Error of semiclassical eigenvalues in the semiclassical limit - an asymptotic analysis of the Sinai billiard

1999 J. Phys. A: Math. Gen. 32 7317

(http://iopscience.iop.org/0305-4470/32/42/307)

J. Phys. A: Math. Gen. **32** (1999) 7317–7344. Printed in the UK PII: S0305-4470(99)00872-0

Error of semiclassical eigenvalues in the semiclassical limit—an asymptotic analysis of the Sinai billiard

Per Dahlqvist
Mechanics Department, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received 13 January 1999, in final form 13 August 1999

Abstract. We estimate the error in the semiclassical trace formula for the Sinai billiard under the assumption that the largest source of error is due to penumbra diffraction: namely, diffraction effects for trajectories passing within a distance $R \cdot \mathrm{O}((kR)^{-2/3})$ to the disc and trajectories being scattered in very forward directions. Here k is the momentum and R the radius of the scatterer. The semiclassical error is estimated by perturbing the Berry–Keating formula. The analysis necessitates an asymptotic analysis of very long periodic orbits. This is obtained within an approximation originally due to Baladi, Eckmann and Ruelle. We find that the average error, for sufficiently large values of kR, will exceed the mean level spacing.

1. Introduction

During the early days of the trace formula nobody really believed that it could be used to predict individual eigenvalues, at least not in the strict semiclassical limit. The experience from numerical computations is that the Gutzwiller–Voros zeta function [1] does exhibit complex zeros quite close to the (real) quantum eigenvalues, at least in the lower part of the spectrum [2–4]. The Berry–Keating formula performs even better [4, 5]. The successful numerical results in the bottom of the spectra left room for some optimism concerning the error in the strict semiclassical limit, because common estimates have suggested that the average semiclassical error, measured relative to the mean level spacing, should tend to a constant as $\overline{h} \to 0$, for systems in two degrees of freedom. For a nice review, see [6].

There are mainly two sources of errors. First, the semiclassical energy domain Green function is obtained by Laplace transforming the Van Vleck propagator [1]. However, quantum evolution follows classical evolution only for a limited time, a time that seems to be longer than first expected [7], but still limited. Computing the Laplace transform of such an object is, of course, adventurous.

Secondly, the trace formula is obtained by taking the trace of this energy domain Green function by a stationary phase technique. This is the procedure that selects out the periodic orbits. Whether or not this stationary phase approximation is justified for a particular cycle depends on $\overline{h}$, when $\overline{h}$ is *sufficiently small* it is justified.

So, the performance of the trace formula depends on the set of cycles that are required to resolve a particular state, and how well the semiclassical approximation is justified for those particular orbits. Consequently, the *semiclassical error* is not a well-defined concept. It depends on the method by which the eigenvalues are extracted. In the Berry–Keating method [8] it depends on how the semiclassical approximation works around periodic orbits up to a certain cutoff length. One thus assumes that the neglected tail *should* be able to put the

0305-4470/99/427317+28$30.00

zero on the real energy axis. In complex methods, based on *cycle expansions* [9], one includes this tail as it is. It contains long orbits, for which the semiclassical approximation obviously is nonsense. They will effect convergence of the cycle expansion and the final position of the eigenvalue. In the present paper we have, for these reasons, found it natural to use the Berry–Keating formula as the starting point for the analysis of the semiclassical error.

A semiclassical error can also be defined in other contexts. One approach is based on the boundary integral method and the semiclassical limit of the characteristic determinant [10]. Bogomolny's transfer matrix method is, in the case of billiards, closely related to this approach [11].

We will be interested in the semiclassical error for fairly realistic systems. The trace formula *is* exact for some systems, such as the Cat map [12, 13] and geodesic flow on surfaces of constant negative curvature [14]. These are uniformly hyperbolic systems and are no good as representatives for generic Hamiltonian systems.

There are two common features of chaotic systems that may cause severe problems to the stationary phase approximation. One is *intermittency*. In terms of periodic orbits, this means that periodic orbit stabilities cannot be exponentially bounded with length, as opposed to *uniformly hyperbolic* systems where they can†. This implies that the Berry–Keating sum may contain periodic orbits for which the semiclassical approximation is far from justified. If a periodic orbit has a low stability eigenvalue, it means that the stationary phase integral cannot be cut off sufficiently far from neighbouring periodic orbits. The problem can often be traced to pairs or families of periodic orbits related by bifurcations. Intermittency can be avoided in billiard systems but hardly in smooth Hamiltonian flows, so we consider it as a generic feature.

Secondly, the complex topology of generic systems make bifurcations abundant (with respect to the variation of some parameter)—generic systems are not structurally stable. There are *almost forbidden* and *almost existing* orbits all over phase space. One can imagine different strategies to deal with this problem. For billiards, it has been suggested to provide the saddle point integrals with cutoffs, which lead to diffraction effects [15]. The lack of simple symbolic coding of periodic orbits indeed proliferates diffraction effects. Uniformly hyperbolic systems with a complete symbolic dynamics such as the Baker map [16] do exhibit diffraction effects, but they are concentrated along the boundary of the generating partition.

Diffraction is not an issue for smooth potentials. But the lack of discontinuities gives us no relief. Periodic orbits close to bifurcation are dense in phase space. A periodic orbit gets stabilized close to bifurcation, which again causes problems to the trace integral. Sufficiently pruned orbits can be included as ghost orbits [17] but close to bifurcations uniform approximations have to be invoked [18, 19]. It seems to be a formidable task to include these corrections in a consistent quantization scheme.

Billiard systems are easier to analyse and in the present paper we have chosen the Sinai billiard as our model system. It is in many respects generic; it suffers from both intermittency and a lack of simple topology. Yet, it is simple enough to allow some analytical treatment.

There are two pioneering studies indicating that many of the terms included in the Berry–Keating sum are way off and that the stationary phase approximation behind them is simply not justified. Tanner [20] studied the dynamics close to the bouncing ball orbit in the stadium billiard. Primack *et al* [15] demonstrated that, for the Sinai billiard, the standard semiclassical approximation breaks down in a $\overline{h}$-dependent region associated with orbits scattering in very forward directions or sneaking very close to the disc. They coined the name *penumbra* for this region.

† Periodic orbit stabilities Λ_p can be exponentially bounded with period T_p, if all cycle Lyapunov exponents $\lambda_p \equiv \log|\Lambda_p|/T_p$ exceed some positive number. In intermittent (or nonuniformly hyperbolic) systems λ_p can be arbitrary small.

The paper [15] by Primack *et al* is the main inspiration for the present study. By estimating the size of this *penumbra* and its scaling in $\hbar$, they concluded that '*the semiclassical approximation fails for the majority of the relevant POs* (periodic orbits) *in the semiclassical limit*'. It thus seems unlikely that the Berry–Keating approach will continue to produce individual eigenvalues in the strict semiclassical limit. Such a conclusion is by no means obvious. Simple conclusions can (maybe) be drawn if the system is uniformly hyperbolic. But cycles contribute with very different weights in intermittent systems, such as the Sinai billiard. If a minority of non-diffractive cycles carried a large part of the semiclassical weight one could perhaps argue that the trace formula could continue forever to produce individual states. But unfortunately the situation is the opposite. The cycles being most prone to penumbra diffraction have large semiclassical weights. The goal of this paper is to make this reasoning more precise.

In a later study Primack *et al* [6] seem to tone down the importance of penumbra diffraction. They suggest that the semiclassical error in the semiclassical limit is of the order of the mean level spacing, or diverges very slowly. The more optimistic interpretation would allow individual states to be resolved, even in the strict semiclassical limit. The study is mainly numerical: it is an ambiguous attempt to extrapolate from finite sets of quantum states and periodic orbits into the semiclassical limit. In the present paper we find that the error will irrevocably increase beyond the mean level spacing. Our finding is consistent with the more pessimistic interpretation of [6]. The numerical results of Boasman [10] for the stadium billiard also allow a slowly growing error.

The main obstacle against exploring the strict semiclassical limit with explicitly computed periodic orbits is that they are so numerous. What is badly needed is an asymptotic theory for the distribution of long periodic orbits, a theory extending beyond standard sum rules [21, 22]. Such a theory has been suggested [23–26], based on an idea of Baladi *et al* [27].

The structure of this paper is very much like a cooking recipe. In section 2 we present all the ingredients, such as the semiclassical zeta function, the Berry–Keating formula, penumbra diffraction, some classical periodic orbit theory and the Baladi, Eckmann and Ruelle (BER) approximation. In section 3 we do the actual cooking. We consider the shift of a zero of the Berry–Keating formula if a perturbation, due to an error, is added. This simple exercise gives us the semiclassical error in terms of finite sums over pseudo-orbits. We then relate these pseudo-orbit sums to various zeta functions. These zeta functions are calculated in the BER approximation. In section 4 we present the outcome of the calculations. Then follows (section 5) a discussion about the validity of the various assumptions and approximations that underlie the results.

2. Ingredients

2.1. The semiclassical zeta function

The starting point will be the Gutzwiller–Voros zeta function [1, 28] whose zeros are to be associated with the quantum eigenvalues. It is usually represented as a product over all primitive periodic orbits p of the systems

$$Z^{sc}(E) = \prod_p \prod_{m=0}^{\infty} \left(1 - \frac{\mathrm{e}^{\mathrm{i}[S_p/\hbar - \mu_p \frac{\pi}{2}]}}{|\Lambda_p|^{1/2}\Lambda_p^m}\right) \tag{1}$$

where $S_p = S_p(E)$ is the action along p, μ_p the Maslov index and Λ_p is the expanding eigenvalue of the Jacobian.

An expansion of this infinite product usually has a larger domain of convergence. To obtain such a *cycle expansion* [29], we first expand the inner Euler product

$$Z^{sc}(E) = \prod_p \sum_{n=0}^{\infty} \frac{\Lambda_p^{-n(n-1)/2}}{\prod_{j=1}^{n}(1-\Lambda_p^{-j})} \left(-\frac{\mathrm{e}^{\mathrm{i}[S_p/\hbar - \mu_p \frac{\pi}{2}]}}{|\Lambda_p|^{1/2}} \right)^n \tag{2}$$

$$= \prod_p \sum_{n=0}^{\infty} C_{p,n} \mathrm{e}^{\mathrm{i}[nS_p - n\mu_p \frac{\pi}{2}]} \tag{3}$$

where

$$C_{p,n} = (-1)^n \frac{\Lambda_p^{-n(n-1)/2}}{|\Lambda_p|^{n/2} \prod_{j=1}^{n}(1-\Lambda_p^{-j})}. \tag{4}$$

If we now also expand the product over p, we obtain the cycle expansion in the form of a Dirichlet series. It is a sum over all pseudo-orbits, that is, all distinct combinations of periodic orbits: $\alpha = p_1^{n_{p_1}} p_2^{n_{p_2}} \dots p_k^{n_{p_k}} \dots$

$$Z^{sc}(E) = \sum_{\alpha} C_\alpha \mathrm{e}^{\mathrm{i}[S_\alpha - \mu_\alpha \frac{\pi}{2}]} \tag{5}$$

where we have defined the quantities

$$C_\alpha = \prod_p C_{p,n_p} \tag{6}$$

$$S_\alpha = \sum_p n_p S_p \tag{7}$$

$$\mu_\alpha = \sum_p n_p \mu_p. \tag{8}$$

We will restrict ourselves to billiards: the cycle action S_p is then given in terms of the geometric length $S_p = L_p \cdot k$ where $k = \sqrt{2E}$. The units are chosen such that $m = \hbar = 1$ and the semiclassical limit $\hbar \to 0$ is replaced by $k \to \infty$. In the following, we use the complex variable k rather than the energy E and the Maslov indices will be absorbed in the amplitudes C_α:

$$Z^{sc}(k) = \sum_{\alpha} C_\alpha \mathrm{e}^{\mathrm{i}L_\alpha k}. \tag{9}$$

For dispersive billiards, such as the Sinai billiard, the Maslov indices are even and the redefined amplitudes C_α will still be real.

Note that the sizes of the amplitudes are to leading order

$$C_{p,n} \sim \frac{1}{|\Lambda_p|^{n^2/2}} \tag{10}$$

and thus decay fast with n. The zeta function is not seriously affected if one restricts the n to $n \in \{0, 1\}$, at least as long as we stay on the real energy axis. This amounts to retaining only the factor $m = 0$ in (1). The resulting type of zeta function is often referred to as a *dynamical zeta function*. The higher m-factors influence the analytic structure of the zeta function far down in the complex k-plane in a quite intricate way.

2.2. The Berry–Keating formula

The spectral determinant, formally defined as $D = \prod_i (E - E_i) \sim \prod_i (k - k_i)(k + k_i)$ where $E_i = k_i^2/2$ are the energy eigenvalues, obviously obeys the functional equation

$$D(k) = D(-k). \tag{11}$$

The semiclassical analogue to the spectral determinant is

$$D^{sc}(k) = \mathrm{e}^{-\mathrm{i}\pi\bar{N}(k)} Z^{sc}(k) = \sum_{\alpha} C_\alpha \mathrm{e}^{\mathrm{i}(L_\alpha k - \pi\bar{N}(k))} \tag{12}$$

where $\bar{N}$ denote the mean spectral staircase function. It is the average of the staircase function $N(k) = \sum_i \theta(k_i - k)$ where $\theta(\cdot)$ is the unit step function. The idea of Berry and Keating [8] was to postulate that the semiclassical determinant also satisfies the functional equation (11). This is, of course, not exactly true† but by insisting on it one can convert equation (12) to a finite sum

$$D^{sc}(k) = 2 \sum_{\alpha: L_\alpha < L_{BK}} C_\alpha \cos(L_\alpha k - \pi\bar{N}(k)) \tag{13}$$

where the cutoff length L_{BK} is related to the mean spectral staircase function $\bar{N}(k)$

$$L_{BK} = \pi \frac{\mathrm{d}\bar{N}(k)}{\mathrm{d}k}. \tag{14}$$

For a billiard the function $\bar{N}(k)$ is, to leading order, given by Weyl's law: $\bar{N} = \frac{Ak^2}{4\pi}$, where A is the billiard area. So the cutoff length is given by

$$L_{BK} = \frac{Ak}{2}. \tag{15}$$

If neutral orbits are present, as they are in the Sinai billiard, their contribution can be included in $\bar{N}(k)$ which is then decorated by oscillation whose amplitude decreases with increasing k.

2.3. The classical zeta function

Another central object in our investigation will be the (weighted) evolution operator [29], whose action on a phase space distribution function $\Phi(x)$ is given by

$$\mathcal{L}^t_w \Phi(x) = \int w(x,t)\delta(x - f^t(y))\Phi(y)\,\mathrm{d}y. \tag{16}$$

The phase space point y is taken by the flow to $f^t(y)$ during time t. $w(x,t)$ is a weight associated with a trajectory starting at x and evolved during time t. It is multiplicative along the flow, that is $w(x, t_1 + t_2) = w(x, t_1)w(f^{t_1}(x), t_2)$. If $w \equiv 1$, the operator just describes the plain classical evolution of phase space densities.

In the following we will restrict ourselves to chaotic 2D billiards, and will use traversed length L as the 'time' variable.

Zeta functions are usually introduced through the trace of the evolution operator. This trace can be represented in terms of periodic orbits in two ways. First as a sum

$$\mathrm{tr}\,\mathcal{L}^L_w = \sum_p L_p \sum_{r=1}^{\infty} w_p^r \frac{\delta(L - rL_p)}{|\det(1 - M_p^r)|} \tag{17}$$

where r is the number of repetitions of primitive orbit p, having period L_p. M_p is the Jacobian of the Poincaré map, its expanding eigenvalue is Λ_p, and the cycle weight is $w_p = w(x_p, T_p)$ where $x_p \in p$ is any point on p (w_p does not depend on which of these points is chosen because of the multiplicative structure of $w(x,t)$).

† In, e.g., the Sinai billiard the semiclassical zeta function has a branch cut along the negative imaginary k-axis and the equation $D^{sc}(k) = D^{sc}(-k)$ cannot hold [30].

The trace can also be written in terms of of a *Fredholm determinant* $F_w(s)$:

$$\operatorname{tr} \mathcal{L}_w^L = \frac{1}{2\pi \mathrm{i}} \int_{a-\mathrm{i}\infty}^{a\mathrm{i}\infty} \mathrm{e}^{sL} \frac{F_w'(s)}{F_w(s)} \,\mathrm{d}s. \tag{18}$$

where $F_w(s)$ is given by

$$F_w(s) = \prod_p \prod_{m=0}^{\infty} \left(1 - w_p \frac{\mathrm{e}^{-sL_p}}{|\Lambda_p| \Lambda_p^m}\right)^{m+1}. \tag{19}$$

Again we will deal only with the $m = 0$ factors, and define instead a classical *zeta function* $Z_w(s)$ as

$$F_w(s) \approx Z_w(s) = \prod_p \left(1 - w_p \frac{\mathrm{e}^{-sL_p}}{|\Lambda_p|}\right). \tag{20}$$

The subtle difference between zeta functions and the full Fredholm determinant is far beyond the scope of this paper. The approximations we will apply are too rough to be able to tell the difference.

If the zeta function is entire, the trace can be written as a sum over zeros s_γ of the zeta function

$$\operatorname{tr} \mathcal{L}_w^L = \sum_\gamma \mathrm{e}^{s_\gamma L} \tag{21}$$

where $\mathrm{e}^{s_\gamma L}$ can be interpreted as the eigenvalues of the evolution operator. The classical zeta function $Z_w(s)$ can, in exact analogy with the semiclassical zeta function, be subject to a cycle expansion

$$Z_w(s) = \sum_\alpha a_\alpha(w) \mathrm{e}^{-sL_\alpha}. \tag{22}$$

We will eventually use the weight $w(x, L)$ to account for diffraction but for the time being it is just an arbitrary weight.

2.4. The BER approximation

We note that the equality between representations (17) and (18) also holds after smearing:

$$\operatorname{tr} \mathcal{L}_\sigma^L \equiv \sum_p L_p \sum_{r=1}^{\infty} w_p^r \frac{\delta_\sigma(L - rL_p)}{|\det(1 - M_p^r)|} = \frac{1}{2\pi \mathrm{i}} \int_{a-\mathrm{i}\infty}^{a+\mathrm{i}\infty} \mathrm{e}^{sL} \frac{F_w'(s)}{F_w(s)} \mathrm{e}^{(\mathrm{i}s\sigma)^2/2} \,\mathrm{d}s \tag{23}$$

where $\delta_\sigma(\cdot)$ are Gaussians of standard deviation σ. Gross features of the periodic sum (23) are encoded in the behaviour of the zeta function for small s and will here be estimated by means of the BER approximation [27].

In our final formulae we will need more complicated cycle sums than traces, but their gross features will still be related to the behaviour of a zeta function for small values of s.

The probabilistic approach that underlies the BER approximation has a long history: see, e.g., [31]. It was put into the context of Ruelle resonances by Baladi *et al* [27]. In [23–26] the formalism was generalized to include zeta functions with general thermodynamic weights, and the evaluation of chaotic averages.

A typical trajectory of an intermittent system fluctuates between chaotic and quasi-regular behaviour. The basic idea is to define a surface of a section such that all trajectories from the section back to itself traverse a chaotic region at least once. A chaotic burst is assumed to obliterate the memory of the preceding quasi-regular period.

The basic object in the BER approximation is the distribution of recurrence times to the surface of section: the coordinate on this surface of section will be denoted x_s. To define it, let $\Delta_s(x_s)$ be the length of such a recurring trajectory starting at x_s. Further, let $w(x_s) \equiv w(x_s, \Delta_s(x_s))$ be the weight associated with this segment. We define the (weighted) distribution of recurrence time as

$$p_w(L) = \int w(x_s)\delta(L - \Delta_s(x_s))\,\mathrm{d}\mu(x_s) \tag{24}$$

where $\mathrm{d}\mu(x_s)$ is the invariant measure of the associated Poincaré map, which we assume to be normalized: $\int \mathrm{d}\mu(x_s) = 1$. The approximate zeta function is then given in terms of the Laplace transform of this function [23–25]:

$$Z(s) \approx \hat{Z}(s) = 1 - \int p_w(L)\mathrm{e}^{-sL}\,\mathrm{d}L. \tag{25}$$

The approximation is expected to work well for small s, provided the mixing of the map is 'much faster' than that of the flow. This requirement can be realized by intermittent flows if the map is constructed according to the prescription above. The BER approach is ideal for the Sinai billiard whose disc-to-disc map is uniformly hyperbolic, whereas the flow is not†.

2.5. Penumbra diffraction

A convenient starting point for deriving the semiclassical trace formula for billiards, and to study its limitations, is the boundary integral method [10, 15, 32]. The eigenvalues of the problem are those for which the following integral equation has a solution:

$$u(\boldsymbol{r}(s)) = 2\int_S \frac{\partial G(\boldsymbol{r}(s), \boldsymbol{r}(s'))}{\partial \hat{n}_s} u(\boldsymbol{r}(s'))\,\mathrm{d}s'. \tag{26}$$

The integral is performed along the boundary of the billiard. The function $u(\boldsymbol{r}(s))$ is the normal derivative of the wavefunction

$$u(\boldsymbol{r}(s)) = \frac{\partial \Psi(\boldsymbol{r}(s))}{\partial \hat{n}_s}\ . \tag{27}$$

The boundary $\boldsymbol{r}(s)$ is parametrized by the Birkhoff coordinate s.

We can write equation (26) symbolically as a matrix equation

$$(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{U} = 0 \tag{28}$$

having a solution when $\det(\boldsymbol{I} - \boldsymbol{A}) = 0$. This determinant is rewritten as

$$\det(\boldsymbol{I} - \boldsymbol{A}) = \mathrm{e}^{\mathrm{tr}\log(\boldsymbol{I}-\boldsymbol{A})} = \mathrm{e}^{-\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{tr}(\boldsymbol{A}^n)} \tag{29}$$

where

$$\mathrm{tr}(\boldsymbol{A}^n) = 2^n \int \mathrm{d}s_1 \ldots \mathrm{d}s_n\, \frac{\partial G(\boldsymbol{r}(s_1), \boldsymbol{r}(s_2))}{\partial \hat{n}_{s_1}} \ldots \frac{\partial G(\boldsymbol{r}(s_{n-1}), \boldsymbol{r}(s_n))}{\partial \hat{n}_{s_{n-1}}} \frac{\partial G(\boldsymbol{r}(s_n), \boldsymbol{r}(s_1))}{\partial \hat{n}_{s_n}}. \tag{30}$$

There is considerable freeedom of choice of the Green function $G(\boldsymbol{r}, \boldsymbol{r}')$. In order to study the problem of penumbra diffraction Primack *et al* [15] suggested using the one-disc Green function (see below) in (26). The integral in (26) then needs to be performed only along the square boundary.

The one-disc Green function reads [33]

$$G(r_1, r_2, \Delta\theta) = \frac{\mathrm{i}}{8}\sum_{\ell=-\infty}^{\infty} (H_\ell^{(2)}(kr_1) + S_\ell(kR)H_\ell^{(1)}(kr_1))H_\ell^{(1)}(kr_2)\mathrm{e}^{\mathrm{i}\ell(\Delta\theta)} \tag{31}$$

† Stability Λ_p can be exponentially bounded with topological length n_p (number of intersections with the disc by cycle p) but not with physical length L_p.

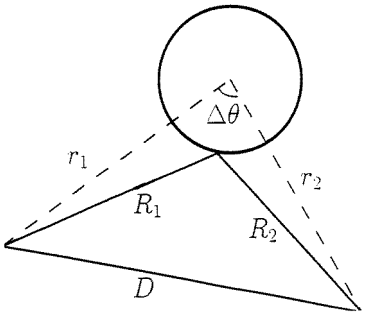

Figure 1. Some notation used in the discussion of the one-disc Green function.

where $H_\ell^{(1)}(z)$ and $H_\ell^{(2)}(z)$ are Hankel functions and r_1, r_2 and $\Delta\theta$ are explained in figure 1. The phase shift function $S_\ell(kR)$ is defined by

$$S_\ell(kR) = -\frac{H_\ell^{(2)}(kR)}{H_\ell^{(1)}(kR)}. \tag{32}$$

Using Poisson resummation we get

$$G(r_1, r_2, \Delta\theta) = \sum_{m=-\infty}^{\infty} G^{(m)}(r_1, r_2, \Delta\theta) \tag{33}$$

where

$$G^{(m)}(r_1, r_2, \Delta\theta) = \frac{\mathrm{i}}{8}\int_{-\infty}^{\infty}(H_\ell^{(2)}(kr_1) + S_\ell(kR)H_\ell^{(1)}(kr_1))H_\ell^{(1)}(kr_2)\mathrm{e}^{\mathrm{i}\ell(\Delta\theta+2\pi m)}\,\mathrm{d}\ell. \tag{34}$$

The determinant $\det(\boldsymbol{I} - \boldsymbol{A})$ reduces to the Gutzwiller–Voros zeta function if we

(1) retain only the integral approximation $G^{(m=0)}$ in (34) for the Green function, semiclassically this means that classically forbidden orbits such as creeping orbits are neglected;
(2) use the Debye approximations for the Hankel functions; and
(3) compute the integrals by stationary phase. The term $\mathrm{tr}(\boldsymbol{A}^n)$ in (30) will now contain contributions from all periodic orbits with n bounces on the square boundary.

In figure 2 we plot the circle Green function together with its semiclassical limit. Semiclassically there is a discontinuity at $d = R$ (where d is the classical impact parameter) marking the transition between the lit region and the shadow. In the exact Green function there is of course a smooth transition through what was called *the penumbra* in [15]. A destructive interference between the rays starts already in the lit region, at some critical impact parameter $d = d_{crit}$, and continues into the classical shadow. In the appendix we show that d_{crit} is given by

$$d_{crit} = (1 + \epsilon(kR))R \tag{35}$$

with $\epsilon(kR)$ (denoted ϵ_{max} in the appendix)

$$\epsilon(kR) = c \cdot (kR)^{-2/3} \tag{36}$$

where c depend only weakly (logarithmically) on kR, r_1, r_2 and R.

Within the penumbra the usually semiclassical look of the Green function is lost. The circle Green function can no longer be written as the sum of a direct and a reflected ray, leading to contributions of the form $C\mathrm{e}^{\mathrm{i}kL}$. To use the periodic orbit apparatus described in section 2.3 we need a weight that is multiplicative along the flow. However, we will not use the cycles to actually compute the spectrum. We are only interested in estimating the error induced by

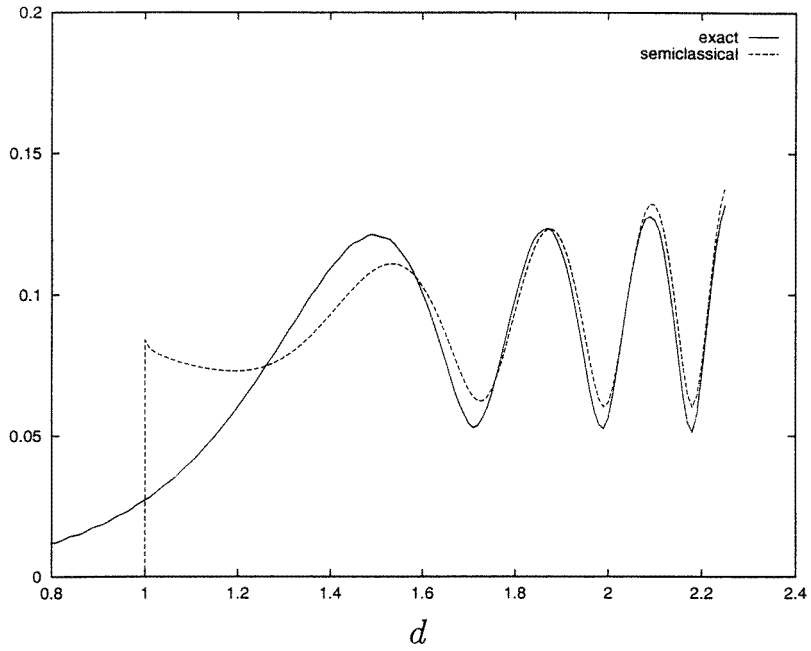

Figure 2. The circle Green function versus impact parameter d for fixed values of $k = 30$, $R = 1$ and $r_1 = r_2 = 3$. The full curve represents the exact result and the dashed curve the semiclassical approximation.

penumbra diffraction. We can then suffice with a rather crude weight, which we will choose to be multiplicative. Before making up our mind exactly how this should be achieved, let us discuss some properties of the periodic orbits in the Sinai billiard.

The cycles can be coded by associating a (coprime) lattice vector $\boldsymbol{q}$ with each disc-to-disc segment, see section 3.3. In the limit of small R any such periodic sequence $p_t = \overline{\boldsymbol{q}_1 \boldsymbol{q}_2 \dots \boldsymbol{q}_n}$ can be realized in the system, except for the rule that two consecutive lattice vectors may not be identical [24]. Suppose now that we increase the size R of the disc. Some segment of the periodic orbit, say $\boldsymbol{q}_1$, would then eventually need to go through the disc which, of course, is prohibited: the cycle is then said to be pruned. Let us say that this happens when $R = R_{bif}$. For $R < R_{bif}$, there is another cycle $p_r = \overline{\boldsymbol{q}_{1a} \boldsymbol{q}_{1b} \boldsymbol{q}_2 \dots \boldsymbol{q}_n}$, identical to p_t except that it does scatter at the the disc when p_t just passes by. When $R = R_{bif}$, p_t and p_r overlap exactly and when $R > R_{bif}$ they are both pruned. This is an analogue of the saddle-node bifurcation in smooth potential, and the only source or pruning in the Sinai billiard. For a lucid discussion of pruning in billiards, see [34].

Obviously there is a close connection between pruning and penumbra diffraction. The semiclassical weights drop suddenly to zero when the pair is pruned. We have learned from our studies of the circle Green function that the quantum pruning is more gradual.

To estimate the error we will say that the pair effectively annihilate each other if they are within the transition region discussed above. Since $|\Lambda_{p_r}| \gg |\Lambda_{p_t}|$, we will in practice only consider the removal of p_t. The pairing argument above guarantees that these uncorrected periodic orbits do not outnumber the one that are actually corrected.

So we define a cycle to be diffractive if it passes the disc within the distance ϵR (as given by (36)). The error of the semiclassical weight C_p for a diffractive cycle is thus defined as

$\delta C_p = -C_p$. A pseudo-orbit is diffractive if at least one of the participating prime cycles is diffractive. The error of the diffractive pseudo-orbit's amplitude is thus $\delta C_\alpha = -C_\alpha$.

To exclude diffractive cycles from cycle sums we introduce a multiplicative cycle weight w_p such that $w_p = 0$ if the orbit is affected by diffraction and $w_p = 1$ if it is not. Exactly how this is done is discussed in section 3. The associated pseudo-cycle weight w_α is consequently zero if one of the participating primes cycle is diffractive, otherwise it is unity.

One can also consider families of cycles with n passages of the penumbra. For each passage the trajectory can either choose to bounce off the disc or not. So such a family thus consists of 2^n members and only one of them is uncorrected according to our rule above. This is the one bouncing at every passage of the penumbra, the semiclassical weight of this one is of course very much suppressed and the neglect of this orbit is completely negligible.

The weight now depends on the parameter ϵ. We will skip the index w on the classical zeta function (20) and instead denote it $Z(s;\epsilon)$. Traces will be denoted tr $\mathcal{L}^L_\epsilon$ and coefficients of the cycle expansion (22) will be denoted $a_\alpha(\epsilon)$.

3. Preparation

3.1. Perturbation of the Berry–Keating zeros

Let k_0 be a zero of $D(k)$ as given by (13) (with superscript omitted):

$$D(k_0) = 2\sum_\alpha C_\alpha \cos(L_\alpha k - \pi \bar{N}(k)) = 0. \tag{37}$$

All pseudo-orbit sums are henceforth subject to the cutoff $L_\alpha < L_{BK}$, which will not be explicitly written out in the sums.

We are interested in how small errors in the amplitudes C_α will effect the location of this zero. We thus add a small perturbation

$$\delta D(k) = 2\sum_\alpha \delta C_\alpha \cos(L_\alpha k - \pi \bar{N}(k)) \tag{38}$$

where $\delta C_\alpha$is the error of C_α, and try to solve

$$D(k) + \delta D(k) = 0. \tag{39}$$

We then expand $k = k_0 + \delta k$ and consider the solution to

$$D'(k_0)\delta k + \delta D(k_0) + \delta D'(k_0)\delta k = 0. \tag{40}$$

We neglect the last term, it provides higher-order corrections, and get

$$\delta k = -\frac{\delta D(k_0)}{D'(k_0)}. \tag{41}$$

We will consider the perturbation of a 'typical' zero, sitting at a distance $\sim \bar{d}^{-1}$ (where $\bar{d} = \frac{\mathrm{d}\bar{N}}{\mathrm{d}k}$) from the neighbouring zeros.

We implement the restriction to a *typical* zero by relating the derivative of $D'(k_0)$ to the average amplitude of the oscillating function $D(k_0)$. If we furthermore assume that these oscillations are sine-like we can write

$$D'(k_0) = \bar{d}\sqrt{2\pi^2\langle D(k_0)^2\rangle}. \tag{42}$$

The averages $\langle\cdot\rangle$ are taken over a range Δk fulfilling $\bar{d}^{-1} \ll \Delta k \ll k_0$. We then get for the mean square of the shift

$$\langle \delta k^2\rangle^{1/2} = \bar{d}^{-1}\left(\frac{\langle \delta D(k_0)^2\rangle}{2\pi^2\langle D(k_0)^2\rangle}\right)^{1/2}. \tag{43}$$

First we focus on the denominator of equation (43). Performing the average we get

$$\langle D(k_0)^2\rangle = \tfrac{1}{2}2^2\sum_\alpha C_\alpha^2 \tag{44}$$

if we assume that cross terms cancel out, cf section 5.3. This function is essentially self-averaging, it increases in small steps as k_0 is increased. It will be related to the zeta function in the next section.

The average perturbation is, by the same arguments, given by

$$\langle \delta D^2(k_0)\rangle = 2\sum_\alpha \delta C_\alpha^2 \tag{45}$$

and

$$\langle \delta k^2\rangle^{1/2} = \left(\frac{\sum_\alpha \delta C_\alpha^2}{\sum_\alpha C_\alpha^2}\right)^{1/2} \bar{d}^{-1}/\sqrt{2\pi^2} \equiv F(k_0)\bar{d}^{-1}/\sqrt{2\pi^2}. \tag{46}$$

Following our reasoning in section 2.5 we put

$$\delta C_\alpha^2 = (1-w_\alpha)C_\alpha^2. \tag{47}$$

This means that the correction is $\delta C_\alpha^2 = C_\alpha^2$ if one of the participating prime cycles in α is diffractive, otherwise the correction is zero $\delta C_\alpha^2 = 0$. We can rewrite the function F as

$$F(k_0) = \left(1 - \frac{\sum_\alpha w_\alpha C_\alpha^2}{\sum_\alpha C_\alpha^2}\right)^{1/2}. \tag{48}$$

Recall that we are only considering pseudo-orbits such that $n_p \leqslant 1$ so

$$|C_{\alpha=p_1p_2\ldots p_k}|^2 = |\Lambda_\alpha|^{-1} \equiv |\Lambda_{p_1}\cdot\Lambda_{p_1}\ldots\Lambda_{p_k}|^{-1} = a_\alpha(\epsilon=0) \tag{49}$$

which we have identified with the coefficients of the cycle expansion (22) (please recall the notational conventions introduced at the end of section 2.5). We can now rewrite the numerator and denominator of (48)

$$\sum_\alpha w_\alpha(\epsilon)C_\alpha^2 = \sum_\alpha |a_\alpha(\epsilon)| \tag{50}$$

$$\sum_\alpha C_\alpha^2 = \sum_\alpha |a_\alpha(0)|. \tag{51}$$

Obviously, the function F is bounded by $0 \leqslant F \leqslant 1$. The perturbative approach taken in this section is valid only if the predicted value of F is small. A semiclassical error for a chaotic system (hence lacking quantum numbers) can only be well defined as long as it is much smaller than the mean spacing, so the restriction $F < 1$, resulting from the perturbative approach, can hardly be considered as a drawback of the perturbative method. If F is approaches unity there is no other sensible interpretation than a failure of the Berry–Keating formula to resolve individual states.

3.2. Treating the pseudo-orbit sums

The goal of this section is to relate the pseudo-orbit sums (50) and (51) to various zeta functions. There is an important distinction between these pseudo-orbit sums and the cycle expansion (22)—the occurrence of the absolute values in equations (50) and (51).

We start with the basic classical zeta function, given by equation (20):

$$Z(s;\epsilon) = \prod_p \left(1 - w_p(\epsilon)\frac{\mathrm{e}^{-sL_p}}{|\Lambda_p|}\right). \tag{52}$$

Consider now the cycle expansion of a slightly different zeta function (note the plus sign!)

$$Z^{+}(s;\epsilon)=\prod_{p}\left(1+w_{p}(\epsilon)\frac{\mathrm{e}^{-sL_{p}}}{|\Lambda_{p}|}\right)=\sum_{\alpha}b_{\alpha}(\epsilon)\mathrm{e}^{-sL_{\alpha}}. \tag{53}$$

All b_α are positive; in fact, they are the summands of equations (50) and (50)

$$b_{\alpha}(\epsilon)=|a_{\alpha}(\epsilon)|. \tag{54}$$

We rewrite $Z^{+}(s;\epsilon)$ as

$$Z^{+}(s;\epsilon)=\int_{0}^{\infty}b(L;\epsilon)\mathrm{e}^{-sL}\,\mathrm{d}L \tag{55}$$

where

$$b(L;\epsilon)=\sum_{\alpha}b_{\alpha}\delta(L-L_{\alpha}). \tag{56}$$

Conversely, the function $b(L;\epsilon)$ is related to the zeta function $Z^{+}(s;\epsilon)$ by means of an inverse Laplace transform

$$b(L;\epsilon)=\frac{1}{2\pi\mathrm{i}}\int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty}Z^{+}(s;\epsilon)\mathrm{e}^{sL}\,\mathrm{d}s. \tag{57}$$

The semiclassical error (48) can now be written in terms of the function $b(L;\epsilon)$

$$F=\sqrt{1-\frac{\int_{0}^{L_{BK}}b(L;\epsilon)\mathrm{d}L}{\int_{0}^{L_{BK}}b(L;0)\mathrm{d}L}} \tag{58}$$

where L_{BK} is given by equation (14) and ϵ is given by equation (36).

If we now define yet another zeta function

$$Z_{2}(s;\epsilon)=\prod_{p}\left(1-w_{p}^{2}\frac{\mathrm{e}^{-2sL_{p}}}{|\Lambda_{p}|^{2}}\right) \tag{59}$$

we can rewrite $Z^{+}(s;\epsilon)$ as

$$Z^{+}(s;\epsilon)=\frac{Z_{2}(s;\epsilon)}{Z(s;\epsilon)} \tag{60}$$

cf [35]. If the involved zeta functions are entire we can write

$$b(L;\epsilon)=\sum_{\gamma}\mathrm{res}\frac{Z_{2}(s;\epsilon)}{Z(s;\epsilon)}\bigg|_{s=s_{\gamma}}\mathrm{e}^{s_{\gamma}L} \tag{61}$$

which should be compared with (21). We will see that this situation will arise if $\epsilon\neq 0$. The asymptotic behaviour of $b(L;\epsilon)$ is (under much milder assumptions on the analytic properties) related to the leading zero $s_0(\epsilon)$

$$b(L;\epsilon)\sim\mathrm{res}\frac{Z_{2}(s;\epsilon)}{Z(s;\epsilon)}\bigg|_{s=s_{0}}\mathrm{e}^{s_{0}L}=\frac{Z_{2}(s_{0}(\epsilon);\epsilon)}{Z'(s_{0}(\epsilon);\epsilon)}\mathrm{e}^{s_{0}L}. \tag{62}$$

This leading asymptotic behaviour is all we need to evaluate the the asymptotics of the integrals in equation (58). Moreover, as discussed in section 2.4, the behaviour of zeta functions close to the origin is insensitive to fine details in the spectrum of periodic orbits. The large-scale structure of periodic orbits will be studied in the next section.

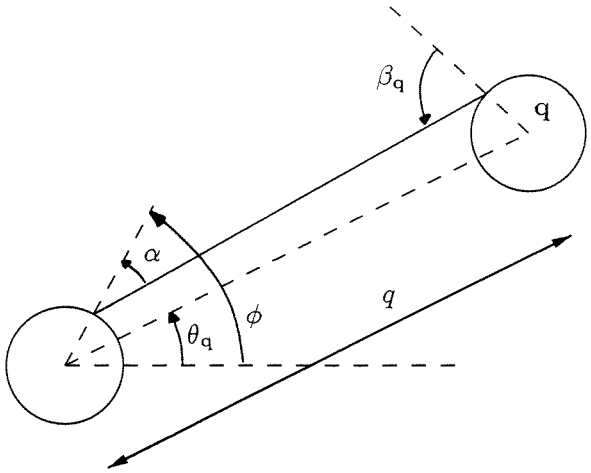

Figure 3. Definition of various angles associated with scattering to disc $\boldsymbol{q}$.

3.3. Implementing the BER approximation for the Sinai billiard

The BER approximation outlined in section 2.4 is very well suited for the Sinai billiard, in particular if the scatterer is small. The obvious choice of the surface of the section is provided by the disc itself [24].

To compute the distribution of recurrence times (24) it is convenient to consider the unfolded representation of the Sinai billiard. A trajectory segment from the disc to itself can thus be considered as going from one disc, associated with lattice vector $(0, 0)$ to some other disc represented by lattice vector $\boldsymbol{q}$. All segments going to $\boldsymbol{q}$ have essentially the same length† $q = |\boldsymbol{q}|$ and equation (24) simplifies to

$$p(L;\epsilon) = \sum_{\boldsymbol{q}} \hat{a}_{\boldsymbol{q}}(\epsilon)\delta(L-q) \tag{63}$$

where

$$\hat{a}_{\boldsymbol{q}}(\epsilon) = \int_{\Omega_q} w(x_s)\,\mathrm{d}\mu(x_s). \tag{64}$$

Ω_q is the set of initial points x_s whose target is $\boldsymbol{q}$. The natural phase space coordinates for the disc-to-disc map are Birkhoff coordinates $x_s = (R\phi, \sin\alpha)$, see figure 3 for definition of the angles ϕ and α. The invariant density is simply uniform, the measure is $\mathrm{d}\mu(x_s) \sim \mathrm{d}x_s = \mathrm{d}(R\Phi)\,\mathrm{d}(\sin\alpha)$. The explicit knowledge of the measure enables us to compute the quantities $\hat{a}_{\boldsymbol{q}}$

$$\hat{a}_{\boldsymbol{q}}(\epsilon) = \int_{\Omega_q} w(x_s)\,\mathrm{d}x_s. \tag{65}$$

The approximate zeta function (25) is then

$$\hat{Z}(s;\epsilon) = 1 - \int_0^\infty p(L;\epsilon)\mathrm{e}^{-sL}\,\mathrm{d}s = 1 - \sum_{\boldsymbol{q}} \hat{a}_{\boldsymbol{q}}(\epsilon)\mathrm{e}^{-sq}. \tag{66}$$

The weight, as defined in section 2.5 will be zero, $w(x_s) = 0$, if the trajectory starting at x_s and heading for disc $\boldsymbol{q}$ passes some other disc within a distance ϵR before actually hitting

† The variation is of order R which will be completely neglible in what follows.

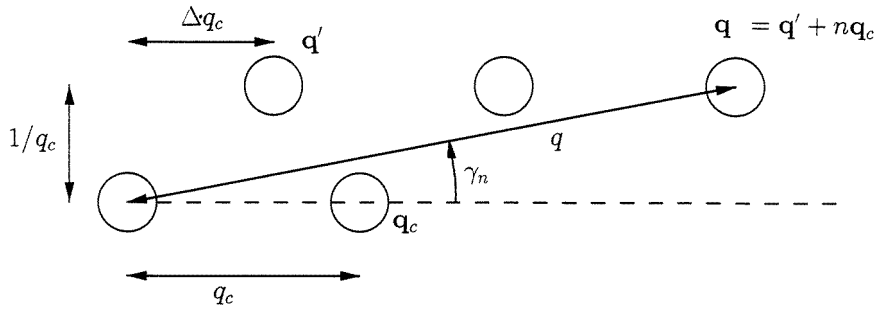

Figure 4. Scattering to disc $\boldsymbol{q} = \boldsymbol{q}' + n\boldsymbol{q}_c$ with $n = 2$ in the corridor $\boldsymbol{q}_c$.

$\boldsymbol{q}$, otherwise it is equal one, $w(x_s) = 1$. If $\epsilon = 0$, then $\hat{a}_q$ is the total fraction of the phase space area corresponding to disc $\boldsymbol{q}$.

The computation of $p(L;\epsilon)$ will unfortunately be long and boring. It is possible for the reader to jump directly to the results (85)–(88) and go on from there without losing too much of the context.

The calculation will run in parallel with [36] in large parts and we will frequently refer to that paper. All error estimates in [36] carry over directly, so we will simply omit them below to make things a little more transparent. Nevertheless, we will without hesitation display equations as strict equalities, but the reader should bear in mind that all expressions are valid in the small R limit.

Keeping track of the accessible discs. It is easy to see that only discs represented by coprime lattice vectors are accessible. In [36] we showed that any coprime lattice vector $\boldsymbol{q}$ can uniquely be written in the form $\boldsymbol{q} = \boldsymbol{q}' + n\boldsymbol{q}_c$, where $n \geqslant 2$ and $\boldsymbol{q}'$ represents the lattice points closest to the line from $(0, 0)$ to $\boldsymbol{q}_c$. We say that disc $\boldsymbol{q}$ lies in the $\boldsymbol{q}_c$ *corridor*, see figure 4.

Actually, there are two such neighbouring lattice points for each corridor vector $\boldsymbol{q}_c$, one with smaller and one with larger polar angle. Below we assume that $\boldsymbol{q}'$ is the one with the larger polar angle, the other case is completely analogous, and is accounted for by multiplying by a factor of two on some strategic occasions, see below.

Computing $\hat{a}_q$. Consider now a trajectory segment hitting disc $\boldsymbol{q}$. The relation between the phase space variables ϕ and α, see figure 3, and the scattering angle β_q on disc $\boldsymbol{q}$ is given by

$$R_q \sin\beta_q - R\sin\alpha = q\sin(\phi - \theta_q - \alpha) \tag{67}$$

where θ_q is the polar angle of the lattice vector $\boldsymbol{q}$, and R_q is the radius of the target disc. The argument of $\sin(\phi - \theta_q - \alpha)$ is small when $\boldsymbol{q}$ is large so we can expand the sine to first order and get

$$R_q \sin\beta_q - R\sin\alpha = q(\phi - \theta_q - \alpha). \tag{68}$$

The disc under observation, $\boldsymbol{q}$, is potentially shadowed only by two discs, namely $\boldsymbol{q} - \boldsymbol{q}_c$ and $\boldsymbol{q}_c$, see figure 4. We implement the weight $w(x_s)$ by simply inflating these two discs from radius R to $(1 + \epsilon)R$.

To see how Ω_{q_c} shadows Ω_q we replace $\boldsymbol{q}$ in (68) by $\boldsymbol{q}_c$ and put

$$\begin{aligned} \beta_{q_c} &= \pi/2 \\ R_{q_c} &= (1+\epsilon)R \\ \theta_{q_c} &= \theta_q - 1/(qq_c). \end{aligned} \tag{69}$$

This gives an equation (in terms of phase space variables α and ϕ) for the relevant boundary of Ω_{q_c}

$$(1+\epsilon)R - R\sin\alpha = q_c\left(\phi - \theta_q + \frac{1}{q\ q_c} - \alpha\right). \tag{70}$$

We can now treat Ω_{q-q_c} in the same way. We replace $\boldsymbol{q}$ in (68) by $\boldsymbol{q}-\boldsymbol{q}_c$ and put

$$\begin{aligned}\beta_{q-q_c} &= -\pi/2\\ R_{q-q_c} &= (1+\epsilon)R\\ \theta_{q-q_c} &= \theta_q + 1/(q\ q_c)\end{aligned} \tag{71}$$

and get the equation for the borderline of Ω_{q-q_c}

$$-(1+\epsilon)R - R\sin\alpha = (q-q_c)\left(\phi - \theta_q - \frac{1}{q\ q_c} - \alpha\right). \tag{72}$$

Next we want to change variables from (α,ϕ) to $(\sin\alpha, \sin\beta_q))$: the relation is given by equation (68).

By combining equations (68) and (70) we see that the shadowing of Ω_{q_c} corresponds to the straight line

$$q_c\sin\beta_q + (q_n - q_c)\sin\alpha = q(1+\epsilon) - \frac{1}{R}. \tag{73}$$

Similarly, by combining equations (68) and (72) we see that the shadowing of Ω_{q-q_c} is given by

$$(q-q_c)\sin\beta_q + q_c\sin\alpha = q(1+\epsilon) - \frac{1}{R}. \tag{74}$$

If $\boldsymbol{q}$ were not shadowed, Ω_q would be given by $-1 < \sin\beta_q < 1$ and $-1 < \sin\alpha < 1$. So the integral (65) is simply the area of the remainder of this square, lying inside the lines given by equations (73) and (74).

The integration element $\mathrm{d}x_s$ in (65) over Ω_q is (to leading order)

$$\mathrm{d}x_s = \frac{R}{4\pi q}\,\mathrm{d}(\sin\alpha)\,\mathrm{d}(\sin\beta_q). \tag{75}$$

It is normalized in such a way that the integral over one octant of the plane is unity.

We arrive at the following results:

$$\hat{a}_q(\epsilon) = \frac{R}{\pi q}\begin{cases} 1 & 0 < q < \dfrac{1}{2R}\dfrac{1}{1+\epsilon/2}\\ 1 - \dfrac{(1/2R - q - \epsilon q/2)^2}{q_c(q-q_c)} & \dfrac{1}{2R}\dfrac{1}{1+\epsilon/} < q < \left(\dfrac{1}{2R}+q_c\right)\dfrac{1}{1+\epsilon/2}\\ \dfrac{(1/2R - q_c - \epsilon q/2)^2}{(q-q_c)(q-2q_c)} & \left(\dfrac{1}{2R}+q_c\right)\dfrac{1}{1+\epsilon/2} < q < \dfrac{2}{\epsilon}\left(\dfrac{1}{2R}-q_c\right)\\ 0 & \dfrac{2}{\epsilon}\left(\dfrac{1}{2R}-q_c\right) < q. \end{cases} \tag{76}$$

The result can be described in the following way. All discs inside a radius $q < \frac{1}{2R}\frac{1}{1+\epsilon/2}$ are unshadowed and the corresponding trajectory segments not diffractive. Outside this horizon the accessible discs are aligned along corridors, each corridor characterized by the vector $\boldsymbol{q}_c$, subject to the condition $q_c < \frac{1}{2R}\frac{1}{1+\epsilon/2}$. Segments longer than $\frac{2}{\epsilon}(\frac{1}{2R}-q_c)$ in a particular corridor $\boldsymbol{q}_c$ are always affected by diffraction. Segments longer than $\frac{2}{\epsilon}\frac{1}{2R}$ are always diffractive.

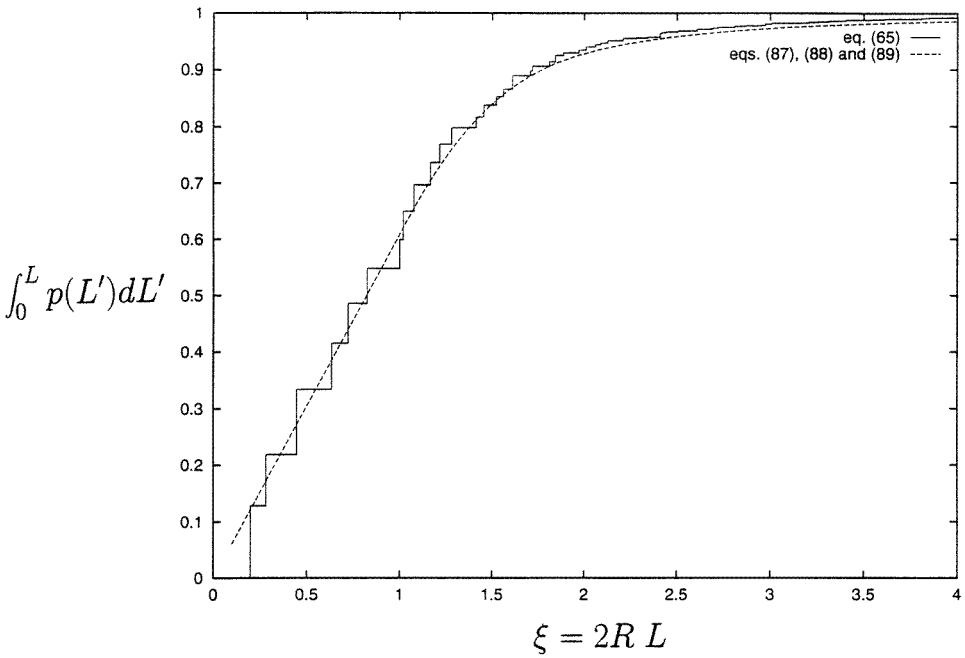

Figure 5. The integral of $p(L(\xi))$ versus ξ. The radius is $R = 0.1$ and $\epsilon = 0$. The (full) staircase curve is obtained from equation (63) with amplitudes $\hat{a}_q(\epsilon)$ given by (76). The (dashed) smooth curve is obtained from equations (86)–(88).

This means that there is a one-to-one correspondence between corridors (beyond the horizon) and accessible discs inside the horizon, so in order to perform the sum (66) we just need to know the finite number of coprime lattice points inside the horizon, exactly how the sum should administrated we be clear below.

We can now obtain an approximate zeta function by plugging these $\hat{a}_q$ into (66). The integrated recurrence time distribution $\int_0^L p(L')\,\mathrm{d}L' = \sum_{q:|q|<L} \hat{a}_q$ is plotted in figures 5 and 6 for the radius $R = 0.1$.

Statistical treatment of the coprime lattice vectors. However, if R is small, there is a vast number of coprime lattice points inside the horizon and, according to number theory, they tend to be uniformly distributed over the plane. As a matter of fact, one can use this uniformity to turn the sum into an integral and write down an explicit formula for the distribution of recurrence times (63).

Asymptotically, there are $6L^2/\pi$ coprime lattice points $\boldsymbol{q}$ such that $|\boldsymbol{q}| < L$ in the first octant so that the mean density of coprime lattice points is, in an asymptotic sense,

$$d_c(L) = \frac{12L}{\pi}. \tag{77}$$

(i) First we consider the case $L < \frac{1}{2R}\frac{1}{1+\epsilon/2}$.
Here $\hat{a}_q$ is a function of q only, cf equation (76). The distribution function $p(L;\epsilon)$ becomes

$$p(L;\epsilon) = \sum_q \hat{a}_q \delta_\sigma(L-q) = \sum_q \frac{R}{\pi q}\delta_\sigma(L-q)$$

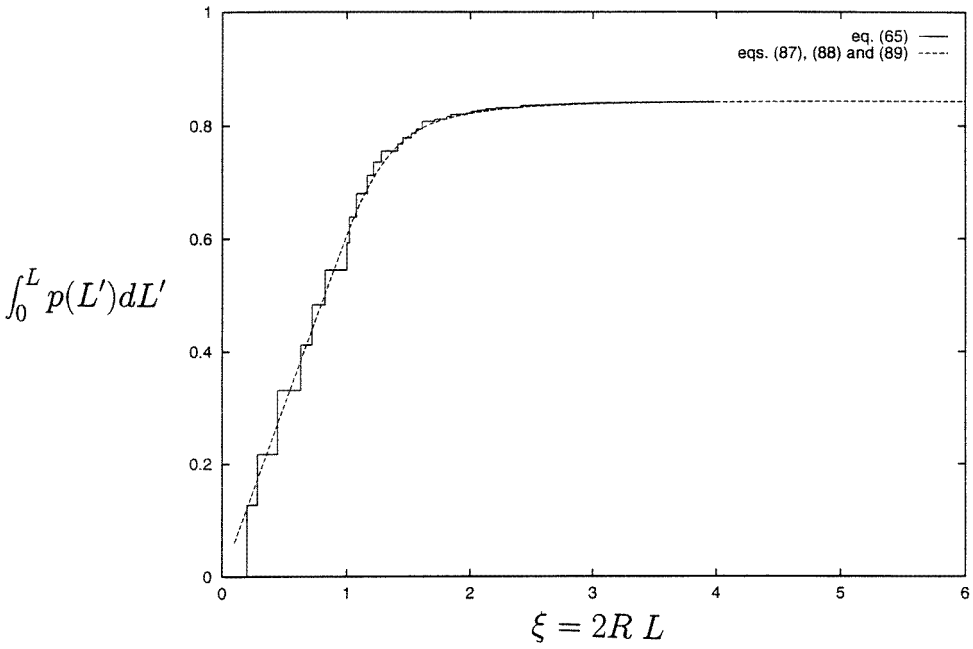

Figure 6. Same as figure 5 but with $\epsilon = 0.2$.

$$= d_c(L)\frac{R}{\pi L} = \frac{12R}{\pi^2}. \tag{78}$$

We introduce the rescaled length $\xi = 2RL$ and express the result as

$$p(L(\xi);\epsilon) = \frac{12R}{\pi^2} \qquad \xi < \frac{1}{1+\epsilon/2}. \tag{79}$$

(ii) Next we consider the *transition region* $1/2R < L < 1/R$.
According to equation (76) the amplitudes $\hat{a}_q$ depend on the length of corridor vector $q_c = |\boldsymbol{q}_c|$ and the length of $\boldsymbol{q}$: $\hat{a}_q = \hat{a}(q_c, q)$. The length q is approximately $q \approx q' + nq_c \equiv \Delta q_c + nq_c$, where Δ is a number such that $0 < \Delta < 1$, see figure 4. We get

$$\begin{aligned} p(L;\epsilon) &= 2\sum_{\boldsymbol{q}_c}\sum_{n=2}^{\infty}\hat{a}(q_c, q)\delta_\sigma(L-q) \\ &= 2\sum_{\boldsymbol{q}_c}\sum_{n=2}^{\infty}\hat{a}(q_c, q)\delta_\sigma(L - q_c n - \Delta q_c) \end{aligned} \tag{80}$$

where we inserted the factor 2 to account for both neighbours of $\boldsymbol{q}_c$.

We will now turn the sums over $\boldsymbol{q}_c$ and n into an integral over the density of coprime lattice vectors. The parameter Δ is uniformly distributed over the interval $0 < \Delta < 1$ [24], this means that we can just integrate n from $n = 2$ to ∞:

$$\begin{aligned} p(L;\epsilon) &= 2\int_0^{1/2R-\epsilon L/2} \mathrm{d}q_c\, d_c(q_c)\hat{a}(q_c, L)\int_{n=2}^{\infty}\delta_\sigma(L - q_c n) \\ &= 2\int_0^{1/2R-\epsilon L/2} \mathrm{d}q_c\, d_c(q_c)\hat{a}(q_c, L)\frac{1}{q_c}\theta_\sigma(L - 2q_c) \\ &= 2\int_0^{\min(1/2R-\epsilon L/2, L/2)} \mathrm{d}q_c\, d_c(q_c)\hat{a}(q_c, L)\frac{1}{q_c} \end{aligned} \tag{81}$$

where $\theta(x)$ is the unit step function.

Since we are considering the region $\frac{1}{2R}\frac{1}{1+\epsilon/2} < q < \frac{1}{R}\frac{1}{1+\epsilon}$, we have $\min(1/2R - \epsilon L/2, L/2) = L/2$. Next we insert the expression for $\hat{a}(q_c, q)$ from equation (76) and $d_c(q_c)$ from equation (77)

$$p(L;\epsilon) = 2\int_{l+\epsilon L/2-1/2R}^{l/2} \mathrm{d}q_c \, \frac{12q_c}{\pi}\frac{1}{q_c}\frac{R}{\pi L}\left(1 - \frac{(1/2R - L - \epsilon L/2)^2}{q_c(L-q_c)}\right) + 2\int_0^{L-\epsilon L/2-1/2R} \mathrm{d}q_c \, \frac{12q_c}{\pi}\frac{1}{q_c}\frac{R}{\pi L}\frac{(1/2R - q_c - \epsilon L/2)^2}{(L-q_c)(l-2q_c)}. \tag{82}$$

We change the integration variable to $\eta = 2Rq_c$ and use, as before, $\xi = 2RL$. This leaves us with the following integral to solve:

$$p(L(\xi);\epsilon) = \frac{24R}{\pi^2}\left(\int_{\xi+\epsilon\xi/2-1}^{\xi/2} \frac{\mathrm{d}\eta}{\xi}\left(1 - \frac{(1-\xi-\epsilon\xi/2)^2}{\eta(\xi-\eta)}\right) + \int_0^{\xi+\epsilon\xi/2-1} \frac{\mathrm{d}\eta}{\xi}\frac{(1-\eta-\epsilon\xi/2)^2}{(\xi-\eta)(\xi-2\eta)}\right). \tag{83}$$

The result of this integral is displayed below (87).

(iii) There now remains only the case $\frac{1}{R}\frac{1}{1+\epsilon} < L < \frac{2}{\epsilon}$.

The calculation is completely analogous to the previous case and we get

$$\begin{aligned} p(L(\xi);\epsilon) &= 2\int_0^{1/2R-\epsilon L/2} \mathrm{d}q_c \, d_c(q_c)\frac{1}{q_c}a(q_c, q = L) \\ &= 2\int_0^{1/2R-\epsilon L/2} \mathrm{d}q_c \, \frac{12q_c}{\pi}\frac{1}{q_c}\frac{R}{\pi L}\frac{(1/2R - q_c - \epsilon L/2)^2}{(L-q_c)(L-2q_c)} \\ &= \frac{24R}{\pi^2}\int_0^{1-\epsilon\xi/2} \frac{\mathrm{d}\eta}{\xi}\frac{(1-\eta-\epsilon\xi/2)^2}{(\xi-\eta)(\xi-2\eta)}. \end{aligned} \tag{84}$$

Results. It is time to summarize our results so far. It is natural to display the final results in terms of the distribution of the rescaled recurrence lengths $\xi = 2RL$ rather than L. We call the distribution $f(\xi;\epsilon)$ and it is trivially related to $p(L;\epsilon)$ according to

$$p(L;\epsilon)\,\mathrm{d}L = f(\xi;\epsilon)\,\mathrm{d}\xi. \tag{85}$$

Inside the horizon we already have

$$f(\xi;\epsilon) = \frac{6}{\pi^2} \qquad \xi < \frac{1}{1+\epsilon/2}. \tag{86}$$

Beyond the horizon we get, after having performed the integrals (83) and (84)†

$$\begin{aligned} f(\xi;\epsilon) = \frac{3}{\pi^2\xi^2}&(2\xi - \epsilon\xi^2 + [4\xi - 3\xi^2 - 2\epsilon\xi^2]\log\xi \\ &+[4\xi^2 - 8\xi^2 + 4 - 4\epsilon\xi^2 - \epsilon^2\xi^2 - 4\epsilon\xi]\log(\xi(1+\epsilon/2) - 1) \\ &+[4\xi - \xi^2 - 4 - 2\epsilon\xi^2 - \epsilon^2\xi^2 + 4\epsilon\xi]\log|\xi(1+\epsilon) - 2| \\ &\frac{1}{1+\epsilon/2} < \xi < 2/\epsilon. \end{aligned} \tag{87}$$

This is valid up to the point $\xi = 2/\epsilon$ where the function chokes. After that we have

$$f(\xi;\epsilon) = 0 \qquad 2/\epsilon < \xi. \tag{88}$$

† The results of these two integrals can be summarized in one formula. Note the absolute value in one of the logarithms!

The integral $\int_0^\xi f(\xi';\epsilon)\,\mathrm{d}\xi'$ of these expressions is plotted in figures 5 and 6. We note that the statistical treatment of the lattice vectors works surprisingly well, even for such a 'large' radius as $R = 0.1$.

Zeros and traces. It is also natural to let the zeta function (66) depend on a rescaled variable z as defined by

$$s = 2Rz. \tag{89}$$

A power series expansion of the the zeta function is related to the moments of the distribution $f(\xi;\epsilon)$

$$\hat{Z}(z;\epsilon) = 1 - \int \mathrm{e}^{-z\xi} f(\xi;\epsilon)\,\mathrm{d}\xi = 1 - \sum_{m=0}^{\infty} \frac{(-z)^m}{m!} \int \xi^m f(\xi;\epsilon)\,\mathrm{d}\xi. \tag{90}$$

These moments can be computed from equations (86)–(88) and are found to be

$$\int \xi^m f(\xi;\epsilon)\,\mathrm{d}\xi = \begin{cases} 1 - \epsilon + \mathrm{O}(\epsilon^2 \log\epsilon) & m = 0 \\ (1 + \mathrm{O}(\epsilon\log\epsilon)) & m = 1 \\ (\mathrm{O}(\log\epsilon)) & m = 2 \\ \mathrm{O}(1/\epsilon^{m-2}) & m \geqslant 3. \end{cases} \tag{91}$$

The leading zero z_0 of $\hat{Z}(z;\epsilon)$, occurring in equation (62), can be computed from equations (90) and (91), and is found to be

$$z_0 = -\epsilon + \mathrm{O}(\epsilon^2 \log\epsilon). \tag{92}$$

We also need the derivative of the zeta function evaluated at this zero, cf equation (62),

$$\hat{Z}'(z_0) = 1 + \mathrm{O}(\epsilon\log\epsilon). \tag{93}$$

In figures 7 and 8 we compare these asymptotic formulae with results from numerical computation of the BER zeta function as obtained from equations (86)–(88).

Traces will not be a direct concern to us. But, by computing traces in the BER framework, we can get an idea of the influence of non-leading zeros by computing the trace in the BER approximation. That information is relevant for the next section. The computation is done numerically by the fast Fourier transform (FFT) technique.

The trace formula is in our rescaled units

$$\mathrm{tr}\,\mathcal{L}_\epsilon^\xi = \frac{1}{2\pi\mathrm{i}} \int_{a-\mathrm{i}\infty}^{a+\mathrm{i}\infty} \mathrm{e}^{z\xi} \left(\frac{\mathrm{d}}{\mathrm{d}z} \log \hat{Z}_w(z) \right) \mathrm{d}z. \tag{94}$$

Traces for $\epsilon = 0$ and $\epsilon = 0.1$ are plotted in figures 9 and 10. For small ξ the trace (in the BER approximation) is given by

$$\mathrm{tr}\,\mathcal{L}_\epsilon^\xi = \exp(6\xi/\pi^2) - 1 \qquad \xi < \frac{1}{1+\epsilon/2}. \tag{95}$$

If the reader wants to verify this, the following hint should be useful: the trace $\mathrm{tr}\,\mathcal{L}_\epsilon^\xi$ in the range $\xi < \Xi$ depends only on the behaviour of the distribution $f(\xi)$ in the same range ($\xi < \Xi$) where f is constant.

In the large ξ limit, the asymptotic behaviour is given by the leading zero z_0, and thus

$$\mathrm{tr}\,\mathcal{L}_\epsilon^\xi \sim \exp(z_0\xi) \sim \exp(-\epsilon\xi) \tag{96}$$

see figure 10. We observe that this asymptotic result settles down very early, that is long before the natural scale $2/\epsilon$, cf equation (88).

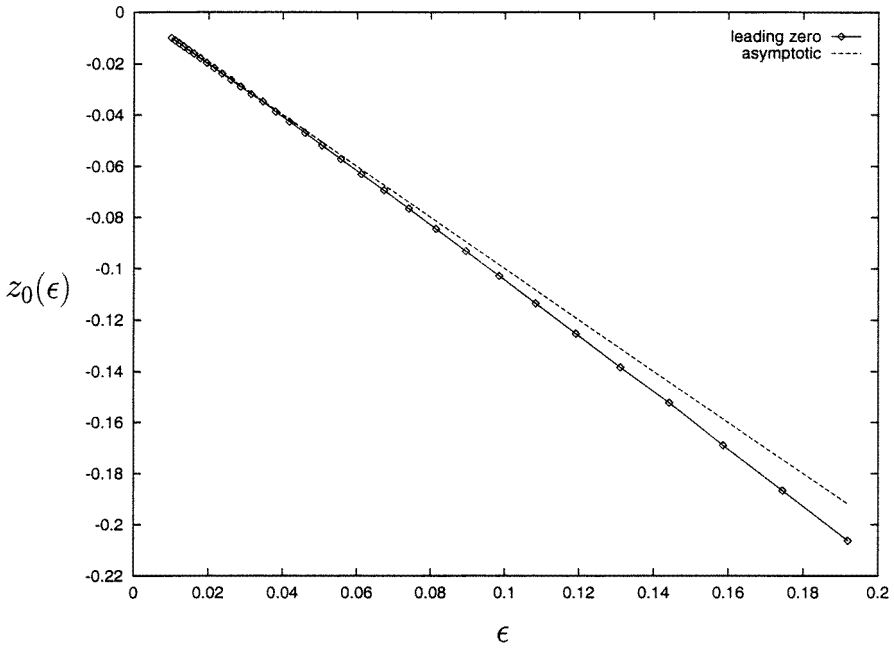

Figure 7. The leading zero $z_0(\epsilon)$ versus ϵ compared with the asymptotic formula (92).

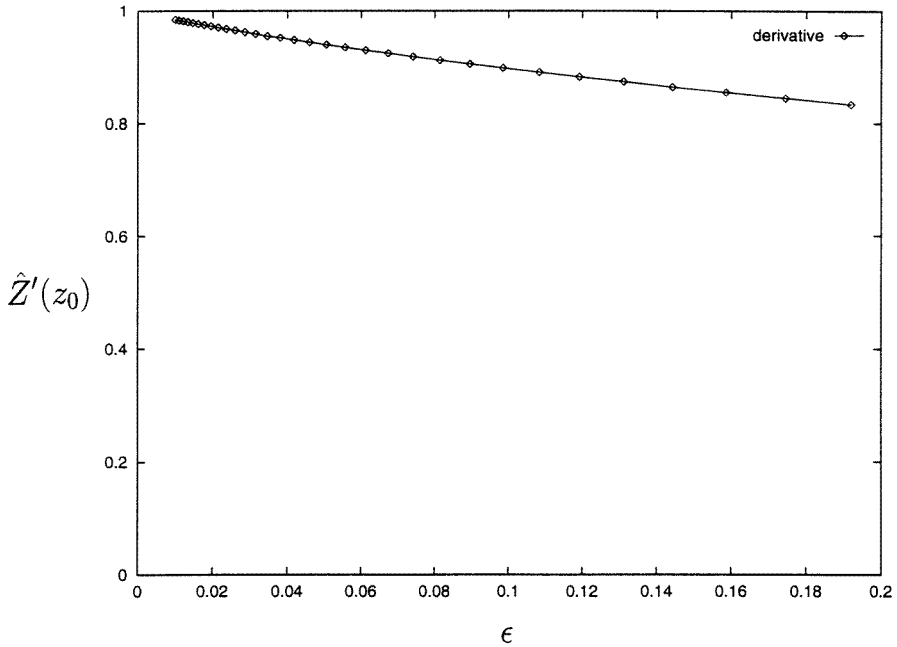

Figure 8. The derivative $\hat{Z}'(z_0)$ evaluated at the leading zero versus ϵ. It approaches unity as $\epsilon \to 0$ as predicted by equation (93).

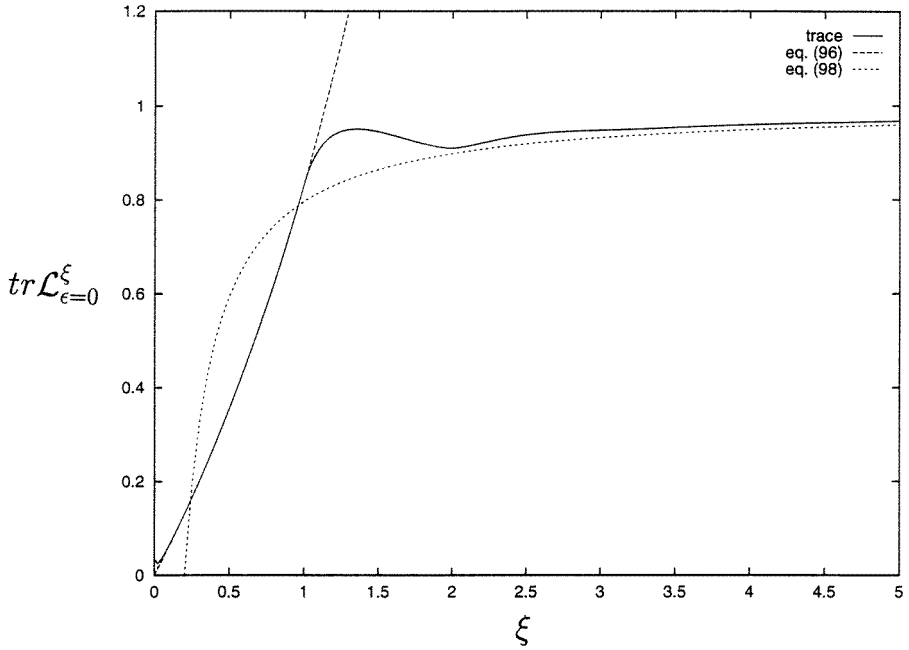

Figure 9. The trace (94) for $\epsilon = 0$ (full curve). Comparison is made with equation (95) for small values of ξ (dashed curve) and equation (97) for large values of ξ (dotted curve).

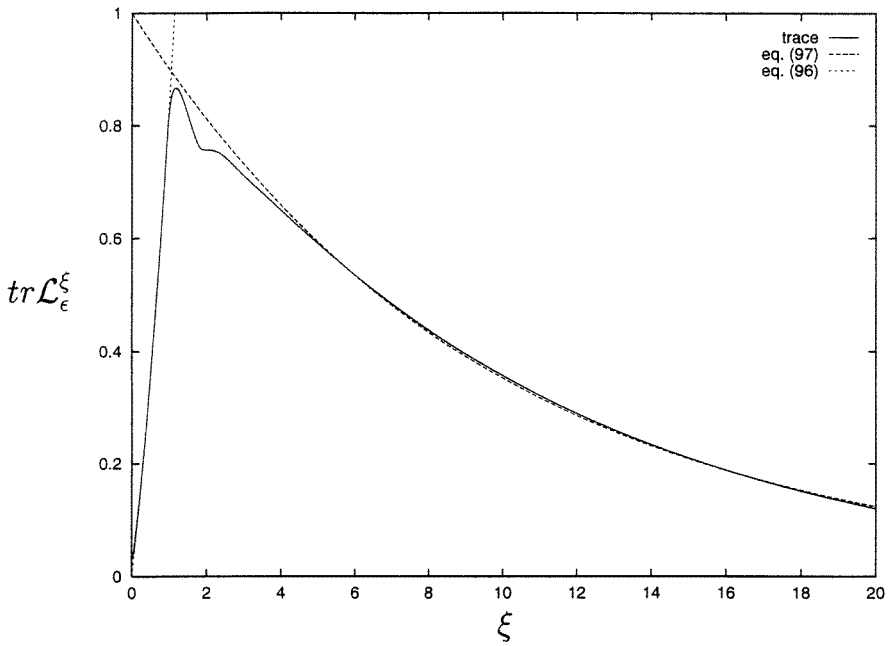

Figure 10. The trace (94) for $\epsilon = 0.1$. Comparison is made with equation (95) for small values of ξ (dotted curve) and equation (96) for large values of ξ (dashed curve).

For any finite ϵ, the zeta function is entire since $f(\xi;\epsilon)$ has compact support. But when $\epsilon \to 0$, zeros will accumulate along the negative real z-axis, building up a branch cut. For the limiting case $\epsilon = 0$, this will lead to a power law correction [24, 26]

$$\operatorname{tr} \mathcal{L}^{\xi} \sim 1 - \frac{2}{\pi^2 \xi} \tag{97}$$

see figure 9. This power law will not be essential in the following.

We also need to know the value of $Z_2(z_0;\epsilon)$ as defined in (59). This zeta function contains the square of $|\Lambda_p|$ in the denominator. This means that at $z = z_0$ (which is close to the origin for small ϵ) the value of the zeta function will be dominated by the shortest cycles, which, for small ϵ, will be non-diffractive. This implies that $Z_2(z_0;\epsilon)$ tends to a constant faster than $\hat{Z}'(z_0)$ do, as $\epsilon \to 0$. A simple estimate, based on the methods in [26] suggests that

$$Z_2(z_0,\epsilon) = 1 + \mathrm{O}(R^2 \log R) + \mathrm{O}(R^2)\mathrm{O}(\epsilon) \tag{98}$$

which should be compared with equation (93).

4. Results

We now possess all the tools we need to finally be able to compute the asymptotic limit of the error estimate F. From section 3.2 we fetch

$$F = \sqrt{1 - \frac{\int_0^{\xi_{BK}} b(\xi;\epsilon)\,\mathrm{d}\xi}{\int_0^{\xi_{BK}} b(\xi;0)\,\mathrm{d}\xi}} \tag{99}$$

where

$$\xi_{BK} = 2RL_{BK} = AkR \tag{100}$$

cf equation (14). This expression for F was obtained by perturbing the Berry–Keating formula. The function $b(\xi;\epsilon)$, introduced in section 3.2, captures the average behaviour of the coefficients of cycle expansions (50) and (51).

To begin with, we are only interested in the leading asymptotic behaviour of the function $b(\xi;\epsilon)$. In section 3.2 we related it to various zeta functions, and got to leading order

$$b(\xi;\epsilon) \sim \frac{Z_2(z_0;\epsilon)}{Z'(z_0;\epsilon)} \mathrm{e}^{z_0\xi} \tag{101}$$

valid for large values of ξ.

In section 3.3 we used the BER approximation to find

$$z_0 \sim -\epsilon \tag{102}$$

$$Z'(z_0;\epsilon) \sim 1 \tag{103}$$

$$Z_2(z_0;\epsilon) \sim Z_2(0,0) \tag{104}$$

to leading order in ϵ, for error bounds please go back to section 3.3.

Finally, from section 2.5 we have

$$\epsilon = c(kR)^{-2/3}. \tag{105}$$

We are then in the position to compute the large kR limit of the error estimate, which we easily evaluate to

$$F \sim \left(1 - \frac{1 - \mathrm{e}^{-\epsilon\xi_{BK}}}{\epsilon\xi_{BK}}\right)^{1/2} = \left(1 - \frac{1 - \mathrm{e}^{-cA(kR)^{1/3}}}{cA(kR)^{1/3}}\right)^{1/2}. \tag{106}$$

We observe that this function will definitely approach unity which implies that individual eigenstates will cease to be resolved. This final collapse will occur on the scale

$$kR \sim (cA)^{-3}. \tag{107}$$

But how is the asymptotic expression (106) approached?

Actually we know that $F = 0$ if kR is less than some critical value which is given by $\xi_{BK} = 1/(1+\epsilon/2)$ (cf equation (95)) or written out explicitly

$$AkR = \frac{1}{1+\frac{c}{2}(kR)^{-2/3}}. \tag{108}$$

The solution is given by

$$(kR)_{\text{threshold}} = (\tfrac{1}{6}\alpha^{1/3} - c\alpha^{-1/3})^3 \tag{109}$$

where

$$\alpha = \frac{108}{A} + 6\sqrt{6c^3 + 324/A^2}. \tag{110}$$

Unstable periodic orbits below this threshold are never diffractive. Of course, neutral orbits in this range are subject to diffraction corrections but we only consider the error due to unstable orbits. We are thus unable to estimate the semiclassical error for small values of kR.

To estimate the intermediate behaviour we introduce a further approximation of equation (57)

$$b(\xi;\epsilon) = \frac{1}{2\pi\mathrm{i}}\int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty}\frac{Z_2(z;\epsilon)}{Z(z;\epsilon)}\mathrm{e}^{z\xi}\,\mathrm{d}z \approx Z_2(0,0)\int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty}\frac{1}{Z(z;\epsilon)}\mathrm{e}^{z\xi}\,\mathrm{d}z \tag{111}$$

which is highly reasonable, cf equation (98). We replace the classical zeta function $Z(z,\epsilon)$ by the approximate one $\hat{Z}(z,\epsilon)$, the transforms are again computed by the FFT technique and the resulting function $F(kR)$ is plotted in figure 11. In the computation we use $A = \frac{1}{8}$ and an arbitrary value of $c = 1$. What we see is a very steep ascent at the threshold $(kR)_{\text{threshold}}$ as discussed above and a fast approach to the asymptotic formula (106). It is likely that individual states cease to be resolved long before the scale (107). However, it is hard to make any safe prediction regarding the crudeness of our treatment of the penumbra problem, see also the discussion of c below. However, the function F *does* approach unity, sooner or later, and it is difficult to find an prevarication of this fact.

We note from figure 11 that above the threshold, and up to fairly large values of k, the function $F(k)$ is approximately $F(k) \approx 0.4$. This corresponds to a semiclassical being 10% of the mean spacing, which is consistent with the findings of [6].

There is an issue which we appear to have forgotten: the constant c in equation (126) is not really a constant, it depends weakly on kR, R, r_1 and r_2, see figure 1 and the appendix. First, this dependence is too weak to be able to alter our general conclusions. Moreover, it is not obvious how to implement the dependence on r_1 and r_2. In the BER application we consider disc-to-disc segments whereas in the study of the circle Green function in section 2.5 we consider (square) boundary-to-boundary segments. To make the exact connection one has to convolute the circle Green function with itself, cf equation (30), the outcome of this operation is not obvious. So we used c as if it was a constant when implementing the BER approximation. Therefore, the 'constant' c coming out of the other end of the BER calculation should have some residual weak dependence on R and kR.

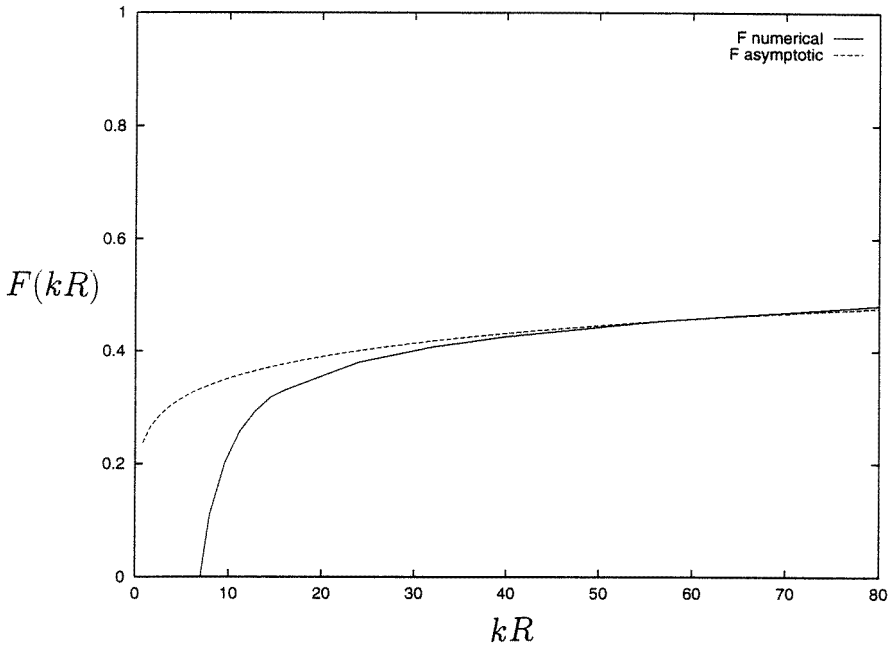

Figure 11. The error estimate $F(kR)$ versus kR according to equation (111) (full line) compared with the asymptotic expression (99) (dashed line). $A = \frac{1}{8}$ and c is chosen to be $c = 1$.

5. Discussion of the validity of the various approximations involved

No chain is stronger that the weakest link. The results presented here are based on a long series of approximations and assumptions. Some of them may be readily justified and should hardly be controversial but some may seem much cruder and one may ask if they will allow the chain to break.

5.1. The correction to the semiclassical weights

On one hand we argued that penumbra diffraction cannot be accounted for by multiplicative corrections but on the other hand we *needed* multiplicative corrections to be able to use the machinery of the periodic orbit theory and evolution operators. So we simply constructed a multiplicative weight inspired by the penumbra diffraction that should be able to provide us with an estimate of the error in the Berry–Keating formula. The procedure was discussed at some length in section 2.5.

One could object that it is too crude to approximate the gradual transition of the circle Green function with a step function, and suggest a smoother weight. This is in principle possible, but that would make the calculation in section 3.3 immensely more complicated, and hardly change the result in any significant way.

5.2. The BER approximation

It is natural that a theory for the asymptotics of the periodic orbits is hard to check numerically. In [24] we compared the exact trace formulae with those of the BER approximation for lengths up to roughly the horizon $1/2R$ for the Sinai billiard. The results were encouraging but hardly

asymptotic.

However, a range of asymptotic predictions of the BER approximation appears to be correct. It does provide the suggested exact diffusion behaviour in the regular Lorentz gas (with unbounded horizon) [26, 37] and the related correlation decays [38] as well as the small radius limit of the Lyapunov exponent [36]. We therefore feel confident that the approach works even if more rigorous results are called for.

We introduced some extra approximations, valid in the small R limit, and our experience is that they work excellently even for rather large R.

One could again raise the objection that the diffractive weight is discontinuous and that this could cause problems for averages to settle down. However, we saw in section 3.3 that the effect is just to change the pruning rules slightly, the billiard is as discontinuous as before and the fluctuations of the pseudo-orbit sums appearing in the numerator and denominator (58) are similar. The integrals in equation (58) are self-averaging and the fluctuations of the integrands irrelevant for the estimate of the error.

5.3. The diagonal approximation

The diagonal approximation underlying equation (44) can be verified assuming that the spectral statistics is given by random matrix theories [39]. Random matrix results are expected to apply to the Sinai billiard for large values of kR and our attempts to make predictions have already been restricted to this limit, mainly because of the problems posed by neutral orbits.

The diagonal approximation on the diffractive sum behind (45) is natural, in particular since the majority of pseudo-orbits are diffractive for small ϵ, but by no means obvious.

Acknowledgments

I am grateful to Jon Keating for a very useful discussion. This work was supported by the Swedish Natural Science Research Council (NFR) under contract no F-AA/FU 06420-313.

Appendix. Stationary phase analysis of the circle Green function

We will consider scatterings in extreme forward angles so only the second term in the integral (34) ($m = 0$) contributes. This means that we only need to evaluate the integral

$$G^{(0)}(r_1, r_2, \Delta\theta) = \frac{\mathrm{i}}{8}\int_{-\infty}^{\infty} S_\ell(kR) H_\ell^{(1)}(kr_1) H_\ell^{(1)}(kr_2) \mathrm{e}^{\mathrm{i}\ell\Delta\theta}\,\mathrm{d}\ell \tag{112}$$

where $0 \ll \Delta\theta < \pi$. Actually, this integral is divergent but as long as we study the stationary phase approximation, it serves our purposes. We can still safely use the Debye approximation for the Hankel functions $H_\ell^{(1)}(kr_1)$ and $H_\ell^{(1)}(kr_2)$

$$H_\ell^{(1)}(z) \sim \sqrt{\frac{2}{\pi\sqrt{z^2-\ell^2}}}\mathrm{e}^{\mathrm{i}[\sqrt{z^2-\ell^2}-\ell\arccos(\ell/z)-\pi/4]} \tag{113}$$

because $kr_1 \gg \ell$ and $kr_2 \gg \ell$. However, the phase shift function needs a more careful analysis. The phase shift function $S_\ell(kR)$ is of unit modulus and we call the phase $\gamma(kR, \ell)$

$$S_\ell(kR) = -\frac{H_\ell^{(2)}(kR)}{H_\ell^{(1)}(kR)} \equiv \mathrm{e}^{\mathrm{i}\gamma(kR,\ell)}. \tag{114}$$

The Green function now reads

$$G^{(0)}(r_1, r_2, \Delta\theta) = \int_{-\infty}^{\infty} A(r_1, r_2)\mathrm{e}^{\mathrm{i}(\ell\Delta\theta - \Psi(\ell))}\,\mathrm{d}\ell \tag{115}$$

where the slowly varying amplitude $A(r_1, r_2)$ is composed of the Hankel functions $H_\ell^{(1)}(kr_1)$ and $H_\ell^{(1)}(kr_2)$. The asymptotics of the integral is determined by the phase function

$$\Psi(\ell) = -\gamma(kR, \ell) - \left[\sqrt{(kr_1)^2 - \ell^2} - \ell \arccos(\ell/(kr_1))\right] - \left[\sqrt{(kr_2)^2 - \ell^2} - \ell \arccos(\ell/(kr_2))\right] + \pi/2. \tag{116}$$

We will now investigate the phase of $S_\ell(kR)$ in detail. To this end we will use the uniform approximation for Hankel functions relating the phase of Hankel functions to the phase of Airy functions according to [40]

$$S_\ell(kR) = -\frac{\mathrm{Ai}(-x) + \mathrm{iBi}(-x)}{\mathrm{Ai}(-x) - \mathrm{iBi}(-x)} \equiv \mathrm{e}^{\mathrm{i}\gamma(\ell, kR)} \tag{117}$$

where

$$x(kR, \ell) = \begin{cases} \left[\frac{3}{2}(\sqrt{(kR)^2 - \ell^2} - \ell \arccos(\ell/(kR)))\right]^{2/3} & \ell < kR \\ -\left[\dfrac{3}{2}\left(\ell \cdot \log\left(\dfrac{\ell + \sqrt{\ell^2 - (kR)^2}}{kR}\right) - \sqrt{\ell^2 - (kR)^2}\right)\right]^{2/3} & \ell > kR. \end{cases} \tag{118}$$

Defining

$$\ell = kR(1 + \epsilon) \tag{119}$$

we have for small ϵ

$$-x = (\sqrt{2}kR)^{2/3}(\epsilon + \mathrm{O}(\epsilon^2)). \tag{120}$$

Using this expression for x one gets the so-called transition region approximation. This approximation fails to yield the Debye approximation as its asymptotic limit but the nice thing is that there is a considerable overlap because whenever $x \ll \ell^{2/3}$ we can use the transition region approximation and when $x \gg 1$ we can use the Debye approximation.

$\gamma(x)$ is a complicated function but one has the following asymptotic expressions [40]:

$$\gamma(x) = -\mathrm{e}^{-\frac{4}{3}(-x)^{3/2}}(1 + \mathrm{O}(1/(-x)^{3/2})) \qquad x \to -\infty \tag{121}$$

$$\gamma(x) = -\frac{\pi}{2} - \frac{4}{3}x^{3/2} + \mathrm{O}(x^{-3/2}) \qquad x \to +\infty \tag{122}$$

so we obtain the Debye approximation when $x \to \infty$ as expected.

The stationary phase condition will simply read $\Psi_\ell(\ell) = \Delta\theta$ (subscripts denote differentiation). In figure 12 we plot the function $\Psi_\ell(\ell)$ versus ℓ for some arbitrary chosen values of $r_1/R = r_2/R = 3$ and $kR = 30$ together with its Debye approximation. The plot gives a picture of how the stationary phase will perform for all $\Delta\theta$, just place the ruler horizontally at $\Delta\theta$. If it goes below the maximum, it intersects the curve twice. The corresponding saddle points correspond to the direct (larger ℓ) and reflected ray (smaller ℓ), respectively.

We are interested in the location of the maximum of $\Psi_\ell(\ell)$. This turns out to lie in the region where equations (120) and (121) apply. To leading order in (kR) the location of the maximum is the solution to the equation

$$-x\mathrm{e}^{-\frac{4}{3}(-x)^{3/2}} = 2^{-8/3}(kR)^{-1/3}\left(\frac{R}{\sqrt{r_1^2 - R^2}} + \frac{R}{\sqrt{r_2^2 - R^2}}\right) \equiv (kR)^{-1/3}C(r_1, r_2). \tag{123}$$

By elementary methods one can show that the solution to this equation lies in the range

$$\frac{9}{16}\left(\log\frac{(kR)^{1/3}}{C(r_1, r_2)}\right)^{2/3} < (-x)_{max} < \left(\log\frac{(kR)^{1/3}}{C(r_1, r_2)}\right)^{2/3}. \tag{124}$$

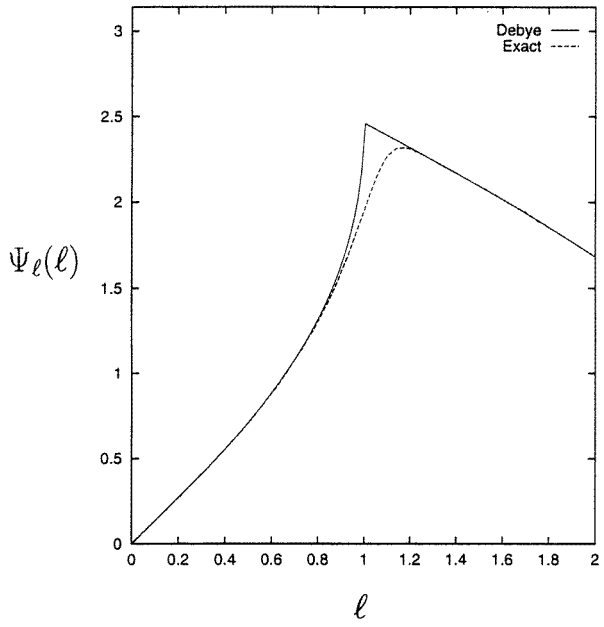

Figure 12. The function $\Psi_\ell(\ell)$ versus ℓ for fixed values of $k = 30$, $R = 1$ and $r_1 = r_2 = 3$.

Where in this range the solution lies is completely irrelevant for us. We use this liberty and choose ϵ in (35) to be (cf equation (120))

$$\epsilon_{max} = 2^{-1/3}(kR)^{-2/3}\left(\log\frac{(kR)^{1/3}}{C(r_1, r_2)}\right)^{2/3} \equiv c(kR)^{-2/3}. \tag{125}$$

where

$$c = c(kR, R, r_1, r_2) = 2^{-1/3}\left(\frac{1}{3}\log kR + \frac{8}{3}\log 2 - \log\left(\frac{R}{\sqrt{r_1^2 - R^2}} + \frac{R}{\sqrt{r_2^2 - R^2}}\right)\right)^{2/3}. \tag{126}$$

References

[1] Gutzwiller M C 1990 *Chaos in Classical and Quantum Mechanics* (New York: Springer)
[2] Dahlqvist P 1991 *J. Phys. A: Math. Gen.* **24** 4763
[3] Christiansen F and Cvitanović P 1992 *Chaos* **2** 61
[4] Tanner G, Scherer P, Bogomolny E B, Eckhardt B and Wintgen D 1991 *Phys. Rev. Lett.* **67** 2410
[5] Sieber M and Steiner F 1991 *Phys. Rev. Lett.* **67** 1941
[6] Primack H and Smilansky U 1998 *J. Phys. A: Math. Gen.* **31** 6253
[7] Heller E J, Tomsovic S and Sepélveda M A 1992 *Chaos* **2** 105
[8] Berry M V and Keating J P 1990 *J. Phys. A: Math. Gen.* **23** 4839
[9] Cvitanović P and Eckhardt B 1989 *Phys. Rev. Lett.* **63** 823
[10] Boasman P A 1994 *Nonlinearity* **7** 485
[11] Bogomolny E B 1992 *Nonlinearity* **5** 805
[12] Hannay J H and Berry M V 1980 *Physica* D **1** 267
[13] Keating J P 1991 *Nonlinearity* **4** 309
[14] Gutzwiller M C 1980 *Phys. Rev. Lett.* **45** 150
[15] Primack H, Schanz H, Smilansky U and Ussishkin I 1996 *Phys. Rev. Lett.* **76** 1615

[16] Saraceno M and Voros A 1992 *Chaos* **2** 99
[17] Sundaram B and Scharf R 1995 *Physica* D **83** 257
[18] Sieber M, Pavloff N and Schmit C 1997 *Phys. Rev.* E **55** 2279
[19] Schomerus H and Sieber M 1997 *J. Phys. A: Math. Gen.* **30** 4537
[20] Tanner G 1997 *J. Phys. A: Math. Gen.* **30** 2863
[21] Hannay J H and Ozorio de Almeida A M 1984 *J. Phys. A: Math. Gen.* **17** 3429
[22] Nielsen S F, Dahlqvist P and Cvitanović P Periodic orbit sumrules: accelerating cycle expansions *J. Phys. A: Math. Gen.* submitted
[23] Dahlqvist P 1994 *J. Phys. A: Math. Gen.* **27** 763
[24] Dahlqvist P 1995 *Nonlinearity* **8** 11
[25] Dahlqvist P 1997 *J. Technol. Phys.* **38** 189
[26] Dahlqvist P 1996 *J. Stat. Phys.* **84** 773
[27] Baladi V, Eckmann J P and Ruelle D 1989 *Nonlinearity* **2** 119
[28] Voros A 1988 *J. Phys. A: Math. Gen.* **21** 685
[29] Cvitanović P *et al* 1997 *Classical and Quantum Chaos: a Cyclist Treatise* webpage: http://www.nbi.dk/ChaosBook/ (Copenhagen: Niels Bohr Institute)
[30] Dahlqvist P 1997 *Chaos Solitons Fractals* **8** 1011
[31] Karney C F F 1983 *Physica* D **8** 360
[32] Berry M V and Wilkinson M 1984 *Proc. R. Soc.* A **392** 15
[33] Levy B R and Keller J B 1959 *Commun. Pure Appl. Math.* **12** 159
[34] Hansen K T 1900 Symbolic dynamics in chaotic systems *PhD Thesis* Oslo
[35] Artuso R, Aurell E and Cvitanović P 1990 *Nonlinearity* **3** 325
Artuso R, Aurell E and Cvitanović P 1990 *Nonlinearity* **3** 361
[36] Dahlqvist P 1997 *Nonlinearity* **10** 159
[37] Bleher P M 1992 *J. Stat. Phys.* **66** 315
[38] Dahlqvist P and Artuso R 1996 *Phys. Lett.* A **219** 212
[39] Keating J P Private communication
[40] Abramovitz M and Stegun I A 1964 *Handbook of Mathematical Functions* (Washington, DC: National Bureau of Standards)