are included.

sider the phase, *i.e.*

$$\cos\left(\pi\overline{N}_R(E) - \operatorname{Im}\log\zeta\left(\frac{1}{2} - iE\right)\right) = 0 . \tag{57}$$

So, using the Euler product in (56) just corresponds to finding where $N_R(E)$, given by the sum over primes (21), is equal to $n + 1/2$. Of course, this is a natural condition since at a level the true staircase jumps from n to $n + 1$. The divergence of the Euler product on the critical line now plays an important role. It causes the divergence of (21), and hence the growth in the oscillations discussed above. Eventually, as more primes are included, these are expected to grow so big that they will exceed the height, $n + 1/2$, of the middle of the steps and so will give rise to extra, spurious solutions of (57).

The method of using the product to locate zeros has, in fact, already been applied to finding semi-classical approximations to the energy levels of Artin's billiard [36] and of the anisotropic Kepler problem [37]. The results are exceedingly good. One advantage is that it appears that no levels are missed in this way. The reason for this is clear when the method is viewed in terms of the staircase: all approximations to $N(E)$ are smooth and, on average, increasing. Hence there are solutions to $N_R(E) = n + 1/2$ for all n. The problem is that, if the periodic formula diverges, then there may be too many.

Having seen that the formulae for d_R and N_R diverge, and seen the consequences of this, there remains the question: is there any cure? In fact, there are at least two. The first that I will consider was introduced by DELSARTE [38]. It was noted above that using (23) in the complex plane corresponds to having Lorentzians rather than δ-functions in the density of states. Delsarte's idea was to use Gaussians, *i.e.*

$$d_R^{(g)}(E) = \frac{1}{\varepsilon} \int_{-\infty}^{\infty} \mathrm{d}E' \exp[-\pi(E - E')^2/\varepsilon^2]\, d_R(E'), \tag{58}$$

or

$$d_R^{(g)}(E) = \frac{1}{\varepsilon} \sum_n \exp[-\pi(E - E_n)^2/\varepsilon^2]. \tag{59}$$

Substituting (23) into (58) gives

$$d_R^{(g)}(E) = \frac{1}{\varepsilon} \int_{-\infty}^{\infty} \mathrm{d}E' \exp[-\pi(E - E')^2/\varepsilon^2]\, \bar{d}_R(E') - \\ - \frac{1}{\pi} \sum_p \sum_{k=1}^{\infty} \frac{\log p}{\sqrt{p^k}} \cos(Ek \log p) \exp\left[-\frac{\varepsilon^2 k^2 \log^2 p}{4\pi}\right]. \tag{60}$$

The sum now converges for all $\varepsilon > 0$, *i.e.* for arbitrarily narrow peaks in the density of states, and so there is no limit to the number of zeros which can be identified. Of course, the closer the zeros, and thus on average the higher they are, the more primes are needed in the sum. This method of Gaussian smoothing also works for the general semi-classical formula (29), where it was introduced in [8].

The second method for making the staircase and density-of-states sums for the Riemann zeros converge is due to GUINAND [39]. A convenient form of this result is given in [34]:

$$d_R(E) = \bar{d}_R(E) - \frac{1}{\pi} \lim_{T^* \to \infty} \left\{ \sum_{p,k}^{k \log p < T^*} \frac{\log p}{\sqrt{p^k}} \cos(Ek \log p) - \right. \\ \left. - \frac{2 \exp[T^*/2]}{1 + 4E^2} [2E \sin(ET^*) + \cos(ET^*)] \right\}. \tag{61}$$

The sum is now convergent as $T^* \to \infty$, essentially because the leading divergence is subtracted off by the second term in the brackets. This result can be derived in the following manner (for a rigorous proof see Guinand's original paper).

Consider the density-of-states formula in the complex E plane:

$$d_R(E) = \bar{d}_R(E) - \frac{\mathrm{Re}}{\pi} \sum_{p,k}^{k \log p < T^*} \frac{\log p}{\sqrt{p^k}} \exp[iEk \log p] - \\ - \frac{\mathrm{Re}}{\pi} \sum_{p,k}^{k \log p \geq T^*} \frac{\log p}{\sqrt{p^k}} \exp[iEk \log p], \tag{62}$$

where the sum has been split into two parts. If $\mathrm{Im}\, E > 1/2$, then, as shown above, the infinite sum converges, and as $T^* \to \infty$ it may be approximated by an integral with the appropriate weighting obtained from (39):

$$-\frac{\mathrm{Re}}{\pi} \int_{T^*}^{\infty} \mathrm{d}T \exp[T] \exp[-T/2] \exp[iET]. \tag{63}$$

This then gives as an approximation to the second sum in (62)

$$\frac{2\,\mathrm{Re}}{\pi} \left\{ \frac{\exp[T^*/2 + iET^*]}{1 + 2iE} \right\} \tag{64}$$

if $\mathrm{Im}\, E > 1/2$. However, this quantity is well defined everywhere in the complex plane and so it may, by analytic continuation, be used on the real E axis. The result is (61).

This simple procedure thus leads to a convergent form for the divergent sum

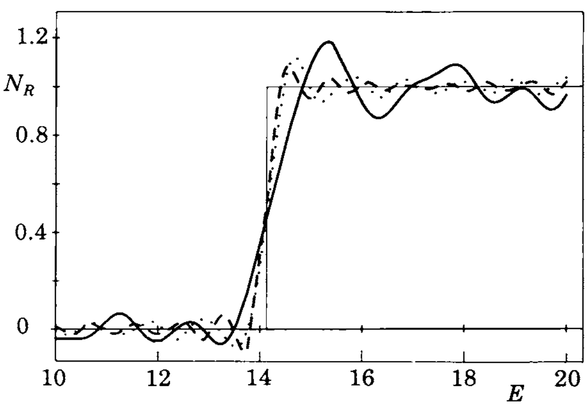

Fig. 5. – The staircase function N_R plotted using Guinand's formula—the integral of (61)—with *a)* T^* corresponding to the first 5 primes (black line), *b)* T^* corresponding to the first 50 primes (dotted line), and *c)* T^* corresponding to the first 250 primes (dashed line).

(23). The same method also works for the staircase function as given by (21). When it is used in this case, the oscillations, whose growth is the signature of the divergence of the bare sum, actually decrease in size as T^* increases, *i.e.* as more primes are included in (61). This behaviour is shown in fig. 5, which is to be compared with fig. 4. A generalization of Guinand's result to general chaotic systems was conjectured in [34]. It has recently been tested by AURICH and STEINER [40], who found it to work remarkably well.

6. – Riemann-Siegel resummation.

In the last section, I described ways in which the density-of-states and the staircase functions for the Riemann zeros could be made to converge. In this section I will consider the same problem for the Dirichlet series representation of Δ_R.

The method used to derive Guinand's result (61) can also be applied in this case. That is, the sum in (16) can be split at some point, and then the second, infinite sum can be approximated by an integral. If $\operatorname{Im} E > 1/2$, the integral converges and can be evaluated trivially. The resulting approximation may then, by analytic continuation, be used on the real E line. This method can be made rigorous by using the Euler-Maclaurin summation formula for the second infinite sum rather than just approximating it by an integral. In fact, the resulting formula was, for many years, the basis for all actual computations of the zeta-function on the critical line [16]. I will not, however, go into the details of this method here, although it certainly has a semi-classical analogue. Instead, I want to concentrate on a much more powerful and sophisticated approach which is intimately related to the functional equation described in sect. 2.

The result that I will concentrate upon is known as the *approximate functional equation* [15]. For Δ_R, it is that

$$\Delta_R(E) = B_R(E)\left\{\exp[-i\pi\overline{N}_R(E)]\sum_{n=1}^{X}\frac{1}{\sqrt{n}}\exp[iE\log n] + \right.$$
$$\left. + \exp[i\pi\overline{N}_R(E)]\sum_{n=1}^{E/2\pi X}\frac{1}{\sqrt{n}}\exp[-iE\log n] + O(X^{-1/2}) + O(\sqrt{2\pi X/E})\right\} \tag{65}$$

when E is real. In fact, the general result for complex E is almost identical to the above, but has different error estimates [15]. Taking $X = \sqrt{E/2\pi}$ in (65), the second sum is simply the complex conjugate of the first, *i.e.*, for real E,

$$\Delta_R(E) = B_R(E)\left\{2\sum_{n=1}^{\sqrt{E/2\pi}}\frac{1}{\sqrt{n}}\cos(\pi\overline{N}_R(E) - E\log n) + O(E^{-1/4})\right\}. \tag{66}$$

Two points to note about this approximation are, first, that it involves only a finite sum and thus suffers from no divergence problems, and, second, that it is real and so manifestly satisfies the functional equation (12). This second fact is hardly surprising since the methods used to derive (65) are essentially the same as those used to derive the functional equation [15]. For completeness, I now sketch the necessary extensions to the two derivations given in sect. 2.

The first method effectively involved applying the Poisson summation formula to the Dirichlet series. In order to derive (65), the Dirichlet series is split at $n = X$, and the Poisson summation formula is applied only to the tail:

$$\Delta_R(E) = B_R(E)\left\{\exp[-i\pi\overline{N}_R(E)] \sum_{n=1}^{X} \frac{1}{\sqrt{n}} \exp[iE \log n] + \right. \\ \left. + \exp[-i\pi\overline{N}_R(E)] \sum_{k=-\infty}^{\infty} \int_{X+\delta}^{\infty} dx \frac{\exp[iE \log x]}{\sqrt{x}} \exp[2\pi ikx]\right\}, \tag{67}$$

where $0 < \delta < 1$. The approximation then comes from evaluating the integrals by the method of stationary phase. This approach is treated rigorously in [15]; it is also discussed in some detail in [31].

In the second method, the contour of (5) was expanded to infinity, leaving only the contributions of the poles through which it passed. To derive (65) the denominator of the integrand is partially expanded to give

$$\Delta_R(E) = -\frac{B_R(E)\exp[-i\pi\overline{N}_R(E)]}{(1+\exp[-2i\pi E])\Gamma\left(\frac{1}{2} - iE\right)} \int_C dz\, z^{-1/2-iE} \cdot \left\{\sum_{n=1}^{X} \exp[-nz] + \frac{\exp[-Xz]}{\exp[z]-1}\right\}. \tag{68}$$

The integrals for the terms in the first sum can be evaluated directly to give the first sum in (65). This then leaves the integral of the second contribution. In deriving the functional equation we had $X = 0$ and so the contour could be expanded to infinity and the integral evaluated as described above. If $X \neq 0$, this is no longer possible. It is then natural to expand the circle of the contour out to the stationary point which lies, for $E \gg 1/2$, at $z \approx -iE/X$. In doing so we pick up the contributions from the poles at $\pm 2\pi mi$, $m < E/2\pi X$, and these give the second sum in (65). The integral around the contour can then be estimated to give the error term quoted [15]. Of course, one could even try to evaluate this integral directly by expanding about the stationary point, but this turns out to be very subtle and complicated due to the presence of the poles of the integrand. Nevertheless, RIEMANN was able to do just this and so obtain an explicit series, asymptotic as $E \to \infty$ (*i.e.* in what corresponds to the semi-classical limit), for the error term in (65). However, he never published it. The result was eventually found, scribbled amongst other calculations in his private papers, by SIEGEL in 1932. Having only the answer, SIEGEL was then able to find a proof and the result is known as the *Riemann-Siegel formula* (for a beautiful account see [16]). I will not write down the whole series, but, so as to give some idea of its form, the first term corresponds to replacing $O(E^{-1/4})$ in (66) by

$$(-1)^N \left(\frac{E}{2\pi} \right)^{-1/4} \frac{\cos 2\pi(p^2 - p - 1/16)}{\cos 2\pi p} + O(E^{-3/4}), \tag{69}$$

where N is the integer part of $\sqrt{E/2\pi}$, and p its fractional part.

The Riemann-Siegel formula represents one of the deepest and most important results for the Riemann zeta-function. It provides a systematic method for rewriting the (infinite) Dirichlet series as a finite sum, which exactly satisfies the functional equation, plus explicit correction terms. The role of these correction terms is, in fact, not difficult to see. The sum in (66) is discontinuous when $\sqrt{E/2\pi}$ is an integer, but Δ_R has no such discontinuities. It is the job of the m-th correction term to remove the discontinuity in the $(m-1)$-th derivative of the finite sum, and this can be seen immediately for the first of these terms, as given by (69). One of the most interesting features of the Riemann-Siegel formula is the factor of 2 which appears in front of the principal sum (66). This is the crucial sign that the result depends upon *resummation.* If we take the original Dirichlet series and split the sum at $\sqrt{E/2\pi}$, then the infinite second series may be resummed to give, approximately, the complex conjugate of the finite first series. Thus the factor of 2 marks the fact that (66) really contains contributions from all integers; that is, it cannot be derived by truncating the Dirichlet series and then imposing the functional equation. In fact, the approximate functional equation (65) represents an even stronger and more general statement of this resummation property: truncating the Dirichlet series at any place X, the infinite second series $(X < n < \infty)$ can be resummed to give an explicit finite answer involving of the complex conjugates of the contributions to the first sum $(1 \leq n < X)$. This is a truly amazing analytical property for a series to possess. It is, of course, explicit in both the derivations of the approximate functional equation given above. It is worth emphasizing that, once the contributions from all integers are dealt with consistently, the result (66) automatically satisfies the functional equation (12).

Not only does the Riemann-Siegel formula represent a remarkable analytical property of Δ_R, it also provides the most efficient method for actually computing the function when E is real. It is, for example, much more efficient than the Euler-Maclaurin method described above. Very roughly, a feel for the number of terms required in the two cases can be obtained by considering the sums in (65) and (66). The Riemann-Siegel sum (66) obviously requires $\sqrt{E/2\pi}$ integers. The Euler-Maclaurin method involves the direct evaluation of the Dirichlet series, subtracting away the leading divergence. Now, formally, (65) with $X > E/2\pi$ gives an approximation for Δ_R involving just the truncated original Dirichlet series, because the second sum is then empty. Hence in some sense $E/2\pi$ terms of the original series are required to approximate the function if their complex conjugates are not systematically incorporated. Thus one would expect, and indeed it can be shown [15], that at least $E/2\pi$ terms must be included in the sum involved in the Euler-Maclaurin method. And, since we are interested in $\Delta_R(E)$ as $E \to \infty$, this is significantly more than in the Riemann-Siegel formula. It is again worth emphasizing the point that (65), which was the result of the resummation of contributions from integers $n > X$, leads to the intuition that, in some sense, only the integers $n < E/2\pi$ are needed to approximate Δ_R using (16).

Having seen that the approximate functional equation and the Riemann-Siegel formula are of central importance in the theory of the Riemann zeta-function, the obvious question now is: do these results have analogues in quantum chaology? Certainly the approximate functional equation is very closely related to the actual functional equation and this does have a semi-classical analogue which, it might then be hoped, would have associated with it a finite approximation to the Dirichlet series (48). The main problem is, of course, that the derivations of (65) given above were essentially number-theoretical, as noted when the associated derivations of the functional equation were given. What is needed is a method of obtaining (65) directly from the functional equation which can also be generalized to the semi-classical formulae. This is the problem that I will now discuss.

The starting point of this approach is the equation obtained by substituting (16) directly into (12):

$$\exp[-i\pi\overline{N}_R(E)] \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \exp[iE \log n] = \exp[i\pi\overline{N}_R(E)] \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \exp[-iE \log n]. \tag{70}$$

It is, however, immediately apparent that this equation has only formal meaning because the two sides share no common domain of convergence. Hence any derivation based upon it will be only formal. This then obviously requires some discussion.

In quantum chaology, the use of formal arguments is standard and, as things stand at present, almost inevitable if nontrivial analytical results are to be obtained for general systems. Thus, for example, the manipulation of divergent series is an essential part of the well-known work of GUTZWILLER [3] and BERRY [10]. The aim usually is, by systematic operation on the divergent series, to obtain a finite result. This is what will be done below. Of course, the method to be used is nonrigorous, but certainly no more so than in the cases cited above. Perhaps it is best viewed as «experimental mathematics», the results to be tested numerically later. If these tests are in support, one might then feel that a «truth» has been uncovered, if not proved. One might even feel that the original result was, in fact, «derived», at least in the sense of Newton, Euler and Ramanujan, if not in the sense of Bourbaki.

Having made the step of writing down (70), the result follows directly [41]. If the two sides of the equality are denoted by $F(E)$, then consider the operation

$$\int\limits_{\log(\sqrt{2\pi}X/\sqrt{E})}^{\infty} \mathrm{d}z \int\limits_{-\infty}^{\infty} \mathrm{d}y\, F(E+y) \exp[-izy] \exp[-y^2/\varepsilon^2(E)]. \tag{71}$$

If $\varepsilon(E)$ is chosen so that $\varepsilon(E)/E \to 0$ as $E \to \infty$, then in this limit it is consistent to expand $\overline{N}_R(E+y)$ to first order in y. Substituting both sides of (70) into (71) and interchanging the orders of summation and integration then gives

$$\exp[-i\pi\overline{N}_R(E)] \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \exp[iE \log n] \cdot \frac{1}{2} \operatorname{erfc}\left[\frac{\varepsilon(E)}{2} \log\left(\frac{X}{n}\right)\right] \approx \tag{72}$$

$$\approx \exp[i\pi\overline{N}_R(E)] \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \exp[-iE \log n] \cdot \frac{1}{2} \operatorname{erfc}\left[\frac{\varepsilon(E)}{2} \log\left(\frac{2\pi n X}{E}\right)\right],$$

where

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int\limits_{z}^{\infty} \mathrm{d}y \exp[-y^2] \tag{72a}$$

and the approximation comes because here I will retain only the first two terms in the expansion of $\overline{N}_R$ (it would be exact if all terms of the expansion in y are retained) and also because the approximate form (27) has been used for $\overline{d}_R$. If now the further condition is made that $\varepsilon(E) \to \infty$ as $E \to \infty$, then in this limit the erfc-function terms approach step functions, the one on the left selecting terms in the sum with $n > X$, and the one on the right selecting terms in that sum with $n < E/2\pi X$ (as $x \to \infty$, $\operatorname{erfc} x \to 0$ and $\operatorname{erfc}(-x) \to 2$). If for the moment these terms are indeed approximated by steps, then combining (72) with the original Dirichlet series gives precisely the sums in the approximate functional equation (65). Furthermore, in approximating them by step functions, we are failing to treat correctly those terms in the sums which lie asymptotically close (as $E \to \infty$) to the actual steps. Thus the error might be expected to be of the order of these mistreated terms; that is, $O(X^{-1/2})$ and $O(\sqrt{X/2\pi E})$, respectively. This matter will be considered in more detail below, but it is worth noting that these are just the errors in (65).

Now setting $X = \sqrt{E/2\pi}$ in (72) and using the symmetry and normalization of the integrals, we obtain the final result

$$(73) \qquad \Delta_R(E) \approx B_R(E)\left[2\sum_{n=1}^{\infty}\frac{1}{\sqrt{n}}\cos(\pi\overline{N}_R(E) - E\log n)\cdot \cdot\frac{1}{2}\operatorname{erfc}\left\{\frac{\varepsilon(E)}{2}\log\left(n\sqrt{\frac{2\pi}{E}}\right)\right\}\right].$$

If the erfc-function term in (73) is approximated, as above, by the step function which it approaches in the limit $E \to \infty$, then only those contributions coming from integers $n \leqslant \sqrt{E/2\pi}$ are retained and the sum reduces to the finite approximation in (66). The error then comes because the step is not sharp on asymptotically fine scales; in fact $\varepsilon(E)$ may obviously be chosen so that the transition region captures asymptotically many contributions as $E \to \infty$. Thus these contributions which lie on either side of the cut-off in (66) are mistreated. They are of the order $(E/2\pi)^{-1/4}$ and so again one might guess that this would be the error in the approximation. It is just the order quoted in (66).

This error is precisely what is given by the asymptotic expansion in the Riemann-Siegel formula. Thus the argument presented above suggests that this expansion might be obtained by combining a careful consideration of the terms mistreated when the erfc-function term in (73) is approximated by a step function with the systematic inclusion of the complete Taylor series for $\overline{N}_R(E+y)$ into (71). Indeed, in the first case this would be a natural guess since, as discussed earlier, the role of this expansion is to remove the discontinuities in all of the derivatives (including the zeroth) of the finite sum in (66) and so it is understandable that these corrections would depend on the terms in (73) near the cut-off. As evidence that this is indeed the case, I now offer a crude treatment of these terms which yields the first correction term (69) in the Riemann-Siegel formula.

An explicit formula for the error made in the approximation (66) is

$$(74) \qquad R(E) = \exp[i\pi\overline{N}_R(E)]\sum_{n>[\sqrt{E/2\pi}]}\frac{1}{\sqrt{n}}\exp[-iE\log n] - - \exp[-i\pi\overline{N}_R(E)]\sum_{n\leqslant[\sqrt{E/2\pi}]}\frac{1}{\sqrt{n}}\exp[iE\log n],$$

where $[x]$ represents the integer part of x. Now changing to the variable $t = = \sqrt{E/2\pi}$, using the asymptotic form (26), and expanding around the end-point $[t]$, the first sum is approximately

$$(75) \qquad \text{first term} \approx \frac{1}{\sqrt{t}}\exp[i\pi(t^2\log t^2 - t^2 + 7/8)]\sum_{m=1}^{\infty}\exp\left[-2\pi i t^2\log([t]+m)\right],$$

where the expectation that the result only depends on terms near the cut-off has been used explicitly by taking for the amplitude $n^{-1/2}$ its value at the cut-off. The logarithm in the exponent of the summand is now expanded around $[t]$ to second order in m (all higher orders tend to 0 as $t \to \infty$) and the sum can then be evaluated to give

(76) first term $\approx$

$$\approx -\frac{1}{\sqrt{t}} \exp[i\pi(t^2 \log t^2 - t^2 + 7/8 - 2t^2 \log[t])] \frac{\exp[-2\pi i t^2/[t]]}{1 + \exp[-2\pi i t^2/[t]]}$$

as $t \to \infty$. Similarly the second term in (74) can be approximated in this limit by

(77) second term $\approx$

$$\approx \frac{1}{\sqrt{t}} \exp[-i\pi(t^2 \log t^2 - t^2 + 7/8 - 2t^2 \log[t])] \frac{1}{1 + \exp[-2\pi i t^2/[t]]}.$$

Then combining these two contributions gives precisely the expression in (69).

Hence approximating the erfc-function term in (73) by a step function we obtain the approximation (66) and, moreover, we can even see how the Riemann-Siegel expression for the error term arises. But, of course, it was not actually necessary to make this approximation. That is, not only does (73) provide a route to a formal derivation of (66), it also represents a *new approximation* for Δ_R in its own right. This new result clearly takes into account the contributions near the cut-off, and it has no discontinuities. Thus it represents a different way of encoding the end-point contributions which, at least when continued to all orders, would be equivalent to the Riemann-Siegel formula. This is important because it provides a way of testing the formal manipulations which led to it; that is, not only do these manipulations give a formal derivation of the known approximation (66), they also lead to a new approximation which, from the above discussion, we would expect to be better. And indeed numerical tests suggest that it is. These are shown in fig. 6-8, which plot the actual function Δ_R, its approximation (66) and the new approximation (73).

In fig. 8 the first approximation actually misses two zeros which are captured by the second (missing zeros are an important problem—it is a phenomenon that cannot happen with the quantization rule (56), as noted in the discussion). In (73) the function $\varepsilon(E)$ is unspecified beyond the asymptotic growth laws already given. In the above tests we chose $\varepsilon(E) = 2\sqrt{E}$, but the results are not much affected by the actual choice. If I had also plotted the Riemann-Siegel formula using just the first correction term (69), the result would have been indistinguishable from the actual function on these scales. Finally, in [41] a different type of smoothing to the Gaussian used in (71) was employed.

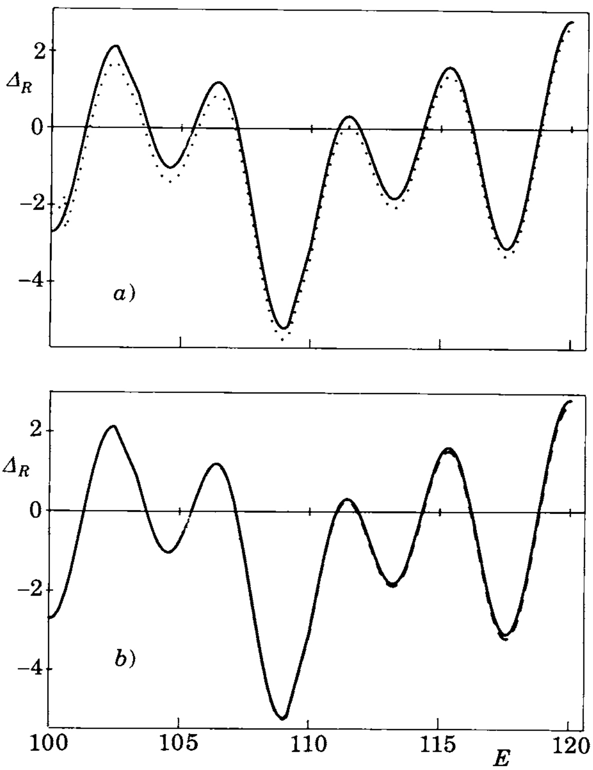

Fig. 6. – The function Δ_R (black line) with in a) the approximation (66) (dotted line) and in b) the approximation (73) (dashed line). In a) a discontinuity of (66) can be seen at $E = 2\pi \cdot 4^2 = 100.53\ldots$. In b) the dashed line is almost completely hidden by the bold line.

Indeed, as noted there, any method of effectively restricting the integration range to $\pm\varepsilon(E)$ is allowed.

The point of the method outlined above is that it immediately extends to the general semi-classical formulae for the spectral determinant. Thus the representation of Δ as a sum over pseudo-orbits (48) and the semi-classical functional equation (45) can be combined to give the analogue of (70) called the *formal functional equation*. Then applying exactly the same methods as above yields the analogue of (73):

$$(78)\qquad \Delta(E) \approx B(E)\left[2\sum_n C_n(E)\cos(\pi\bar{N}(E) - \mathcal{S}_n(E)/\hbar)\cdot \frac{1}{2}\,\mathrm{erfc}\left\{\frac{\alpha(\hbar)}{2\hbar}(\mathcal{T}_n(E) - \pi\hbar\bar{d}(E))\right\}\right],$$

where $\alpha(\hbar)$ is some function satisfying the conditions $\alpha(\hbar) \to 0$ but $\alpha(\hbar)/\hbar \to \infty$ as $\hbar \to 0$. Approximating the erfc-function term by a step function now gives a fi-

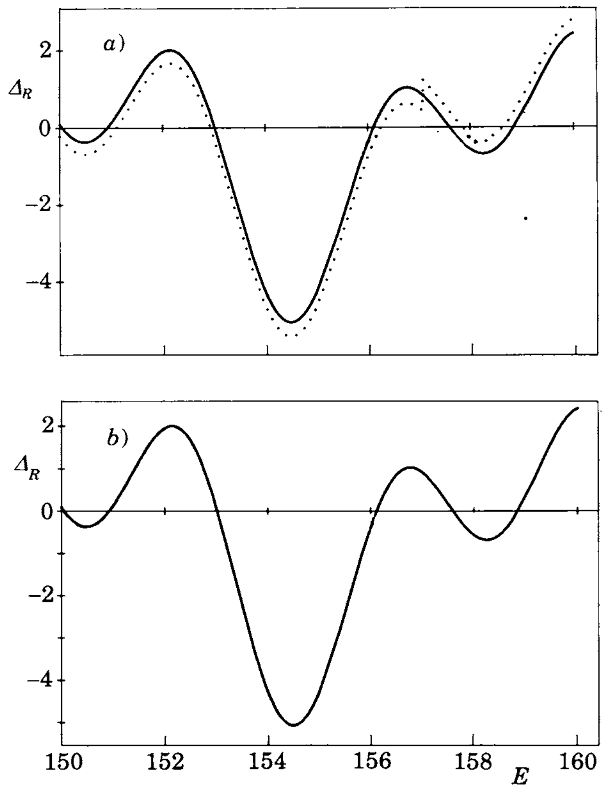

Fig. 7. – As for fig. 6 *a*) and *b*). In *b*) the dashed line is almost completely hidden by the black line. In *a*) a discontinuity of (66) can be seen at $E = 2\pi \cdot 5^2 = 157.08\ldots$.

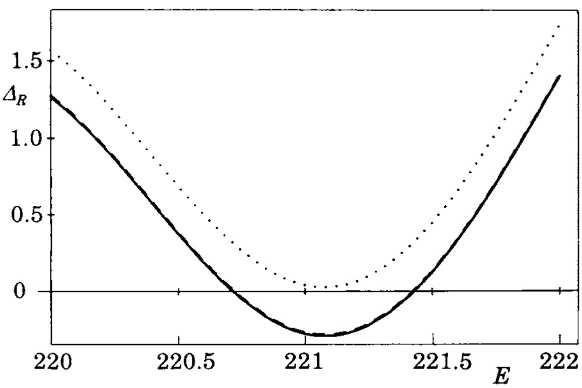

Fig. 8. – As for fig. 6 and 7.

nite sum over pseudo-orbits with period $\mathscr{T}_n \leqslant \pi\hbar\bar{d}$. This finite sum was, elaborating the earlier suggestion in [31], first conjectured by BERRY and KEATING in [25], where it was called the *Riemann-Siegel lookalike* formula. Furthermore, one can also go beyond (78) and derive, in this formal sense, its generalization, the analogue of the approximate functional equation (65) [41]. The most important aspect of these results is that they exhibit exactly the same kind of resummation as do (65) and (66). Hence all of the conclusions drawn about the analytic structure of the Dirichlet series (16) may be transferred directly to (48). The idea of this kind of periodic-orbit resummation had earlier been discussed in [10], where it was called *bootstrapping*. An example of this will be given in the next section.

Finally, the lookalike formula has recently been tested for the hyperbola billiard [30]. In this case it can actually be compared to the complete pseudo-orbit sum, for which there is strong numerical evidence of conditional convergence (see sect. 5). The result is good agreement, especially with regard to the factor of 2—the important hallmark of resummation.

7. – The pair correlation of the zeros.

One of the most interesting and stimulating discoveries relating to the semi-classical properties of energy levels concerns their statistical distribution. It has been found that in typical quantum systems the energy levels are distributed in one of a small number of universal ways and that, furthermore, in a given system the particular distribution taken depends only upon whether the underlying classical dynamics is chaotic or not and whether it possesses any anti-canonical symmetries (*e.g.*, time-reversal symmetry). This remarkable universal behaviour is revealed by studying the statistics of the energy levels. For nonfermionic quantum systems which have chaotic classical limits these have been found to be either the GUE statistics of random-matrix theory, if the system does not possess an anticanonical symmetry, or the GOE statistics if it does (for a review see, for example, [9]). The reason for the particular interest in this result here is that, amazingly, the nontrivial zeros of the Riemann zeta-function have also been found to have distributional statistics which are exceedingly close to the GUE forms [11].

This last important discovery significantly strengthens the analogy between the equations describing the zeta-function and the formulae of quantum chaology. It suggests that, if there is a system whose quantum energy levels are the zeros E_n and whose periodic orbits have the properties discussed in sect. 4, then its classical mechanics will possess no anticanonical symmetries. For the purposes of this lecture, however, its most important consequence is that the zeta-function is not only of considerable use in understanding the analytic structure of the periodic-orbit formulae, but it is also an invaluable mathematical model for studying the semi-classical connection between the periodic orbits and the energy level statistics. In particular, it represents an ideal example for investigating which properties of the orbits give rise to the observed universality.

In this section, I will consider how the zeta-function can be used to provide important insights into precisely how GUE universality arises in statistics which are bilinear in the density of states. The reason for choosing these particular statistics is that there already exists a semi-classical theory, due to BERRY [10], which relates the statistical universality directly to universal properties of the long periodic orbits. There are, however, still several questions left unanswered by this theory, and it is on these that the zeta-function can, because a great deal is known about the distribution of the prime numbers, cast some new light.

The particular function that I will concentrate upon is the *spectral form factor* $K(\tau)$. Consider scaling a spectrum $\{E_n\}$ so that it has unit mean density (*i.e.* multiplying each energy level by the mean density). Let the scaled spectrum be denoted by $\{e_n\}$ and have density of states

$$d(e) = \sum_n \delta(e - e_n). \tag{79}$$

Then the form factor is essentially the Fourier transform of the pair correlation function of the scaled spectrum:

$$K(\tau) \equiv \int_{-\infty}^{\infty} \mathrm{d}y \exp[2\pi i y \tau]\langle [d(e + y/2) - 1][d(e - y/2) - 1]\rangle_e , \tag{80}$$

where the average in the correlation function extends over many levels. The more commonly studied bilinear statistics, such as the number variance $\Sigma^2(L)$, may be expressed as simple transforms of the form factor; for example,

$$\Sigma^2(L) = \frac{2}{\pi^2}\int_0^{\infty} \mathrm{d}\tau \frac{K(\tau)}{\tau^2}\sin^2(\pi L \tau). \tag{81}$$

A semi-classical representation of $K(\tau)$ is obtained by substituting the periodic-orbit formula for the density of states (29), when appropriately scaled, directly into (80). The result may be written in the form [10]

$$K(\tau) \approx \frac{4\pi^2}{h\bar{d}} \cdot \tag{82}$$

$$\cdot \left\langle \sum_{p,\,q} \sum_{j,\,k=1}^{\infty} A_{p,k} A_{q,j} \exp[i(kS_p - jS_q)/\hbar - i\pi(k\alpha_p - j\alpha_q)/2]\delta(T - [kT_p + jT_q]/2)\right\rangle,$$

where

$$T = \tau h \bar{d} \tag{83}$$

and has the dimensions of time. We now need to evaluate the double sum over orbits in (82). Unfortunately, this is not possible in general because the pairwise distribution of the periodic-orbit actions is not known. However, BERRY argued that, for sufficiently small values of τ, the double sum could be approximated by the sum of its diagonal terms, *i.e.*

$$K(\tau) \approx \frac{4\pi^2}{h\bar{d}} \sum_p \sum_{k=1}^{\infty} A_{p,k}^2 \delta(T - kT_p). \tag{84}$$

This periodic-orbit sum can now be evaluated using the classical sum rule of Hannay and Ozorio de Almeida [24], eq. (40). Note that this sum rule represents universal classical information about long periodic orbits, and it is this that gives rise to the quantum universality in the spectral statistics. It is worth pointing out explicitly that, although we are interested in small τ, in $D \geqslant 2$ dimensions T is still semi-classically large because $\bar{d}$ in (83) is of the order h^{-D} [10]. One further point is that the classical sum rule relates a weighted string of δ-functions, each centred on the period of one of the periodic orbits, to a smooth function. Thus it represents the average behaviour of the distribution. If, however, this distribution is probed on scales corresponding to the mean separations of the δ-functions, then their discrete nature will become important and it is this which ultimately gives rise to the breakdown of universality in spectral statistics [10]. This breakdown also has very important consequences for the Riemann zeros [12], but I will not pursue it any further here.

The result of using (40) is that, when τ is sufficiently small,

$$K(\tau) \approx \tau \tag{85}$$

for systems which are chaotic and possess no anticanonical symmetries [10]. However, this behaviour cannot continue for all τ since it would cause the integral for the number variance in (81) to diverge. In fact, irrespective of the semi-classical representation (82), the pair correlation function in (80) must behave like $\delta(y)$ for $y \ll 1$ (the mean spacing in the scaled spectrum), unless the energy levels exhibit some kind of systematic degeneracy. This then means that $K(\tau) \to 1$ for $\tau \gg 1$.

The conclusion of the above considerations is thus that $K(\tau) \approx \tau$ for $\tau \ll 1$, and $K(\tau) \approx 1$ for $\tau \gg 1$. Now the form factor for the GUE is known to be

$$K_{\mathrm{GUE}}(\tau) = \begin{cases} \tau, & \text{if } 0 \leqslant \tau < 1, \\ 1, & \text{if } \tau \geqslant 1, \end{cases} \tag{86}$$

and so, not only does the semi-classical theory show how long periodic orbits give rise to universality, but, in the regions where $K(\tau)$ is either known or can be evaluated, it agrees with the GUE form.

Of course, what we would like to be able to do now is to address the questions left unanswered by the above theory. Thus we would like to have a way of probing the transition region between the one governed by the classical sum rule ($\tau \ll 1$) and the one governed by the discreteness of the quantum spectrum

($\tau \gg 1$). This is the region where the off-diagonal terms in (82) must start to become important. It would be interesting to see if in general there is a discontinuity of slope, as there is in the GUE form factor (86). Furthermore, we would also like a truly semi-classical theory for the behaviour when $\tau \gg 1$. That is, we would like to understand how the off-diagonal terms cancel the contributions from the diagonal terms, leaving just the necessary remainder.

In fact, the region $\tau \gg 1$ is interesting for another reason. Since we know that $K(\tau) \approx 1$, this means that

$$\bar{d} \approx \frac{4\pi^2}{h}\left\langle \sum_{p,q} \sum_{j,k=1}^{\infty} A_{p,k} A_{q,j} \exp\left[i(kS_p - jS_q)/\hbar - i\pi(k\alpha_p - j\alpha_q)/2\right] \cdot \delta(T - [kT_p + jT_q]/2) \right\rangle \tag{87}$$

for $T \gg h\bar{d}$. This is a semi-classical sum rule of the type discussed by BERRY in [10]. It represents an expression for the mean density of states, the first term in the semi-classical formula (29), in terms of pairs of orbits with mean period $T \gg h\bar{d}$. In fact, it will be shown below that these orbits will typically both have periods very much greater than $h\bar{d}$; that is, they are both semi-classically very long. Hence the expression relates the first term in (29) to some of the late terms. This then is an example of the bootstrapping of periodic orbits mentioned at the end of the last section. One can also write down expressions like (87) which relate individual short periodic orbits to sets of asymptotically long ones [13]. Hence it may be hoped that an understanding of the general topic of bootstrapping, and perhaps of the resummation results derived in the last section, may be achieved by studying the relationship between the periodic orbits and energy level statistics.

My aim now is to show that the analogue of (82) for the Riemann zeros can be evaluated for all values of τ. This problem was first studied by MONTGOMERY [42], who conjectured that in this case the form factor is exactly given by the GUE result (86). Unfortunately, he did not give any details of the analysis which led him to this conclusion for the important region $\tau > 1$. Therefore, given the considerable importance of the general problem in quantum chaology, I will present my own «derivation». This will be seen immediately to be based in number theory, but to be heuristic in nature (and hence nonrigorous). The aim will be to keep as closely as possible to semi-classical language and concepts in the hope that the results obtained may then provide clues as to how progress can be made on the general problem. Many of the manipulations to be employed are discussed in more detail in [13, 14]. All of the number-theoretical analysis is contained in the appendix.

Substituting into (82) the zeta-function analogues (27) and (33)-(36), the form factor for the zeros of Δ_R is given, as $E \to \infty$, by

$$(88) \qquad K_R(\tau) \approx \frac{1}{\log\left(\dfrac{E}{2\pi}\right)} \left\langle \sum_{p,q} \sum_{j,k=1}^{\infty} \frac{\log p \log q}{\sqrt{p^k q^j}} \exp[iE(k \log p - j \log q)] \cdot \delta\left(T - \frac{1}{2}\log(p^k q^j)\right) \right\rangle_E ,$$

where the first sum is over all pairs of primes, p and q, and $T = \tau \log(E/2\pi)$. The average may be taken to extend over a range $\gamma(E)$ around E, where $\gamma(E) \to \infty$ but $\gamma(E)/E \to 0$ as $E \to \infty$. It may then be seen that this will wash away terms in the sum for which $\gamma(E)\log(p^k/q^j) \gg 1$ and so as $E \to \infty$ the main contributions will come from terms with $p^k \sim q^j$. If either $k > 1$ or $j > 1$, then the sum over these contributions will converge, and so (88) is dominated by the terms with $j = k = 1$. Hence we have that

$$(89) \qquad K_R(\tau) \approx \frac{1}{\log\left(\dfrac{E}{2\pi}\right)} \left\langle \sum_{p,q} \frac{\log^2 p}{p} \exp[iE \log(p/q)]\, \delta(T - \log p) \right\rangle_E ,$$

where the fact that $p \sim q$ has been used in all of the slowly varying amplitude terms.

We can now split (89) into contributions from the diagonal and off-diagonal terms in the double sum. The diagonal contribution is given by

$$(90) \qquad K_R^{(\mathrm{d})}(\tau) = \frac{1}{\log\left(\dfrac{E}{2\pi}\right)} \sum_p \frac{\log^2 p}{p}\, \delta(T - \log p),$$

which can be evaluated using the prime-number theorem (39) to give

$$(91) \qquad K_R^{(\mathrm{d})}(\tau) \approx \tau .$$

This is the equivalent of the general result (85) obtained from the semi-classical sum rule.

Now, of course, we must deal with the off-diagonal contributions to (89), in order to see how good an approximation (91) is to the full form factor. These are given by

$$(92) \qquad K_R^{(\mathrm{off})}(\tau) \approx \frac{2}{\log\left(\dfrac{E}{2\pi}\right)} \left\langle \sum_{p>q} \frac{\log^2 p}{p}\, \delta(T - \log p) \cos\{E \log(p/q)\} \right\rangle_E .$$

To evaluate this expression, we obviously need to know about the pairwise correlation properties of the prime numbers, and it is here that we can make use of important number-theoretical information.

In sect. **1**, the function $\pi(X)$—the number of primes less than or equal to X—was mentioned. The prime-number theorem, previously given in the form (39), is that as $X \to \infty$

$$\pi(X) \sim \frac{X}{\log X}\,. \tag{93}$$

The pair correlation properties of the primes can then be represented by a similar function $\pi_m(X)$, defined to be the number of primes p not exceeding X and such that $p - m$ is also a prime. There is a strong conjecture, due to HARDY and LITTLEWOOD [43], for the asymptotic behaviour of $\pi_m(X)$ as $X \to \infty$:

$$\pi_m(X) \sim \alpha(m) \frac{X}{\log^2 X}\,, \tag{94}$$

where, if m is even, $\alpha(m)$ is given in terms of two products, one over all odd primes q, and the other over odd primes p which divide m:

$$\alpha(m) = 2 \prod_{q>2} \left(1 - \frac{1}{(q-1)^2}\right) \prod_{\substack{p>2 \\ p|m}} \left(\frac{p-1}{p-2}\right). \tag{95}$$

If m is odd, then $\alpha(m) = 0$. It is this conjecture that will form the basis of the method to evaluate (92). It is, however, not yet quite in the required form, because $\alpha(m)$ is an irregular number-theoretical function of m and its functional behaviour is far from obvious. What is needed is its «average» behaviour. This can be obtained, rigorously, by complicated but standard number-theoretical manipulations (D. R. HEATH-BROWN: personal communication). However, it can also be obtained nonrigorously by more familiar probabilistic methods. I will outline this procedure in the appendix, where, for completeness, I also give a probabilistic derivation of (95). The result is that, as $m \to \infty$, the average behaviour of $\alpha(m)$ is approximately given by

$$\langle \alpha(m) \rangle \approx 1 - \frac{1}{2m}\,. \tag{96}$$

It follows from (93) and (94), and it is shown explicitly in the appendix, that $\alpha(m)$ is the ratio of the mean density of asymptotically large prime pairs with separation m to the square of the mean density of primes. Hence, if the primes were truly randomly (*i.e.* Poisson) distributed, then only the first term would be present in (96). The second term thus represents a small correction, vanishing as $m \to \infty$, to their pairwise randomness.

We are interested in the behaviour of (92) as $E \to \infty$, and this corresponds, for all τ, to $T \to \infty$. Hence, using the above interpretation of $\alpha(m)$, we have, approximating the sums by integrals with the appropriate weighting obtained from (93), that

(97) $K_R^{(\text{off})}(\tau) \approx$

$$\approx \frac{2}{\log\left(\frac{E}{2\pi}\right)} \left\langle \int_0^{\infty} \frac{dx}{x}\, \delta(T - \log x) \int_{y_0}^{x-1} dy \langle \alpha(y) \rangle \cos\left\{E \log\left(\frac{x}{x-y}\right)\right\} \right\rangle_E ,$$

where the lower limit of the y integral denotes the fact that the primes are limited as to how close they can be to each other. Of course, this lower limit is effectively 2, however it turns out that it is only the fact that there is such a limit, and not its actual value, which is important for the result. As noted earlier, the averaging process means that only values of $y \ll x$ contribute in (97). Thus the logarithm in the integrand may be expanded, and the upper limit of the y integral taken to be infinity (see [13]). This then gives

$$K_R^{(\text{off})}(\tau) \approx \frac{2}{\log\left(\frac{E}{2\pi}\right)} \left\langle \int_{y_0}^{\infty} dy \langle \alpha(y) \rangle \cos\left\{2\pi y \left(\frac{E}{2\pi}\right)^{1-\tau}\right\} \right\rangle_E , \tag{98}$$

or

$$K_R^{(\text{off})}(\tau) \approx \frac{2}{\log\left(\frac{E}{2\pi}\right)} \left(\frac{E}{2\pi}\right)^{\tau-1} \left\langle \int_{y_0\left(\frac{E}{2\pi}\right)^{1-\tau}}^{\infty} dz \left\langle \alpha\left\{z\left(\frac{E}{2\pi}\right)^{\tau-1}\right\}\right\rangle \cos\{2\pi z\} \right\rangle_E . \tag{99}$$

Now (99) shows two different types of behaviour depending upon whether $\tau < 1$, or $\tau > 1$. If $\tau < 1$, then as $E \to \infty$ the lower limit of the integral approaches the upper limit and the prefactor tends to zero. Hence, for $\tau < 1$,

$$K_R^{(\text{off})}(\tau) \to 0 \tag{100}$$

as $E \to \infty$. This is just the behaviour assumed in Berry's semi-classical theory to hold for $\tau \ll 1$, thus allowing the form factor to be approximated by the diagonal sum (84). For the Riemann zeta-function we have here shown that this is indeed the case, and that this behaviour actually persists up to $\tau = 1$.

For $\tau > 1$, the argument of α tends to infinity as $E \to \infty$ for all values of the integration variable. Hence we can use the result (96). The first term in this expression leads to a simple oscillatory integral and hence gives 0. The second term leads to an integral which diverges as $E \to \infty$ because the lower limit then approaches zero. Evaluating it asymptotically in this limit gives that

$$K_R^{(\text{off})}(\tau) \approx 1 - \tau \tag{101}$$

for all $\tau > 1$. Combining (101) and (91) then gives for K_R the GUE result (86).

Hence K_R can indeed be shown to have the complete GUE form, even in the intermediate range which was inaccessible to the general semi-classical theory. The discontinuity in slope of the result is obviously related to the fact that there is a finite lower limit in the integral (97); that is, to the fact that the primes have a size-independent, finite minimum separation. Furthermore, the behaviour when $\tau > 1$, where the off-diagonal terms somehow conspire to cancel off the contributions from the diagonal terms to give the bootstrapping formula (87), may in this case be seen to arise directly from small corrections to pairwise randomness in the distribution of the primes; the random element played no part.

It might be hoped that some of these features will also apply in the general case, but here there is an obvious difficulty due to the fact that we do not yet know enough about the periodic orbits of chaotic systems to be able to derive correlation information analogous to (96). Nevertheless, it turns out that, guessing a statistical distribution of the form (96) for the exponentials of the periods of periodic orbits, the off-diagonal terms do indeed cancel the contributions from the diagonal terms in (82). There is, however, a problem with evaluating the remainder. In the case of the zeta-function, its value (the first term in (101)) is related to the form of the lower limit in (97). For general systems the corresponding lower limit must depend on $\hbar$ if the correct answer is to be obtained. But it is a classical quantity and so there is obviously an important difference in this case.

One way of investigating this is to observe that the direction of the above calculation can be reversed. It was noted in sect. 3 that the formula for d_R can be Fourier-transformed to give a weighted density-of-states formula for the primes. Then *assuming* that the Riemann zeros are pairwise GUE distributed, one can follow the steps of the above calculation to find the pairwise distribution of the primes. This has, in fact, been done rigorously by GALLAGHER and MUELLER [44]. The result is a correlation function which is, at least for our purposes, equivalent to (96) over long ranges. Over short ranges it gives a result consistent with the fact that there is a lower limit to the separation between the primes. Mimicking this calculation for general systems, one can obtain a formula for a certain periodic-orbit correlation function [45]. This shows a number of similarities to (96), but there are also some significant differences which explain the difficulties encountered above. Numerical evidence seems to confirm the existence of these important correlations and it would be interesting to see how they could be derived from a purely classical mechanical analysis.

This problem of using energy level statistics to understand periodic-orbit statistics is obviously important and needs much more study, perhaps using the resummed pseudo-orbit formula (78). In fact, I expect that their statistical distribution will be of increasing interest in the near future and that the primes will play a central role as a model «orbit spectrum».

* * *

It is a pleasure to acknowledge innumerable discussions with M. BERRY and J. HANNAY on most of the topics covered here. I am also very grateful to H. MADER and J. ROBBINS for their help in preparing these notes. The research described was carried out whilst I was funded by the Royal Society.

APPENDIX

Probabilistic number theory and the pairwise distribution of the primes.

The use of probabilistic methods in number theory is a well-established technique. In some important cases their application can be made rigorous (see, for example, [46-49]), but they are also frequently used simply to derive heuristic estimates for the behaviour of number-theoretical functions (see, for example, [50, 51]). It is this second kind of approach that I will be employing below. I have discussed this method of analysis and applied it to the study of periodic orbits elsewhere [52] and so I will here give only a brief summary of the actual results that will be needed.

From (93), the density of primes near a given large integer n is $1/\log n$. This density can be interpreted as a probability on the set of integers (which are thus taken to be the sample space). Hence the probability $P(n)$ that n is a prime may be taken to be, as $n \to \infty$,

$$P(n) \sim \frac{1}{\log n}. \tag{A.1}$$

Similarly, it follows from the conjecture (94) that the probability that both n and $n-m$ are primes is, as $n \to \infty$,

$$P(n, n-m) \sim P^2(n)\,\alpha(m). \tag{A.2}$$

The use of probabilistic language applies not only to the distribution of the primes, but to the divisibility of integers by them as well. Of the set of positive integers, a fraction $1/p$ are divisible by the prime p (for later convenience we can restate this as saying that the integers take p possible values modulo p, but only those congruent to 0 modulo p are divisible by p). Hence, given a large randomly picked integer, the probability that it is divisible by p will be precisely $1/p$. Similarly, the probability that an integer is divisible by p but not by p^2 is $(p-1)/p^2$. Now the important point is that the fundamental theorem of arithmetic means that divisions by different primes are independent; hence the probability that a given integer is divisible both by the prime p and the prime q is $1/pq$, which factorizes into a product of the independent probabilities. Thus, with respect to probability of divisibility, the primes are *statistically independent* and it is essentially this point that allows for the nontrivial use of probabilistic methods.

We are now in a position to derive the form (95) of $\alpha(m)$. The first point is trivial. If n is a prime, then, apart from one exceptional case, it must be odd. Thus, if $n-m$ is also to be prime, m must be even. Hence $\alpha(m) = 0$ if m is odd.

Now for the case when m is even, consider the ratio $P(n, n-m)/P^2(n)$. The aim is to find a formula for this ratio using the above probabilistic ideas. The main point is that, if an integer is to be prime, then it must not be divisible by any smaller prime. But, because the primes are statistically independent, its divisibility by any one prime is independent of its divisibility by the others. So consider a given prime p. The probability that n is not divisible by p is $(p-1)/p$ (there are p possible values of n modulo p, and any will do except 0). Similarly, we can write down the probability that neither n nor $n-m$ is divisible by p. This is just the fraction of integers n such that $n \not\equiv 0$ modulo p and $n \not\equiv -m$ modulo p. There are now two possibilities: if $m \equiv 0 \pmod p$, there is just one condition upon n, hence the probability is $(p-1)/p$; if $m \not\equiv 0 \pmod p$, then there are two values of n modulo p that are not allowed, and so the probability is $(p-2)/p$. Finally, the product over all primes p less than n must be taken. However, it turns out that the product for $P(n, n-m)/P^2(n)$ converges (this ratio was specifically chosen for this reason—extra subtleties may have arisen if the product had not converged [47]) and so the result may be approximated, as $n \to \infty$, by the infinite product

$$\text{(A.3)} \qquad \frac{P(n, n-m)}{P^2(n)} \approx \prod_{p \nmid m} \left[\frac{(p-2)/p}{(p-1)^2/p^2} \right] \prod_{p \mid m} \left[\frac{(p-1)/p}{(p-1)^2/p^2} \right].$$

When m is even, this reduces to the expression (95).

As discussed above, what we now want is the average dependence of $P(n, n-m)$ upon m. Specifically we will need this in the limit of large m. Thus effectively we are to consider picking at random some large m and then finding the expectation value of $\alpha(m)$. To do this we first write

$$\text{(A.4)} \qquad \alpha(m) = \alpha \prod_{\substack{p>2 \\ p \mid m}} \left(1 + \frac{1}{p-2}\right),$$

where α denotes the m-independent factors in (95). Then, expanding the product, $\alpha(m)$ is given as a sum over the divisors of m:

$$\text{(A.5)} \qquad \alpha(m) = \alpha \sum_{d \mid m} \beta(d),$$

where $\beta(d)$ is zero if d is even or divisible by the square of any prime; otherwise

$$\text{(A.6)} \qquad \beta(d) = \prod_{\substack{p>2 \\ p \mid d}} \frac{1}{p-2}.$$

Now consider

$$\text{(A.7)} \qquad \sum_{m=1}^{X} \alpha(m) = \alpha \sum_{d \leq X/2} \beta(d) \left[\frac{X}{2d} \right],$$

where the square brackets denote the integer part. This can obviously be written as

$$\sum_{m=1}^{X} \alpha(m) = \frac{\alpha X}{2} \sum_{d \leqslant X/2} \frac{\beta(d)}{d} - \alpha \sum_{d \leqslant X/2} \beta(d) \left\{ \frac{X}{2d} \right\}, \tag{A.8}$$

where the curly brackets denote the fractional part. The object now is to estimate these two sums.

It appears from its definition (A.6) that $\beta(d)$ might behave on average as $1/d$. Hence consider the expectation value of $d\beta(d)$ for large d. Obviously this is zero if d is even. If d is odd, we can compute the expectation value by the probabilistic methods outlined above, because both d and $\beta(d)$ can be written as products over primes. With regard to a given prime p, if p does not divide d, it contributes a factor 1 to the product—this has probability $(p-1)/p$. If p divides d but p^2 does not, the contribution is $p/(p-2)$—this has probability $(p-1)/p^2$. If p^2 divides d, then the contribution to the product is 0, because $\beta(d)$ is then zero. Hence as $d \to \infty$

$$\langle d\beta(d) \rangle \approx \frac{1}{2} \prod_{p>2} \left[\left(1 - \frac{1}{p}\right) + \frac{p}{p-2} \left(\frac{p-1}{p^2} \right) \right] = \frac{1}{\alpha}, \tag{A.9}$$

where again, because the product converges, we have included all primes. Thus as $d \to \infty$ $\beta(d)$ behaves, on average, like $(\alpha d)^{-1}$. This means that the first sum in (A.8) converges as $X \to \infty$ and so it may be approximated by the infinite sum

$$\sum_{d=1}^{\infty} \frac{\beta(d)}{d} = \prod_{p>2} \left(1 + \frac{1}{p(p-2)} \right) = \frac{2}{\alpha}. \tag{A.10}$$

The second sum may be approximated by assuming that, as d runs through its range, the values of $\{X/2d\}$ are uniformly and randomly distributed between 0 and 1. Hence replacing this fractional part by its mean value, $1/2$, and using the above result for the average behaviour of $\beta(d)$, we have as $X \to \infty$ that

$$\sum_{d \leqslant X/2} \beta(d) \left\{ \frac{X}{2d} \right\} \sim \frac{1}{2\alpha} \log X. \tag{A.11}$$

This then gives as the final result that

$$\sum_{m=1}^{X} \alpha(m) \sim X - \frac{1}{2} \log X \tag{A.12}$$

as $X \to \infty$; or that, in this limit, $\alpha(m)$ behaves on average like

$$\langle \alpha(m) \rangle \approx 1 - \frac{1}{2m}, \tag{A.13}$$

which is the result quoted in the text.

REFERENCES

[1] M. C. Gutzwiller: *J. Math. Phys. (N.Y.)*, **12**, 343 (1971).
[2] M. C. Gutzwiller: *Chaos in Classical and Quantum Mechanics* (Springer, Berlin, 1990).
[3] M. C. Gutzwiller: *Physica D*, **5**, 183 (1982).
[4] J. P. Keating: *Nonlinearity*, 4, 309 (1991).
[5] A. M. Ozorio de Almeida and M. Saraceno: *Ann. Phys. (N.Y.)*, **210**, 1 (1991).
[6] M. C. Gutzwiller: *Phys. Rev. Lett.*, 45, 150 (1980).
[7] N. L. Balazs and A. Voros: *Phys. Rep.*, **143**, 109 (1986).
[8] R. Aurich, M. Sieber and F. Steiner: *Phys. Rev. Lett.*, **61**, 483 (1988).
[9] O. Bohigas: *Random matrix theories and chaotic dynamics*, in *Les Houches Lecture Series*, Vol. **52**, edited by M. J. Giannoni, A. Voros and J. Zinn-Justin (North-Holland, Amsterdam, 1991), p. 89.
[10] M. V. Berry: *Proc. R. Soc. London*, **400**, 229 (1985).
[11] A. M. Odlyzko: *The* 10^{20}*-th zero of the Riemann zeta-function and* 70 *million of its neighbors*, preprint (1989).
[12] M. V. Berry: *Nonlinearity*, **1**, 399 (1988).
[13] J. P. Keating: *The semiclassical sum rule and Riemann's zeta-function*, in *Quantum Chaos*, edited by H. A. Cerdeira, R. Ramaswamy, M. C. Gutzwiller and G. Casati (World Scientific, Singapore, 1991), p. 280.
[14] J. P. Keating: in preparation (1991).
[15] E. C. Titchmarsh: *The Theory of the Riemann Zeta-Function*, second edition (Clarendon Press, Oxford, 1986).
[16] H. M. Edwards: *Riemann's Zeta Function* (Academic, New York, N.Y., 1974).
[17] M. C. Gutzwiller: *Physica D*, **7**, 341 (1983).
[18] P. Cvitanovic and B. Eckhardt: *Phys. Rev. Lett.*, **63**, 823 (1989).
[19] G. Polya: *J. Reine Angew. Math.*, **158**, 6 (1927).
[20] A. Voros: *Commun. Math. Phys.*, **110**, 439 (1987).
[21] H. P. Baltes and E. R. Hilf: *Spectra of Finite Systems* (Wissenschaftsverlag, Mannheim, 1976).
[22] V. P. Maslov and M. V. Fedoriuk: *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston, Mass., 1981).
[23] J. M. Robbins: *Nonlinearity*, 4, 343 (1991).
[24] J. H. Hannay and A. M. Ozorio de Almeida: *J. Phys. A*, **17**, 3429 (1984).
[25] M. V. Berry and J. P. Keating: *J. Phys. A*, **23**, 4839 (1990).
[26] E. B. Bogomolny: *Semiclassical quantization of multidimensional systems*, preprint (1991).
[27] A. Selberg: *Commun. Pure Appl. Math.*, **25**, 225 (1956).
[28] J. P. Keating: *Chaos*, **2** (1), 15 (1992).
[29] G. Tanner, P. Scherer, E. B. Bogomolny, B. Eckhardt and D. Wintgen: *Quantum eigenvalues from classical periodic orbits*, preprint (1991).
[30] M. Sieber and F. Steiner: *On the quantization of chaos*, preprint (1991).
[31] M. V. Berry: *Riemann's zeta function: a model for quantum chaos?*, in *Quantum Chaos and Statistical Nuclear Physics*, edited by T. H. Seligman and H. Nishioka, *Lecture Notes in Physics*, Vol. **263** (Springer, Berlin, 1986), p. 1.
[32] B. Eckhardt and E. Aurell: *Europhys. Lett.*, **9**, 509 (1989).
[33] M. Sieber: Ph. D. Thesis, University of Hamburg (1991).
[34] M. V. Berry: *Some quantum-to-classical asymptotics*, in *Les Houches Lecture Series*, Vol. **52**, edited by M. J. Giannoni, A. Voros and J. Zinn-Justin (North-Holland, Amsterdam, 1991), p. 251.

[35] M. V. BERRY, E. B. BOGOMOLNY and J. P. KEATING: unpublished (1991).
[36] C. MATTHIES and F. STEINER: *Selberg'z zeta function and the quantization of chaos*, preprint (1991).
[37] G. TANNER and D. WINTGEN: personal communication (1991).
[38] J. DELSARTE: *J. Anal. Math. (Jerusalem)*, **17**, 419 (1966).
[39] A. P. GUINAND: *Proc. London Math. Soc.*, **50**, 107 (1948).
[40] R. AURICH and F. STEINER: *From classical periodic orbits to the quantization of chaos*, preprint (1991).
[41] J. P. KEATING: *Proc. R. Soc. London, Ser. A*, **436**, 99 (1992).
[42] H. L. MONTGOMERY: *Proc. Symp. Pure Math.*, **24**, 181 (1973).
[43] G. H. HARDY and J. E. LITTLEWOOD: *Acta Math.*, **44**, 1 (1923).
[44] P. X. GALLAGHER and J. H. MUELLER: *J. Reine Angew. Math.*, **303/304**, 205 (1978).
[45] N. ARGAMAN, E. DORON, J. P. KEATING, A. KITAEV, M. SIEBER and U. SMILANSKY: *Correlations in the actions of periodic orbits derived from quantum chaos*, preprint (1992).
[46] M. KAC: *Statistical Independence in Probability, Analysis and Number Theory* (The Mathematical Association of America, Providence, R.I., 1959).
[47] M. KAC: *Probability and Related Topics in Physical Sciences* (Interscience, London, 1959).
[48] J. KUBILIUS: *Probabilistic Methods in the Theory of Numbers, Translations of Mathematical Monographs*, **11** (Auger Mathematical Society, Providence, R.I., 1964).
[49] P. D. T. A. ELLIOTT: *Probabilistic Number Theory*, Vol. **1** and **2** (Springer, Berlin, 1979).
[50] M. R. SCHROEDER: *Number Theory in Science and Communication* (Springer, Berlin, 1984).
[51] P. RIBENBOIM: *The Book of Prime Number Records*, second edition (Springer, Berlin, 1989).
[52] J. P. KEATING: *Nonlinearity*, **4**, 277 (1991).