DEUTSCHES ELEKTRONEN-SYNCHROTRON

DESY 93-044
April 1993

Some Studies on Arithmetical Chaos in Classical and Quantum Mechanics

J. Bolte
II. Institut für Theoretische Physik, Universität Hamburg

ISSN 0418-9833

NOTKESTRASSE 85 · D-2000 HAMBURG 52

DESY 93-044
April 1993

ISSN 0418-9833

Some Studies on Arithmetical Chaos in Classical and Quantum Mechanics

Jens Bolte

II. Institut für Theoretische Physik
Universität Hamburg
Luruper Chaussee 149
2000 Hamburg 50
Federal Republic of Germany

Abstract

Several aspects of classical and quantum mechanics applied to a class of strongly chaotic systems are studied. The latter consists of single particles moving without external forces on surfaces of constant negative Gaussian curvature whose corresponding fundamental groups are supplied with an arithmetic structure.

It is shown that the arithmetical features of the considered systems lead to exceptional properties of the corresponding spectra of lengths of closed geodesics (periodic orbits). The most significant one is an exponential growth of degeneracies in these geodesic length spectra. Furthermore, the arithmetical systems are distinguished by a structure that appears as a generalization of geometric symmetries. These pseudosymmetries occur in the quantization of the classical arithmetic systems as Hecke operators, which form an infinite algebra of self-adjoint operators commuting with the Hamiltonian.

The statistical properties of quantum energies in the arithmetical systems have previously been identified as exceptional. They do not fit into the general scheme of random matrix theory. It is shown with the help of a simplified model for the spectral form factor how the spectral statistics in arithmetical quantum chaos can be understood by the properties of the corresponding classical geodesic length spectra. A decisive role is played by the exponentially increasing multiplicities of lengths. The model developed for the level spacings distribution and for the number variance is compared to the corresponding quantities obtained from quantum energies for a specific arithmetical system.

Finally, the convergence properties of a representation for the Selberg zeta function as a Dirichlet series are studied. It turns out that the exceptional classical and quantum mechanical properties shared by the arithmetical systems prohibit a convergence of this important function in the physically interesting domain.

Contents

1 Introduction

Two classes of theoretical activities in the field of natural sciences may be distinguished. The first one consists of efforts towards unraveling the kinematics of a given system and of finding the underlying dynamics governing the time evolution of the system; the second one is formed by the development of methods and techniques to obtain numerical data from the theory describing the dynamics in order to be able to compare them with empirical data. A sensible and satisfying theory of natural phenomena then has to meet these two requirements: it should be logically consistent and ought produce testable predictions. Both of these aspects have to be well distinguished. One consequence of this consideration is that a theory containing deterministic dynamics need not be predictive with respect to its time evolution. This realization lies at the heart of what nowadays is widely known as *deterministic chaos.*

To be more specific we now consider Hamiltonian dynamical systems (with a finite number N of degrees of freedom) in classical mechanics. These may be described on N–dimensional configuration manifolds M with Riemannian metrics $ds^2 = g_{ij}dq^i dq^j$ defined on them. ($\vec{q} = (q^1,\dots,q^N)$ are local coordinates on M.) The dynamics is specified by providing M with a Lagrangian function $L(\vec{q},\dot{\vec{q}}) = \frac{1}{2}(\frac{ds}{dt})^2 - V(\vec{q})$. The equations of motion for this system may then be obtained by Hamilton's principle as the Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0\ , \quad i = 1,\dots,N\ . \tag{1.1}$$

Specifying initial values $\vec{q}_0$ and $\dot{\vec{q}}_0$ at a time t_0 then uniquely fixes the time evolution $\vec{q}(t)$ of the system for all times $t \geq t_0$. This is the manifestation of determinism in classical mechanics. As yet, nothing is said, however, about predictability. Since in practice it is never possible to prepare a system at an initial time t_0 to be in a definite state $(\vec{q}_0,\dot{\vec{q}}_0)$, one has to allow the initial values to be taken from $[\vec{q}_0-\vec{\varepsilon},\vec{q}_0+\vec{\varepsilon}]\times[\dot{\vec{q}}_0-\vec{\delta},\dot{\vec{q}}_0+\vec{\delta}]$ for some small uncertainties $\vec{\varepsilon}$ and $\vec{\delta}$. Predictability would now require the uncertainties to grow only modestly under the time evolution dictated by (1.1). By this we mean an increase of $|\vec{\varepsilon}|$ and $|\vec{\delta}|$ at most like a power of t. It is, however, by no means clear from the information provided so far that this will be the case.

The class of dynamical systems possessing the most regular kind of time evolution is given by the *integrable* ones. For them there exist N independent constants of motion with pairwise vanishing Poisson brackets. If one considers their time evolution in phase space (i.e. on the cotangent bundle T^*M), this is found to take place on an N–dimensional torus. The equations of motion can then be integrated by quadratures. Integrable systems are predictable in the sense just introduced.

The other extreme is given by irregular systems sharing the property of *ergodicity.* Almost all of their phase space trajectories fill the $(2N-1)$–dimensional hypersurface of constant energy in phase space densely. The probability of finding an arbitrary phase point in some bounded region of the hypersurface of constant energy is proportional to the volume of that region. This means that the trajectories of ergodic systems are uniformly distributed in phase space. Among such systems one can find a hierarchy of even higher irregularities: mixing systems, Anosov–systems, K–systems, ...; see e.g. [37] as a reference.

In between the two extremes of integrable and ergodic systems there exists any kind of intermediate behaviour. These systems, however, will not be dealt with in the course of the present investigation.

It now appears to be useful to reformulate the setting a little bit. Suppose the system under study is in a state of energy E. According to the conservation of energy it can visit only those parts of M during its time evolution, where $V(\vec{q}) \leq E$ is fulfilled. On this domain of M one introduces the *Jacobi metric* $dS_E^2 := [E - V(\vec{q})]ds^2$. Maupertuis' principle of least action, being equivalent to the equations of motion (1.1), is now also equivalent to the statement that the trajectories of the system on M are geodesics in the Jacobi metric dS_E^2, see e.g. [1] for details. Hence every Hamiltonian dynamics can be viewed as the geodesic flow on some Riemannian manifold. Hopf [56] has shown that a negative curvature associated with the Jacobi metric is a sufficient condition to render the system ergodic.

A guiding principle in the present study will be to keep things as simple as possible, without giving up the essential structures that determine the properties of a –hopefully– general enough class of systems. Hence we agree to restrict our attention to the following kind of examples. They will have two degrees of freedom, since this is the minimal dimensionality required for a Hamiltonian system to be non–integrable; the reason for this being that conservation of energy renders every system of one degree of freedom integrable. Having in mind the above remark on the Jacobi metric and its curvature, we choose geodesic flows on two dimensional Riemannian surfaces of constant negative Gaussian curvature as our prime examples. We are then going to study the interplay between the classical and quantum dynamics for these systems.

A central issue of this investigation will be to identify fingerprints of the classical properties of a given dynamical system in its quantum version. Since the time evolution in (non-relativistic) quantum mechanics, which is governed by the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi(t,\vec{q}) = H\psi(t,\vec{q}) \tag{1.2}$$

for the time dependent wave function $\psi(t,\vec{q})$, is linear, no chaotic phenomena like an exponential sensibility to initial conditions can occur. Thus the question arises, how a potential irregularity of the classical system can show up in the semiclassical limit $\hbar \to 0$. Although this limit is not "smooth" in that at the value $\hbar = 0$ typical quantum mechanical structures and quantities no longer exist, it is expected that for small values of $\hbar$ one should be able to detect classical properties of the system. One such sign of the structure of the classical phase space in quantum mechanics is the applicability of semiclassical quantization rules such as the WKB–method (for one-dimensional systems) or the EBK–method (for multi-dimensional systems), which is restricted to the integrable case. Therefore a first and obvious question would be to find a semiclassical quantization procedure e.g. for strongly chaotic classical systems.

An answer to this question is provided by Gutzwiller's *periodic-orbit theory* [42, 44], in which the spectrum of the quantum Hamiltonian H is determined in a semiclassical approximation by the set of actions evaluated along the classical periodic trajectories (*periodic orbits*) of the system. *Plane billiards* seem to be well suited for an explanation of the general procedure. Let therefore $D \subset I\!R^2$ be some connected domain on the euclidean plane. Its boundary ∂D is assumed to be piecewise smooth. A plane billiard then consists of a particle moving freely inside D and being elastically reflected on ∂D, which will be chosen such that the classical dynamics is strongly chaotic. This should mean that the system is ergodic and that all periodic orbits are isolated in phase space and are unstable. The action S_γ along a periodic orbit γ is then given by $S_\gamma = pl_\gamma$, where p denotes the particle's momentum and l_γ is the length of the orbit γ. The quantum mechanical Hilbert space for such a billiard is given by the space of square integrable functions $\psi(x,y)$ on D that vanish on ∂D. In units where $\hbar = 1 = 2m$ the quantum Hamiltonian is $H = -\Delta_E = -(\partial_x^2 + \partial_y^2)$ and possesses a purely discrete spectrum $0 < E_1 \leq E_2 \leq E_3 \ldots$, $E_n = p_n^2$. Weyl's famous law for the number $N(E)$ of energy eigenvalues

not exceeding E states that asymptotically for $E \to \infty$

$$N(E) \sim \frac{\text{area}(D)}{4\pi}E\,. \tag{1.3}$$

Thus $E_n \sim \frac{4\pi}{\text{area}(D)}n$, $n \to \infty$, and hence the resolvent operator $(H - E)^{-1}$ is not of trace class. Introducing, however, a suitable smearing allows to investigate trace class operator valued functions of H. In [88] one can find a regularized version of *Gutzwiller's trace formula.* Asymptotically in the semiclassical limit it reads

$$\sum_{n=1}^{\infty} h(p_n) \sim 2\int_0^{\infty} dp\,p\,\bar{d}(p)h(p) + \sum_\gamma\sum_{k=1}^{\infty} \frac{\chi_\gamma^k\, l_\gamma\, g(kl_\gamma)}{e^{kl_\gamma\lambda_\gamma/2} - \sigma_\gamma^k e^{-kl_\gamma\lambda_\gamma/2}}\,. \tag{1.4}$$

In this formula $h(p)$ is an even function, holomorphic in the strip $|Im\,p| \leq \tau - \frac{\bar{\lambda}}{2} + \varepsilon$, $\varepsilon > 0$, that decreases faster than $|p|^{-2}$ for $|p| \to \infty$. $g(x) := \int_{-\infty}^{+\infty}\frac{dp}{2\pi}e^{ipx}h(p)$, and $\bar{d}(p)$ is a mean spectral density expressed as a function of momentum $p = \sqrt{E}$. Under these assumptions the integral and the sums involved in (1.4) are absolutely convergent. $\chi_\gamma = (-1)^{j_\gamma}$, where j_γ is the number of reflections on ∂D when traversing γ once. $\lambda_\gamma = u_\gamma/l_\gamma$ is the Lyapunov exponent of γ and u_γ is its stability exponent. σ_γ denotes the sign of the trace of the monodromy matrix for γ. The *topological entropy* τ describes the proliferation of the number $\mathcal{N}(l)$ of periodic orbits of lengths not exceeding l by

$$\mathcal{N}(l) \sim \frac{e^{\tau l}}{\tau l}\,, \quad l \to \infty\,. \tag{1.5}$$

$\bar{\lambda}$ denotes the asymptotic average of the λ_γ's and is also known as the *metric entropy.*

The trace formula (1.4) can be viewed as an identity in the sense of distributions for the spectral density $d(E) = \sum_{n=1}^{\infty}\delta(E - E_n)$ (expressed by the variable p) of the form $d(p) \sim \bar{d}(p) + d_{fl}(p)$. $d_{fl}(p)$ is an oscillatory correction to the mean density $\bar{d}(p)$, determined by the periodic-orbit sum in (1.4). The conditions on $h(p)$ stated after (1.4) then define the space of test functions.

It is possible to use Gutzwiller's trace formula in order to derive semiclassical quantization rules that yield approximations for the quantum energies $\{E_n\}$ in terms of the actions of periodic orbits in the corresponding classical system, see e.g. [44] and the references therein. In some sense the so-obtained quantization rules are a substitute for the EBK-quantization of classically integrable systems. A fundamental difference between the semiclassical quantization schemes for classically integrable and chaotic systems is provided by the computational effort to be spent in order to resolve some quantum energy E_n. In the integrable case this is independent of the energy, whereas (1.5) requires for chaotic systems an exponentially increasing number of periodic orbits to be taken into account when trying to compute higher and higher energies. This observation reflects the high degree of irregularity and complexity of chaotic systems.

Historically, the first chaotic systems under investigation were of mathematical nature and mainly served as examples to develop the mathematical theory of dynamical systems. Hadamard [46] studied the geodesic flow on closed surfaces endowed with Riemannian metrics of constant negative Gaussian curvature, which also go under the notion of *hyperbolic surfaces*, and discovered an instability of the trajectories of such flows. Since the surfaces he studied cannot be realized as being embedded in $I\!R^3$, they were considered as purely mathematical examples. For a specific hyperbolic surface Artin could later prove [2] the ergodicity of the geodesic flow on it. Moreover, at this occasion he introduced a symbolic dynamics for this system. This was the first time that the property of ergodicity could ever be rigorously demonstrated for a dynamical system. One could thus view Artin's paper as the foundation of the modern theory of dynamical systems. Only after the development of powerful computers that allow for a quantitative analysis of irregular dynamical systems these subjects arose interest among a considerable number of physicists. Starting with the series of papers [42] by Gutzwiller a discussion of the quantization of chaotic classical systems was rendered possible. Again it was Gutzwiller who noticed [43] that his trace formula turned into an exact identity when applied to geodesic flows on surfaces of constant negative curvature. In mathematics this identity had long before been introduced by Selberg [84], who intended to understand the Riemann zeta function and the Riemann hypothesis with the help of his trace formula.

Since Gutzwiller's observation hyperbolic surfaces have been intensively studied in the context of the quantization of classically chaotic systems. The first numerical results on quantum energies for a hyperbolic triangle obtained from a solution of the stationary Schrödinger equation by Schmit have been presented in [31, 17]. Aurich and Steiner [10] first calculated lengths of periodic orbits and also quantum energies [9, 11] on a specific hyperbolic surface, the so-called *regular octagon*. The lengths were used as an input to evaluate the Selberg trace formula [9, 11]. By chance, the regular octagon that had been chosen for the numerical studies turned out to be very specific: its fundamental group is of an arithmetical origin. It was realized [10, 4] that the arithmetic structure underlying the regular octagon leads to exponentially increasing multiplicities of lengths of periodic orbits. As was later noticed [12, 13], this property is exceptional and has remarkable consequences. One of these is the occurrence of unexpected statistical properties of the related quantum energy spectra. Although the classical system is as chaotic as it could be, the quantum spectral statistics are more alike those of classically integrable systems than those expected for generic classically chaotic ones. Namely, due to a conjecture of Bohigas, Giannoni and Schmit [30] which was mainly based on numerical observations, generic chaotic systems excel by quantum energy spectra that can be well described by eigenvalues of large random matrices. The resulting spectral statistics then differ drastically from those in the integrable case as obtained before by Berry and Tabor [24]. Somewhat later it was shown [12], however, that the energy fluctuations of 30 non-arithmetic octagons are in agreement with the predictions of random matrix theory. At the time the unexpected statistical properties of the eigenvalues for the regular octagon were first observed it was, however, not clear that these follow from the arithmetic properties of the fundamental group. This was later clearly expressed to hold for all arithmetical systems in the two simultaneously appearing papers [27] and [32].

The objective of the present investigation now is to study the class of dynamical systems arising from geodesic flows on hyperbolic surfaces with arithmetic fundamental groups both from a classical as well as from a quantum mechanical point of view. The ultimate goal then will be to understand the exceptional statistics of the quantum energy spectra. The problem of studying the wave functions arising on the quantum mechanical side, however, will not be treated here. It seems that the eigenfunctions of the arithmetical systems are not as exceptional as the eigenvalue spectra, see [54, 79, 16].

The organization of this paper is as follows. Chapter 2 reviews some general properties of (discrete) quantum energy spectra and introduces the necessary means and notions for their investigation in order to prepare for studying the spectral statistics of arithmetical systems. Since in order to be ready for the latter the classical aspect of the problem has to be understood in quite some detail, chapter 3 is devoted to an investigation of geodesic length spectra on arithmetic hyperbolic surfaces. A central result, obtained in section 3.4, is the observation of exponentially growing degeneracies in arithmetic length spectra. Another peculiarity arising from the arithmeticity of a fundamental group is the existence of infinitely many *pseudosymmetries* on the corresponding surface. It will be attempted to give a geometric picture of this structure in section 3.5. Chapter 3 will be closed by an investigation of fluctuations of the lengths of closed geodesics on both arithmetic and non-arithmetic hyperbolic surfaces. These fluctuations are closely related to the fluctuation properties of the corresponding quantum energies. Using the latter in order to learn about the former is also known as *inverse quantum chaology*.

Chapter 4 then deals with quantum energy spectra of arithmetical systems. At first *Hecke operators* are discussed as representations of pseudosymmetries on the wave functions and it is found that this structure results in constraints on the eigenvalue spectra. This realization is taken as a first hint towards exceptional spectral statistics. Thereafter the spectral form factor is reviewed and its properties in the arithmetical case, derived from the exponentially increasing multiplicities of lengths of periodic orbits, are employed to introduce a simplified model form factor. The latter can then be used to derive a model for the level spacings distribution as well as for the number variance in arithmetical chaos. The model will also be compared to existing numerical data in order to test its quality. Finally, convergence properties of the Selberg zeta function [84], which arises from the Selberg trace formula and plays an important role for obtaining quantization rules, are studied. It is observed that the spectral statistics influence the convergence properties of this function considerably, and that the arithmetic systems suffer from a lack of convergence in the physically interesting domain.

Chapter 5 finally summarizes the findings of the preceding investigations. Furthermore, three appendices are included. The first one reviews the theory of the Riemann zeta function, which is in many respects similar to the Selberg zeta function. The analogy of these two functions is particularly useful for inverse quantum chaology. Appendix B treats the desymmetrization of a class of hyperbolic surfaces explicitly in order to illustrate the general procedure of getting rid of unwanted symmetries on hyperbolic surfaces. The final appendix collects the definitions of O–, o–, and Ω–estimates for remainder terms that occur at several instants in the main body of the text.

Parts of this thesis have previously been published in [5, 6, 32, 33].

2 Some General Remarks on Discrete Quantum Energy Spectra

Let a classical dynamical system be given with a finite number of degrees of freedom. After quantization its Hamiltonian H should have a discrete spectrum $0 < E_1 \leq E_2 \leq E_3 \ldots$. A *quantization rule* then is the specification of a function $f : \mathbb{N} \to \mathbb{R}$ such that $E_n = f(n)$. The explicit knowledge of such a quantization function completely solves the problem of determining the energy spectrum of the system. If one is merely interested in energy eigenvalues and not in wavefunctions, a quantization rule is equivalent to solving the stationary Schrödinger equation $H\psi_n = E_n\psi_n$. One possibility to obtain a quantization function is to study the *spectral staircase*

$$N(E) := \#\{E_n \, ; \, E_n \leq E\} = \sum_{E_n \leq E} 1 \, . \tag{2.1}$$

Defining then $N_0(E) := \frac{1}{2}\lim_{\epsilon\to 0}[N(E+\varepsilon)+N(E-\varepsilon)]$ yields the quantization rule

$$N_0(E) = n - \frac{1}{2} \, , \qquad n \in \mathbb{N} \, , \tag{2.2}$$

which can be converted to the condition $\cos(\pi N_0(E)) = 0$ for E to be an energy eigenvalue, see [14, 7]. This consideration stresses the importance of studying the spectral staircase thoroughly if one is interested in the quantization of a dynamical system.

There is, however, only a very restricted number of examples where an exact quantization function is explicitly known, among which are the typical textbook examples of a particle in a box, the harmonic oscillator, or the Coulomb potential. Semiclassically, quantization rules are provided by the WKB– or EBK–methods for all classically integrable systems. For these there exist canonical transformations of the classical phase space variables $(\vec{p},\vec{q})$ to action–angle variables $(\vec{I},\vec{\omega})$, such that the Hamiltonian function $H(\vec{p},\vec{q})$ transforms to one being only dependent on the actions $\vec{I}$, $H = H(\vec{I})$. Then

$$E_{\vec{m}} = H(\vec{I}_{\vec{m}}) \, , \quad \vec{I}_{\vec{m}} = \vec{m} + \frac{1}{4}\vec{\alpha} \, , \tag{2.3}$$

is the desired (semiclassical) quantization rule. In (2.3) $\vec{m}$ runs through $\mathbb{Z}^N$ and $\vec{\alpha}$ is the vector of Maslov indices.

It is one of the major goals of *quantum chaology* to obtain an analogue of (2.3) for strongly chaotic systems. Since there do not exist action–angle variables and even no remnants thereof for ergodic systems, the quantization problem has to be tackled in a totally different manner. A starting point one could think of would be to study the spectral staircase in as much detail as possible, thereby keeping in mind the relation (2.2). For such an analysis it proves useful to split $N(E)$ into a smooth part $\overline{N}(E)$, which is an approximation to $N(E)$ in that the steps have been smeared out, and a remaining contribution $N_{fl}(E)$. This then describes the fluctuations of $N(E)$ about its mean $\overline{N}(E)$. A priori there is no preferred way to define the smoothing $\overline{N}(E)$. The splitting

$$N(E) = \overline{N}(E) + N_{fl}(E) \tag{2.4}$$

should only be subject to some general requirements. $\overline{N}(E)$ should be smooth without "too many" oscillations (ideally one would require it to be monotonically increasing), and it should be asymptotically identical to the true staircase, i.e. $\overline{N}(E) \sim N(E)$, for $E \to \infty$. $N_{fl}(E)$ then should describe fluctuations about $\overline{N}(E)$, i.e. $N_{fl}(E) = N(E) - \overline{N}(E)$ should fluctuate about

zero in the limit $E \to \infty$. This requirement can be expressed in a variety of different ways, one of which is

$$\lim_{L\to\infty} \frac{1}{N(L)} \sum_{E_n \leq L} N_{fl}(E_n) = 0 \, , \tag{2.5}$$

i.e. an asymptotic vanishing of an arithmetic mean of $N_{fl}(E)$. Another, although similar, requirement uses an integral instead of a sum, namely

$$\lim_{L\to\infty} \frac{1}{L} \int_0^L dE \, N_{fl}(E) = 0 \, . \tag{2.6}$$

Provided there is a Gutzwiller trace formula available, like (1.4) for chaotic plane billiards, a natural splitting (2.4) is yielded by the two contributions to the r.h.s. of the trace formula. In order to obtain this one introduces the complex variable $s = \frac{\bar{\lambda}}{2} - ip$, i.e. $E(s) = p^2 = -(s-\frac{\bar{\lambda}}{2})^2$. A regularized trace of the resolvent operator may be introduced by choosing the function $h(p) = \frac{1}{E(s)-p^2} - \frac{1}{E(\sigma)-p^2}$, for $Re\, s$, $Re\,\sigma > \tau$, which fulfills the requirements for a test function to be used in (1.4). Inserting this into the trace formula yields

$$\begin{aligned}\sum_{n=1}^{\infty}\left[\frac{1}{E(s)-E_n} - \frac{1}{E(\sigma)-E_n}\right] \; \sim \; & 2\int_0^\infty dp\, p\, \bar{d}(p)\left[\frac{1}{E(s)-p^2} - \frac{1}{E(\sigma)-p^2}\right] \\ & - \frac{1}{2s-\bar{\lambda}} \sum_\gamma \sum_{k=1}^\infty \frac{\chi_\gamma^k l_\gamma \, e^{-(s-\frac{\bar{\lambda}}{2})kl_\gamma}}{e^{kl_\gamma\lambda_\gamma/2} - \sigma_\gamma^k e^{-kl_\gamma\lambda_\gamma/2}} \\ & + \frac{1}{2\sigma-\bar{\lambda}} \sum_\gamma \sum_{k=1}^\infty \frac{\chi_\gamma^k l_\gamma \, e^{-(\sigma-\frac{\bar{\lambda}}{2})kl_\gamma}}{e^{kl_\gamma\lambda_\gamma/2} - \sigma_\gamma^k e^{-kl_\gamma\lambda_\gamma/2}} \, .\end{aligned} \tag{2.7}$$

Choosing now $E(\sigma) \in \mathbb{R}$, i.e. $Re\,\sigma \to \frac{\bar{\lambda}}{2}+$, and keeping $Re\, s > \frac{\bar{\lambda}}{2}$, i.e. $Im\, E(s) > 0$, taking the imaginary part on both sides of (2.7), and multiplying the result with $-\frac{1}{\pi}$ yields ($E = p^2$, $p > 0$)

$$\sum_{n=1}^\infty \delta(E - E_n) \sim \bar{d}(E) - Im\left\{\frac{1}{2ip\pi}\frac{Z'(\frac{\bar{\lambda}}{2}-ip)}{Z(\frac{\bar{\lambda}}{2}-ip)}\right\} \, , \tag{2.8}$$

where the *dynamical zeta function* $Z(s)$ has been introduced. It is for $Re\, s > \tau$ defined by the *Euler product*

$$Z(s) := \prod_\gamma \prod_{n=0}^\infty \left(1 - \chi_\gamma \sigma_\gamma^n e^{-(s+n\lambda_\gamma+\frac{1}{2}(\lambda_\gamma-\bar{\lambda}))l_\gamma}\right) \, . \tag{2.9}$$

Integrating both sides of (2.8) from zero to E yields (assuming $\arg Z(\frac{\bar{\lambda}}{2}) = 0$)

$$N(E) \sim \overline{N}(E) + \frac{1}{\pi}\arg Z\left(\frac{\bar{\lambda}}{2} + i\sqrt{E}\right) \, , \tag{2.10}$$

where $\overline{N}(E) = \int_0^E dE'\, \bar{d}(E')$. One has now obtained, in the semiclassical limit $E \to \infty$, that

$$N_{fl}(E) \sim \frac{1}{\pi}\arg Z\left(\frac{\bar{\lambda}}{2} + i\sqrt{E}\right) \, . \tag{2.11}$$

This relation stresses the importance of dynamical zeta functions for the study of quantum spectral properties of chaotic dynamical systems. In addition, dynamical zeta functions turn out to be powerful tools to obtain quantization rules in periodic-orbit theory directly [89, 68, 95, 26, 63, 23, 13]. Their zeroes s_n on the *critical line* $Re\, s = \frac{\overline{\lambda}}{2}$ yield semiclassical approximations to the quantum energies via $E_n = E(s_n)$. This result may be obtained by integrating both sides of (2.7).

In [11, 14, 7] a periodic-orbit expression for the r.h.s. of (2.11) has been obtained that enabled the authors to compute $N_{fl}(E)$ numerically in some approximation from the lengths of primitive periodic orbits for several chaotic systems. To do so one inserts the Gaussian $h(p) = \frac{1}{\varepsilon\sqrt{\pi}}(e^{-(p-q)^2/\varepsilon^2} + e^{-(p+q)^2/\varepsilon^2})$, with $g(x) = \frac{1}{\pi}\cos(qx)\, e^{-\varepsilon^2 x^2/4}$, into the trace formula (1.4). Then one integrates the resulting formula in q from zero to p, $p^2 = E$. Since $\lim_{\varepsilon\to 0} h(p) = \delta(p-q) + \delta(p+q)$, the l.h.s. of (1.4) yields for $\varepsilon > 0$ a smoothing $N_\varepsilon(E)$ of the spectral staircase $N(E)$ with $\lim_{\varepsilon\to 0} N_\varepsilon(E) = N(E)$,

$$\lim_{\varepsilon\to 0} \sum_{n=1}^{\infty} \frac{1}{\varepsilon\sqrt{\pi}} \int_0^p dq \left(e^{-\frac{(p_n-q)^2}{\varepsilon^2}} + e^{-\frac{(p_n+q)^2}{\varepsilon^2}} \right) = \sum_{n=1}^{\infty} \Theta(E - E_n) = N(E) . \qquad (2.12)$$

On the r.h.s. of (1.4) one obtains $\overline{N}_\varepsilon(E) + N_{\varepsilon,fl}(E)$, with

$$N_{\varepsilon,fl}(E) = \frac{1}{\pi} \sum_\gamma \sum_{k=1}^{\infty} \frac{1}{k} \frac{\chi_\gamma^k\, e^{-\varepsilon^2 k^2 l_\gamma^2/4}}{e^{kl_\gamma\lambda_\gamma/2} - \sigma_\gamma^k e^{-kl_\gamma\lambda_\gamma/2}} \sin(\sqrt{E} k l_\gamma) . \qquad (2.13)$$

As long as $\varepsilon > 0$, the double sum on the r.h.s. of (2.13) converges absolutely, since then the Gaussian damping factor $e^{-\varepsilon^2 k^2 l_\gamma^2/4}$ overcompensates the exponential proliferation (1.5) of periodic orbits.

It will now be shown, using the smoothing $N_{\varepsilon,fl}(E)$ for the fluctuating part $N_{fl}(E)$ of the spectral staircase, that the two conditions (2.5) and (2.6) are satisfied by the splitting (2.10). Thus, in the semiclassical limit, $N_{fl}(E)$ indeed describes the fluctuations of $N(E) - \overline{N}(E)$ about zero, which is a major justification to identify $\overline{N}(E)$ as a mean spectral staircase. The essential part of demonstrating that the requirement (2.5) is met is to study, after inserting (2.13) into (2.5),

$$\frac{1}{N(L)} \sum_{p_n \le \sqrt{L}} \sin(kl_\gamma p_n) \qquad (2.14)$$

for every term in the summation over periodic orbits γ. For the following the reasonable assumption has to be made that the momenta $p_1, \dots, p_N$ are linearly independent over $\mathbb{Z}$ for any $N \in \mathbb{N}$. It is then known [61] (see also Bohigas' contribution in [41]) that the functions $\cos p_1 t, \dots, \cos p_N t$ are statistically independent. If A is any Lebesgue-measurable set in $\mathbb{R}$ and $l(A)$ is its Lebesgue-measure, define the relative measure of A as $l_r(A) := \lim_{T\to\infty} \frac{1}{2T} l\{A \cap (-T, +T)\}$. If now

$$x_N(t) := \sqrt{\frac{2}{N}} \sum_{n=1}^{N} \cos p_n t , \qquad (2.15)$$

and $A_N(a,b) := \{t;\ x_N(t) \in (a,b)\}$ denotes the set of the $t \in \mathbb{R}$ such that $x_N(t)$ takes its values in the interval (a,b), then [61]

$$\lim_{N\to\infty} l_r\{A_N(a,b)\} = \frac{1}{\sqrt{2\pi}} \int_a^b dx\, e^{-\frac{1}{2}x^2} . \qquad (2.16)$$

In other words, the values of the function $x_N(t)$ approach a normal distribution with zero mean and unit variance in the limit $N \to \infty$. If one then replaces $\cos p_n t$ by $\sin p_n t$ (for $t = kl_\gamma$),

which does not change the above argument, the values of the function

$$f_L(kl_\gamma) := \frac{1}{\sqrt{N(L)}} \sum_{p_n \le \sqrt{L}} \sin(kl_\gamma p_n) \qquad (2.17)$$

become, in the limit $L \to \infty$, Gaussian distributed with zero mean and variance $\sigma_f^2 = \frac{1}{2}$. Inserting (2.13) into (2.5) leads to a triple summation. The one over the energies E_n yields (2.14). The remaining ones over periodic orbits γ and their repetitions k sum $f_L(t)$ taken at $t = kl_\gamma$ times a prefactor,

$$\frac{1}{N(L)} \sum_{E_n \le L} N_{\varepsilon,fl}(E_n) = \frac{1}{\pi\sqrt{N(L)}} \sum_\gamma \sum_{k=1}^{\infty} \frac{1}{k} \frac{\chi_\gamma^k\, e^{-\varepsilon^2 k^2 l_\gamma^2/4}}{e^{kl_\gamma\lambda_\gamma/2} - \sigma_\gamma^k e^{-kl_\gamma\lambda_\gamma/2}} f_L(kl_\gamma) . \qquad (2.18)$$

Since $f_L(kl_\gamma)$ to be summed over is Gaussian distributed about zero, this is for $kl_\gamma \to \infty$ an effectively bounded function. Therefore the sum on the r.h.s. of (2.18) converges due to the regularizing Gaussian damping as long as $\varepsilon > 0$. In the limit $L \to \infty$ then the complete expression (2.18) vanishes like $\frac{1}{\sqrt{N(L)}}$. Weyl's law (1.3) thus determines the rate of vanishing to be of the order of $L^{-\frac{1}{2}}$.

In order to proof the validity of (2.6), when $N_{\varepsilon,fl}(E)$ from (2.13) is used instead of $N_{fl}(E)$, one has to calculate

$$\begin{aligned} \frac{1}{L} \int_0^L dE\ \sin(kl_\gamma\sqrt{E}) &= \frac{2}{L} \int_0^{\sqrt{L}} dp\ p\ \sin(pkl_\gamma) \\ &= -\frac{2}{kl_\gamma\sqrt{L}} \cos(kl_\gamma\sqrt{L}) + \frac{2}{(kl_\gamma)^2 L} \sin(kl_\gamma\sqrt{L}) , \end{aligned} \qquad (2.19)$$

which vanishes for $L \to \infty$. The double sum over periodic orbits and their repetitions one is in analogy to (2.18) left with therefore also vanishes. It is interesting to notice that both mean values (2.5) and (2.6) extended over a finite interval of length L vanish as $L^{-\frac{1}{2}}$ for $L \to \infty$. For every $\varepsilon > 0$ the conditions (2.5) and (2.6) are thus fulfilled, and hence also in the limit $\varepsilon \to 0$. It seems that any other reasonable requirement expressing the same statement on the fluctuations of $N_{\varepsilon,fl}(E)$ will also be satisfied, so that the splitting (2.10) seems to be a natural one for a semiclassical analysis of the spectral staircase $N(E)$. This point may stress the importance and usefulness of employing Gutzwiller's trace formula and dynamical zeta functions.

It should be remarked that completely analogous results can be obtained for the spectral density $d(E) = \frac{d}{dE} N(E)$. Splitting $d(E) = \overline{d}(E) + d_{fl}(E)$ according to (2.8) and using the regularization (2.13) yields

$$d_{\varepsilon,fl}(E) = \frac{1}{2\pi\sqrt{E}} \sum_\gamma \sum_{k=1}^{\infty} \frac{\chi_\gamma^k\, l_\gamma\, e^{-\varepsilon^2 k^2 l_\gamma^2/4}}{e^{kl_\gamma\lambda_\gamma/2} - \sigma_\gamma^k e^{-kl_\gamma\lambda_\gamma/2}} \cos(\sqrt{E} k l_\gamma) . \qquad (2.20)$$

Again, the two averaging procedures (2.5) and (2.6) can be employed, leading to the result that $d_{\varepsilon,fl}(E)$ vanishes with respect to those mean values if one extends the interval $(0, L)$ to be averaged over to infinity, i.e. for $L \to \infty$. Compared to $N_{\varepsilon,fl}(E)$ these means of $d_{\varepsilon,fl}(E)$ vanish faster, namely like L^{-1}.

As described above a major objective in quantum chaology is to explain and classify the statistical properties of quantum energy spectra for systems with a chaotic classical limit, and one way to achieve this is to describe the spectral staircase as explicitly as possible. Since $\overline{N}(E)$

is smooth and rather "harmless", the more interesting object is $N_{fl}(E)$. To be able to compare systems with different mean behaviours of their respective spectral staircases one conventionally introduces an *unfolding* of spectra. Define $x := \overline{N}(E)$, $x_n = \overline{N}(E_n)$. The unfolded spectrum $\{x_n\}$ has a spectral staircase that shall be denoted by $N(x)$. By its very definition, $\overline{N}(x) = x$ and thus

$$N(x) = x + N_{fl}(x) \ . \tag{2.21}$$

After being unfolded different spectra can be compared quite easily. They then only differ in the fluctuating parts of their respective spectral staircases.

A conventional measure to investigate spectral statistics is the *spectral rigidity* $\Delta_3(L;x)$, originally introduced by Dyson and Mehta [38] to analyze spectra of complicated atomic nuclei. It is defined as the integrated quadratic deviation of the spectral staircase $N(x)$ from the best fitting straight line over an interval $[x-L, x+L]$,

$$\Delta_3(L;x) := < \min_{(A,B)} \frac{1}{2L} \int_{-L}^{+L} dy\ [N(x+y) - A - By]^2 > \ . \tag{2.22}$$

$< \dots >$ denotes an averaging in x over an interval of length Δx, such that $\Delta x \ll x$ and $\Delta x \gg 1$. The latter condition means that the averaging should take place over many eigenvalues. (The mean distance of neighbouring eigenvalues is one due to $\overline{N}(x) = x$.) At the same time the interval should be much smaller than x itself. This procedure presupposes that $x \gg 1$, i.e. it is only possible in the semiclassical regime, and is thus also referred to as a *semiclassical averaging* [20]. Notice that (2.22) differs from the definition in [20] by the replacement $L \to 2L$.

Introducing the splitting (2.21) in (2.22) one obtains the minimum on the r.h.s. for

$$\begin{aligned} A_{min} &= x + \frac{1}{2L} \int_{-L}^{+L} dy\ N_{fl}(x+y) \\ B_{min} &= 1 + \frac{3}{2L^3} \int_{-L}^{+L} dy\ y\, N_{fl}(x+y) \ . \end{aligned} \tag{2.23}$$

We will soon see that $A_{min} \to x$ and $B_{min} \to 1$ in the limit $L \to \infty$. Thus one finds that

$$\Delta_\infty(x) := \Delta_3(\infty;x) = \lim_{L\to\infty} < \frac{1}{2L} \int_{-L}^{+L} dy\, [N_{fl}(x+y)]^2 > \ , \tag{2.24}$$

which means that $\Delta_\infty(x)$ is the quadratic mean of $N_{fl}(x)$. This expression will also sometimes be denoted as $< N_{fl}(x)^2 >$.

When considering the limit $L \to \infty$ one first encounters the problem to define $N_{fl}(x)$ for $x < 0$, a situation occurring in (2.22) for $L > x$. The only reasonable definition seems to be $N_{fl}(x) = 0$ for $x < 0$. Since $N(E) = 0$ for $E < 0$, an obvious choice for $\overline{N}(E)$ seems to be $\overline{N}(E) = 0$ for $E < 0$. Then also $N_{fl}(E) = 0$ for $E < 0$. This is consistent with the fact that $s = \frac{\lambda}{2} - i\sqrt{E} \in I\!R$ for $E < 0$, and thus $\arg Z(\frac{\lambda}{2} + i\sqrt{E}) = 0$. We therefore cut off, in the limit $L \to \infty$, the integrations on the r.h.s. of (2.23) at $-x$.

As our whole discussion deals with two dimensional problems, Weyl's law (see e.g. (1.3)) yields $\overline{N}(E) \sim const.\ E$, $E \to \infty$. Thus $\overline{d}(E) \sim \overline{d} = const.$, $E \to \infty$. We are interested in the semiclassical limit and hence use $x = \overline{N}(E) \sim \overline{d}\, E$ from now on. Thus

$$\begin{aligned} \frac{1}{2L} \int_{-x}^{L} dy\, N_{fl}(x+y) &\sim \frac{\overline{d}}{2L} \int_0^{E+\frac{L}{\overline{d}}} dE'\ N_{fl}(E') \ , \\ \frac{3}{2L^3} \int_{-x}^{L} dy\, y\, N_{fl}(x+y) &\sim \frac{3\overline{d}^2}{2L^3} \int_0^{E+\frac{L}{\overline{d}}} dE'\ (E'-E)\, N_{fl}(E') \ . \end{aligned} \tag{2.25}$$

A possible way to treat the integrals on the r.h.s. of (2.25) is to use the smoothing (2.12). This means that $N_{\varepsilon,fl}(E)$ for $\varepsilon > 0$ from (2.13) has to be inserted into (2.25), and the limit $\varepsilon \to 0$ has to be taken after the evaluation of the resulting integrals. The vanishing of the r.h.s. of the first line of (2.25) for $L \to \infty$ has already been demonstrated, see (2.19). The second line can be treated similarly, leading to

$$\begin{aligned} \frac{3\overline{d}^2}{2L^3} \int_0^{E+\frac{L}{\overline{d}}} dE'\ E' \sin(\sqrt{E'} kl_\gamma) &= \frac{3\overline{d}^2}{L^3} \left[\frac{6\sqrt{E+\frac{L}{\overline{d}}}}{(kl_\gamma)^3} - \frac{\left(E+\frac{L}{\overline{d}}\right)^{3/2}}{kl_\gamma} \right] \cos\left(kl_\gamma \sqrt{E+\frac{L}{\overline{d}}} \right) \\ &+ \frac{3\overline{d}^2}{L^3} \left[\frac{3(E+\frac{L}{\overline{d}})}{(kl_\gamma)^2} - \frac{6}{(kl_\gamma)^4} \right] \sin\left(kl_\gamma \sqrt{E+\frac{L}{\overline{d}}} \right) \ , \end{aligned} \tag{2.26}$$

which also vanishes for $L \to \infty$. Thus for any $\varepsilon > 0$ the smoothings of (2.25) vanish in the limit $L \to \infty$. From this we conclude that also after performing $\varepsilon \to 0$ the integrals on the r.h.s. of (2.23) vanish in the limit $L \to \infty$, leading to $A_{min} \to x$ and $B_{min} \to 1$, as it has been claimed above.

For finite values of L the spectral rigidity measures deviations of the spectral staircase from the best fitting straight line over an interval of length $2L$. Therefore $\Delta_3(L;x)$ indicates correlations in a quantum energy spectrum on a scale L. For $L \to 0$ the fact that $N(x)$ is a step function causes the rigidity to be $\frac{2}{15}L$. For slightly larger values of L $\Delta_3(L;x)$ begins to measure spectral correlations. A completely uncorrelated spectrum excels by a rigidity of $\Delta_3(L;x) = \frac{2}{15}L$ throughout the whole range of L–values, $0 \le L < \infty$. Such a spectrum is obtained by a Poisson process through $x_{i+1} = x_i + s_i$, where the s_i's are independent outcomes of measurements of the random variable s, which is distributed according to the probability density $P(s) = e^{-s}$. The $s_i = x_{i+1} - x_i$ are called the *nearest-neighbour level spacings*, or briefly *level spacings*, of the unfolded spectrum $\{x_i\}$. Due to its universal behaviour for $L \to 0$ the rigidity is, however, not the appropriate tool to investigate short-range correlations in a spectrum. It becomes useful only for medium- and long-range correlations corresponding to $L \gg 1$, i.e. on scales taking several levels into account. The level spacings are then suitable quantities to measure correlations on small scales $L \approx \frac{1}{2}$.

The systems to be considered from now on shall be completely desymmetrized in order to avoid degeneracies in their spectra that superimpose the effects one is looking for. For such an arbitrary unfolded quantum energy spectrum one can then study the distribution of the level spacings of the $x_n \le x$ and ask the question, whether there will exist a limiting distribution $P(s)$,

$$\lim_{x\to\infty} \frac{\#\{n;\ x_n \le x,\ s_n \in (a,b)\}}{N(x)} = \int_a^b ds\ P(s) \ , \tag{2.27}$$

and what it will look like if it exists. The question for the existence of $P(s)$ is a very subtle one and no mathematically rigorous answer is known in general. For some specific examples, where one does have a rigorous theory, see [90, 25]. Extensive numerical calculations in numerous examples, however, indicate that the existence of a limiting distribution cannot be doubted. Therefore this question will not be pursued any further.

It turns out that answering the question for the precise form of $P(s)$ allows for a classification of quantum systems according to the properties of their respective classical limits. Going further on one can also study medium- and long-range correlations in the spectra and try to extend this classification to the spectral rigidity $\Delta_3(L;x)$. Berry developed in [20] a semiclassical theory of the spectral rigidity based on periodic-orbit theory (see also Berry's contribution in [41]). He found a universal behaviour of $\Delta_3(L;x)$ for $1 \ll L \ll L_{max}$; L_{max} is proportional to $\frac{1}{T_{min}}$, where T_{min} is the period of the shortest periodic orbit of the classical system. (For billiard systems $T_{min} = \frac{l_{min}}{2p}$, so that $L_{max} \propto \sqrt{x}$.) In the range $1 \ll L \ll L_{max}$ classically integrable systems share a rigidity of $\Delta_3(L;x) = \frac{2}{15}L$, whereas for classically chaotic systems possessing a time-reversal invariance Berry found that $\Delta_3(L;x) = \frac{1}{\pi^2}\log L + const.$. For very large values of L, $L \gg L_{max}$, he obtained for both cases a saturation of the rigidity at a value that is determined by the contributions of short periodic orbits to $d_{fl}(x)$. The energy dependence of the saturation value $\Delta_\infty(x)$ now is again characteristic for the classical limit of the quantum system. If this is integrable, Berry can show that $\Delta_\infty(x) \sim const.\sqrt{x}$, $x \to \infty$. Chaotic time-reversal invariant systems, however, are demonstrated to yield $\Delta_\infty(x) \sim \frac{1}{2\pi^2}\log x$, $x \to \infty$. According to (2.24) $\Delta_\infty(x)$ is just the mean square $< N_{fl}(x)^2 >$ of the fluctuation part of the spectral staircase. Hence Berry found, in the semiclassical limit, a universal behaviour for this quantity that only depends on whether the classical limit of the considered system is integrable or chaotic. It should be mentioned that the characterization of systems as being *time-reversal invariant* means that generically the lengths of periodic orbits are twofold degenerate, i.e. the multiplicities $g(l_\gamma)$ of lengths should asymptotically, for $l_\gamma \to \infty$, approach two. There are, however, classically completely chaotic time-reversal invariant systems with exponentially increasing multiplicities of lengths to which Berry's results do not apply. These systems are subsumed under the notion of *arithmetical chaos* and are the main object of the present investigation.

A question that might immediately arise is whether or not this universality of the spectral properties of quantum systems carries over to short-range correlations, i.e. to the level spacings distributions $P(s)$. Berry and Tabor demonstrated [24] that generic classically integrable systems show a Poissonian level spacing, $P(s) = e^{-s}$. Thus their quantum energy spectra are close to totally uncorrelated ones, since the short-, medium-, and long-range correlations are Poissonian. The only difference occurs with the saturation of $\Delta_3(L;x)$ for very large L, $L \gg L_{max}$. Concerning chaotic systems, however, there do not exist theoretical results for the level spacings. But it is generally believed that for these systems $P(s)$ may be well described by the distribution of spacings of eigenvalues in an ensemble of large random matrices. In *random matrix theory* (RMT) the statistical properties of the eigenvalues of random matrices from several ensembles have been thoroughly studied, see e.g. [70, 35, 45] and Bohigas's contribution in [41]. For time-reversal invariant chaotic systems the appropriate ensemble is the Gaussian orthogonal ensemble (GOE) of real symmetric matrices, whose level spacings distribution may be well approximated by Wigner's surmise $P(s) = \frac{\pi}{2}s\,e^{-\pi s^2/4}$. Dropping the requirement of time-reversal invariance one has to consider the Gaussian unitary ensemble (GUE) of complex hermitian matrices. Historically, the conjecture that quantum spectra may be described by RMT goes back to Wigner and Landau and Smorodinsky, who applied this to the resonance levels of complicated atomic nuclei, see [77] for a collection of the original contributions. Bohigas, Giannoni and Schmit extended the conjecture of an RMT-behaviour of quantum energy spectra to systems with only a few degrees of freedom possessing chaotic classical limits [30], supporting this by detailed numerical studies for several systems. In addition, the spectral rigidity $\Delta_3(L;x)$ for $1 \ll L \ll L_{max}$ found by Berry agrees with the RMT–prediction. Again, as in the integrable case, the difference lies in the saturation of the rigidity of the actual spectra for $L \to \infty$, as described by Berry's periodic-orbit theory.

In summary, the quantum spectral properties of classically integrable and classically chaotic systems differ in that for small level spacings, $s \to 0$, the former ones show $P(s) \sim 1-s$, whereas the latter ones behave as $P(s) \sim \frac{\pi}{2}s$. For classically integrable systems hence the phenomenon of *level attraction* occurs, whereas classically chaotic systems excel by a *level repulsion*. On larger scales L rather strong correlations, measured by the spectral rigidity $\Delta_3(L;x)$, are found in the chaotic case leading to an only logarithmically increasing rigidity. In contrast, integrable systems possess nearly uncorrelated spectra; only on very large scales correlations do occur.

Up to now several measures to study the statistical properties of quantum energy spectra have been introduced, most of them involving the fluctuating part $N_{fl}(E)$ of the spectral staircase. In sections 3.6 and 4.5 then the strengths of spectral fluctuations as measured by $N_{fl}(E)$ or $\Delta_\infty(E)$ play an important role and decide on the applicability of certain methods of inverse quantum chaology. It hence seems to be worthwhile to study $N_{fl}(E)$ more thoroughly to get hands on quantum energy spectra of classically chaotic systems. Thus it seems to be justified to devote some fraction of the following investigations of quantum energy spectra to a study of their fluctuations, and to the question how these are being influenced by the properties of the classical limits of the considered systems.

3 Classical Aspects of Arithmetical Chaos

In this chapter the classical dynamics of the systems subsumed as showing *arithmetical chaos* shall be introduced. In order to make the whole presentation self-contained, the first section of this chapter is devoted to a review of important definitions and results in two dimensional hyperbolic geometry. The second section then contains a discussion of geodesic length spectra of hyperbolic surfaces and introduces some useful relations among quantities referring to length spectra. Then, in the following two sections, arithmetic Fuchsian groups are defined and their geodesic length spectra are studied, leading to the final result on the exponential growth of the mean multiplicities of lengths. The results of sections 3.2 to 3.4 have previously been published in [33]. The next section then is devoted to a discussion of pseudosymmetries on arithmetic surfaces. Finally, in section 3.6, statistical properties of length spectra are studied.

3.1 A Brief Review of Hyperbolic Geometry

The classical dynamical systems to be considered from now on are geodesic flows on hyperbolic surfaces of finite area. In physics terms they are given by single point-particles of mass m moving on surfaces of constant negative Gaussian curvature without being subject to any external force. The absence of a potential causes the Jacobi metric to be proportional to the hyperbolic metric the surface has been endowed with. Thus the classical trajectories of a particle on such a surface are the geodesics of the hyperbolic metric. In the following the geometric setting shall be reviewed in a rather sketchy manner in order to recall the necessary notions and to introduce the notations used further on. A rather extensive review of the physical and mathematical aspects of the problem may be found in [17].

A convenient model for hyperbolic geometry in two dimensions is the upper complex half-plane $\mathcal{H} = \{z = x + iy;\ y > 0\}$ endowed with the Poincaré metric $ds^2 = y^{-2}(dx^2 + dy^2)$. In this setting the Gaussian curvature is $K = -1$ everywhere on $\mathcal{H}$. This is dimensionless since the internal length scale R has been normalized to $R = 1$. Otherwise the metric would have to be replaced by $ds^2 \to R^2\, ds^2$. The hyperbolic distance $d(z,w)$ of two points $z, w \in \mathcal{H}$ is the infimum of the set of lengths of curves connecting z and w, measured with ds^2. As $\mathcal{H}$ endowed with ds^2 is geodesically complete the infimum is attained for the geodesic segment connecting z and w. Its length is given by

$$\cosh d(z,w) = 1 + \frac{|z-w|^2}{2\, Im\, z\; Im\, w}\ . \tag{3.1}$$

There exists an operation of the group $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm\mathbb{1}\}$ on $\mathcal{H}$ by fractional linear transformations. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL(2,\mathbb{R})$, and $z \in \mathcal{H}$, then

$$\gamma z = \frac{az+b}{cz+d}\ . \tag{3.2}$$

This operation is transitive and compatible with the group structure of $SL(2,\mathbb{R})$. The stabilizing subgroup of $z_0 = i \in \mathcal{H}$ is the maximal compact subgroup $K = SO(2,\mathbb{R}) \subset SL(2,\mathbb{R})$. Thus the hyperbolic plane may also be realized as the symmetric space $\mathcal{H} \cong SL(2,\mathbb{R})/SO(2,\mathbb{R})$. Since both matrices γ and $-\gamma$ from $SL(2,\mathbb{R})$ map a $z \in \mathcal{H}$ to the same image it is the projective group $PSL(2,\mathbb{R})$, where the centre $\{\pm\mathbb{1}\}$ of $SL(2,\mathbb{R})$ has been factored out, that effectively operates on $\mathcal{H}$. From now on the distinction between matrices $\gamma \in SL(2,\mathbb{R})$ and classes $[\gamma] \in PSL(2,\mathbb{R})$ will be dropped and an identification of γ and $-\gamma$ will be understood

automatically. Matrices $\gamma \in SL(2,\mathbb{R})$ will be chosen such that $tr\,\gamma \geq 0$. This somewhat sloppy notation seems to be convenient and should not cause confusion.

$PSL(2,\mathbb{R})$ operating on $\mathcal{H}$ turns out to be the group of orientation preserving isometries of the Riemannian space $(\mathcal{H}, ds^2)$, i.e. $\gamma:\ \mathcal{H} \to \mathcal{H}$ preserves the lengths of curves on $\mathcal{H}$. Three classes of elements of $PSL(2,\mathbb{R})$ have to be distinguished according to their traces ($\gamma \neq \mathbf{1}$):

1. *Elliptic elements* γ have $0 \leq tr\,\gamma < 2$. Such a γ has one fixed point in the interior of $\mathcal{H}$. It is conjugate within $SL(2,\mathbb{R})$ to an element $\left(\begin{smallmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{smallmatrix}\right)$, $\vartheta \in [0, 2\pi)$.

2. *Parabolic elements* γ have $tr\,\gamma = 2$. Such a γ has one fixed point on the boundary $\partial\mathcal{H}$ of $\mathcal{H}$ ($\partial\mathcal{H} = \mathbb{R} \cup \{i\infty\}$). It is conjugate within $SL(2,\mathbb{R})$ to an element $\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$, $x \in \mathbb{R}$.

3. *Hyperbolic elements* γ have $tr\,\gamma > 2$. Such a γ has two fixed points on $\partial\mathcal{H}$. It is conjugate within $SL(2,\mathbb{R})$ to an element $\begin{pmatrix} N^{\frac{1}{2}} & 0 \\ 0 & N^{-\frac{1}{2}} \end{pmatrix}$. $N > 1$ is called the *norm* of γ.

The geodesics of the Poincaré metric on $\mathcal{H}$ are the half-circles and the straight half-lines perpendicular to the real axis. The two end points of a geodesic are the fixed points of a (unique) hyperbolic $\gamma \in PSL(2,\mathbb{R})$. This γ maps the geodesic onto itself, which is therefore also called the *invariant geodesic* of γ.

A discrete subgroup $\Gamma \subset PSL(2,\mathbb{R})$ gives rise to the orbit space $\Gamma\backslash\mathcal{H} = \{\Gamma z;\ z \in \mathcal{H}\}$. This means that an equivalence relation on $\mathcal{H}$ is introduced by an identification of points that are related by a Γ-transformation. As the Poincaré metric is Γ-invariant it induces a metric on the orbit space $\Gamma\backslash\mathcal{H}$, which will also be called Poincaré (or hyperbolic) metric. If Γ contains no elliptic elements, i.e. no $\gamma \in \Gamma$ has a fixed point on $\mathcal{H}$, $\Gamma\backslash\mathcal{H}$ is a regular surface. If, however, elliptic elements are present in Γ, $\Gamma\backslash\mathcal{H}$ will be a regular surface outside the respective fixed points. Including these turns $\Gamma\backslash\mathcal{H}$ into an *orbifold*. Despite this slight complication any $\Gamma\backslash\mathcal{H}$, irrespective of a possible existence of orbifold-points, will be called a *hyperbolic surface*. If $\Gamma\backslash\mathcal{H}$ has finite area (measured with the Poincaré metric), Γ is called a *Fuchsian group of the first kind*, otherwise a *Fuchsian group of the second kind*. The latter ones will be excluded from the further investigations. If a Fuchsian group of the first kind contains no parabolic elements, $\Gamma\backslash\mathcal{H}$ is compact. Therefore such a Γ is called *cocompact*.

The abstract orbit space $\Gamma\backslash\mathcal{H}$ may be realized explicitly as a *fundamental domain* $\mathcal{F}$ of Γ on $\mathcal{H}$. This is a simply connected region $\mathcal{F} \subset \mathcal{H}$ such that the interior of $\mathcal{F}$ contains no two Γ-equivalent points and the union of all Γ-translates of $\mathcal{F}$ covers all of $\mathcal{H}$. The boundary $\partial\mathcal{F}$ of $\mathcal{F}$ consists of segments that have to be identified under the operation of Γ appropriately in order to yield a closed surface. If Γ is *strictly hyperbolic*, i.e. it contains besides the identity only hyperbolic elements, $\Gamma\backslash\mathcal{H}$ is a compact surface of genus $g \geq 2$. Thus it is topologically equivalent to a sphere with g handles. It is then always possible to find a fundamental domain $\mathcal{F}$ that is bordered by $4g$ geodesic segments (a $4g$-gon). Since the Gaussian curvature is $K = -1$ the Gauß-Bonnet theorem yields for the area of $\Gamma\backslash\mathcal{H}$ area$(\Gamma\backslash\mathcal{H})$ = area$(\mathcal{F}) = 4\pi(g-1)$. If Γ is not cocompact the Γ-conjugacy classes of parabolic elements are called *cusps*. A fundamental domain then extends to infinity (i.e. to $\partial\mathcal{H}$). For each cusp $\partial\mathcal{F}$ contains a point of $\partial\mathcal{H} = \mathbb{R} \cup \{i\infty\}$.

A convenient presentation of a Fuchsian group Γ may be given in terms of generators related to a given fundamental domain $\mathcal{F}$, and relations among them. As generators one can take those transformations that identify pairs of geodesic segments of $\partial\mathcal{F}$, together with their inverses. If γ is strictly hyperbolic, $\partial\mathcal{F}$ has $4g$, $g \geq 2$, components. Thus Γ has $4g$ generators which are conventionally denoted as $a_1, b_1, \ldots, a_g, b_g, a_1^{-1}, \ldots, b_g^{-1}$. One frequently chooses the order of the generators such that the one relation they obey is

$$a_1 b_1 a_1^{-1} b_1^{-1} \ldots a_g b_g a_g^{-1} b_g^{-1} = \mathbf{1} \,. \tag{3.3}$$

This presentation seems to be well suited for explicit calculations since the group may be constructed explicitly in terms of $SL(2,\mathbb{R})$-matrices once a fundamental domain $\mathcal{F}$ is given. (See e.g. [12] and appendix B for examples.) The procedure will then be as follows: draw an arbitrary geodesic $4g$-gon ($g \geq 2$) with area($\mathcal{F}$) $= 4\pi(g-1)$. It only has to share the further property that the geodesic segments have to come in pairs of identical lengths. Construct the $2g$ matrices (from $SL(2,\mathbb{R})$) that identify the pairs of edges. Together with the $2g$ inverse matrices $4g$ generators for the desired Fuchsian group Γ have been found. Γ itself can then be obtained by forming all possible products of the generator matrices. (These are also called *words* in the generators.) In general the $4g$ generators will be "arbitrary" matrices, which means that it does not seem possible to obtain an explicit law that determines their matrix entries. Forming words in the generators will also yield matrices with seemingly arbitrary entries. In general there will thus not exist some (algebraic) structure for the matrix entries of the group elements. A consequence of this fact for practical purposes hence is that there seems to be no explicit enumeration scheme for the group elements other than the generator method just described.

This observation is indeed true for "generic" Fuchsian groups. An exception to this rule, however, does exit. This is formed by the class of *arithmetic Fuchsian groups*. For these the arithmetic structure appearing in their definition determines their matrix entries. The basic and most prominent example for an arithmetic Fuchsian group is the *modular group* $SL(2,\mathbb{Z})$. It consists of all 2×2 matrices with rational integers as entries and being of unit determinant. Leaving aside the determinantal condition the determination of the matrix entries could not be easier. For general arithmetic groups the definition is a bit more involved and will therefore be postponed to section 3.3. But still, the characterization of the set of traces of group matrices is rather simple compared to the generic case.

For applications in the context of quantum chaos the knowledge of the set of traces of elements of a Fuchsian group is of major importance, since it is tightly connected to the knowledge of the geodesic length spectrum. The latter in turn is the input on the classical side of periodic-orbit sum rules. Since the set out of which the traces of an arithmetic Fuchsian group have to be taken is known (see section 3.4), it is possible to develop an algorithm that allows to calculate the geodesic length spectrum (including multiplicities) for an arithmetic group up to a certain length completely, see [4, 81, 68, 73]. For non-arithmetic groups there seems to be no better way to calculate the length spectrum numerically other than forming words in the group generators up to a given number of generators per word. This method, however, does not yield the length spectrum completely up to some value of the length.

Because their importance for periodic-orbit theory the following section will be devoted to a discussion of geodesic length spectra connected with Fuchsian groups.

3.2 Geodesic Length Spectra of Hyperbolic Surfaces

By the very construction of hyperbolic surfaces $\Gamma\backslash\mathcal{H}$, where Γ is a Fuchsian group of the first kind, the hyperbolic plane $\mathcal{H}$ is the universal covering space of $\Gamma\backslash\mathcal{H}$. A closed geodesic c on $\Gamma\backslash\mathcal{H}$ lifts to a geodesic $\hat{c}$ on $\mathcal{H}$, which is invariant under some hyperbolic $\gamma_c \in \Gamma$. For any $\gamma \in \Gamma$ the transformation $\gamma\gamma_c\gamma^{-1}$ has some invariant geodesic $\hat{c}_\gamma$ on $\mathcal{H}$ that projects down to the same closed geodesic c on $\Gamma\backslash\mathcal{H}$. There exists in fact a one-to-one correspondence between

Γ-conjugacy classes of hyperbolic elements of Γ and closed geodesics on $\Gamma\backslash\mathcal{H}$. The hyperbolic length l of c can be related to γ_c as follows: the lift $\check{c}$ of c (being a segment of $\hat{c}$) on $\mathcal{H}$ connects two points z and z' on $\mathcal{H}$ which are Γ-equivalent. Thus there exists a $\gamma \in \Gamma$ such that $z' = \gamma z$. Then $l = d(z,z') = d(z,\gamma z)$. Since $\check{c}$ is a geodesic segment $d(z,\gamma z)$ minimizes the lengths of all curves on $\mathcal{H}$ connecting z and γz. Varying z continuously along $\hat{c}$ cannot change the choice of γ. Since all these z's and γz's lie on $\hat{c}$, this has to be the invariant geodesic of γ, and hence $\gamma = \gamma_c$. To compute $l = d(z,\gamma_c z)$ one can conjugate γ_c in $PSL(2,\mathbb{R})$ to obtain the diagonal matrix $\gamma' = \begin{pmatrix} N^{\frac{1}{2}} & 0 \\ 0 & N^{-\frac{1}{2}} \end{pmatrix} = g\gamma_c g^{-1}$, for some $g \in SL(2,\mathbb{R})$ and $N > 1$. Then $l = d(z,\gamma_c z) = d(z, g^{-1}\gamma' g z) = d(gz, \gamma' gz)$. The invariant geodesic of γ' is obviously the imaginary axis, and choosing gz to be on it one obtains from (3.1) that $\cosh l = 1 + \frac{1}{2N}(N-1)^2$. This finally leads to

$$2\,\cosh\frac{l}{2} = tr\,\gamma' = tr\,\gamma \,. \tag{3.4}$$

Because of the one-to-one correspondence between conjugacy classes of hyperbolic elements $\gamma \in \Gamma$ and closed geodesics c on $\Gamma\backslash\mathcal{H}$ we denote the length l of c by $l = l(\gamma)$, where γ is some representative of the conjugacy class $\{\gamma\}_\Gamma = \{\hat{\gamma}\gamma\hat{\gamma}^{-1};\ \hat{\gamma} \in \Gamma\}$ related to c.

The set of lengths of closed geodesics on $\Gamma\backslash\mathcal{H}$ is called the *geodesic length spectrum* $\mathcal{L}(\Gamma)$ of $\Gamma\backslash\mathcal{H}$, or briefly, of Γ,

$$\mathcal{L}(\Gamma) = \{l(\gamma);\ \gamma \in \Gamma \text{ hyperbolic}\} \,. \tag{3.5}$$

Introducing the notation $\mathcal{L}(\Gamma) = \{l_1 < l_2 < l_3 \ldots\}$ the counting function $\hat{\mathcal{N}}(l)$ for the geodesic length spectrum is

$$\hat{\mathcal{N}}(l) := \#\{n;\ l_n \leq l\} \,. \tag{3.6}$$

In general a length spectrum of a hyperbolic surface will be degenerate, i.e. there exist several closed geodesics on the surface of the same length. If e.g. Γ is strictly hyperbolic, no $\gamma \in \Gamma$ different from the identity is conjugate to its inverse, so that every length at least occurs twice. Interpreted in physics terms this means that the dynamical system possesses a time-reversal symmetry, since the geodesic corresponding to the inverse of a hyperbolic transformation is the original classical trajectory traversed backwards in time. In general the multiplicity of any $l \in \mathcal{L}(\Gamma)$ will be denoted by $g(l) \in \mathbb{N}$. The counting function including multiplicities is the *classical staircase function*

$$\mathcal{N}(l) := \#\{\{\gamma\}_\Gamma;\ \gamma \in \Gamma \text{ hyperbolic and } l(\gamma) \leq l\} \,. \tag{3.7}$$

Asymptotically, for $l \to \infty$, the magnitude of $\mathcal{N}(l)$ is universally determined by *Huber's law* [58, 51], also known as the *prime geodesic theorem* (PGT),

$$\mathcal{N}(l) \sim Ei(l) \sim \frac{e^l}{l}\,, \quad l \to \infty \,. \tag{3.8}$$

The remainder to this asymptotic relation will in more detail be dealt with in section 3.6. $Ei(l)$ is related to the logarithmic integral $li(x) = P\int_0^x \frac{dt}{\log t}$ by $Ei(\log x) = li(x)$, see e.g. [39]. ($P\int dt \ldots$ thereby denotes the principal value of the integral.) Comparing (3.8) with (1.5) yields that the topological entropy for geodesic flows on hyperbolic surfaces is $\tau = 1$.

A closed geodesic is called *primitive*, if it is not a multiple traversal (≥ 2) of some other closed geodesic on the same surface. The corresponding conjugacy class $\{\gamma\}_\Gamma$ is then also primitive, i.e. a $\gamma' \in \{\gamma\}_\Gamma$ is not a power (≥ 2) of some other (hyperbolic) element of Γ. The primitive closed geodesics give rise to the *primitive geodesic length spectrum* $\mathcal{L}_p(l) = \{l_{p,1} < l_{p,2} < l_{p,3} < \ldots\}$. All quantities defined for the full length spectrum can be introduced analogously for the primitive one as well. To distinguish the respective quantities one introduces an index p in all notations referring to the primitive length spectrum. The PGT (3.8) remains true also for $\mathcal{N}_p(l)$. Since

$$\hat{\mathcal{N}}(l) = \sum_{l_n \le l} 1 = \sum_{r=1}^{[l/l_1]} \sum_{r l_{p,n} \le l} 1 = \sum_{r=1}^{[l/l_1]} \hat{\mathcal{N}}_p(l/r) \,, \tag{3.9}$$

and $\hat{\mathcal{N}}_p(l)$ is positive and monotonically increasing, the counting functions for $\mathcal{L}(\Gamma)$ and $\mathcal{L}_p(\Gamma)$, respectively, show the same asymptotical behaviour,

$$\hat{\mathcal{N}}(l) \sim \hat{\mathcal{N}}_p(l) \,, \quad l \to \infty \,. \tag{3.10}$$

In the following sections the multiplicities $g_p(l)$ of lengths of primitive closed geodesics will play a major role. As it seems to be impossible to determine the multiplicities explicitly, the only quantity one can lay hands upon appears to be an average $< g_p(l) >$, and its asymptotic behaviour for $l \to \infty$. In number theory [64] two functions $f(n)$ and $h(n)$, $n \in \mathbb{N}$, are said to be of the same *average order*, if

$$\sum_{n=1}^{N} f(n) \sim \sum_{n=1}^{N} h(n) \,, \quad N \to \infty \,. \tag{3.11}$$

Led by this definition a mean multiplicity $< g_p(l) >$ is introduced as a continuous function of l that is of the average order of $g_p(l)$,

$$\mathcal{N}_p(l) = \sum_{l_{p,n} \le l} g_p(l_{p,n}) \overset{!}{\sim} \sum_{l_{p,n} \le l} < g_p(l_{p,n}) >= \int_0^l d\hat{\mathcal{N}}_p(l') \; < g_p(l') > \,, \tag{3.12}$$

for $l \to \infty$. Substituting the multiplicities by their mean hence does not violate the PGT. Since $\mathcal{N}_p(l) \sim Ei(l)$, $l \to \infty$, one observes the asymptotic relation

$$< g_p(l) > \sim \frac{e^l}{l} \cdot \left[\frac{d\hat{\mathcal{N}}_p}{dl} \right]^{-1} \,, \quad l \to \infty \,, \tag{3.13}$$

by differentiating (3.12). This result may be interpreted in more visual terms as follows. $\mathcal{N}_p(l)$ is a step function with step-width $\Delta l_n := l_{p,n} - l_{p,n-1}$ and step-height $g_p(l_{p,n})$ at $l = l_{p,n}$. A mean of the staircase $\mathcal{N}_p(l)$ should then have a slope of $< g_p(l) > / < \Delta l >$. As $< \Delta l >^{-1}$ is the mean density of primitive lengths it is asymptotically given by $\frac{d\hat{\mathcal{N}}_p}{dl}$. Thus $\frac{d\mathcal{N}_p}{dl} \sim < g_p(l) > \frac{d\hat{\mathcal{N}}_p}{dl}$, which is equivalent to (3.13). From this relation one concludes that one has to gain information on the asymptotics of the counting function $\hat{\mathcal{N}}_p(l)$ of different lengths of primitive closed geodesics in order to be able to determine the asymptotic behaviour of the mean multiplicity $< g_p(l) >$. Notice that the same procedure may also be carried through for the full length spectrum instead of the primitive one. Since both staircase functions $\mathcal{N}(l)$ and $\hat{\mathcal{N}}(l)$ share the same asymptotic behaviour as $\mathcal{N}_p(l)$ and $\hat{\mathcal{N}}_p(l)$, respectively, the mean multiplicities agree asymptotically.

Next we are interested in relating the counting functions $\hat{\mathcal{N}}_p^{(1)}(l)$ and $\hat{\mathcal{N}}_p^{(2)}(l)$ for $l \to \infty$. Thereby $\hat{\mathcal{N}}_p^{(i)}(l)$ shall correspond to the length spectra $\mathcal{L}_p(\Gamma_i)$, where Γ_1 is a subgroup of index $d > 1$ in the Fuchsian group of the first kind Γ_2. This means that Γ_2 decomposes into d cosets of Γ_1 according to

$$\Gamma_2 = \Gamma_1 \,\dot{\cup}\, \Gamma_1\gamma_1 \,\dot{\cup} \ldots \dot{\cup}\, \Gamma_1\gamma_{d-1} \,, \tag{3.14}$$

where $\gamma_1, \ldots, \gamma_{d-1}$ are elements of Γ_2, but not of Γ_1. Let $\mathcal{L}_p(\Gamma_2)$ be given as

$$\mathcal{L}_p(\Gamma_2) = \{ l_{p,1} < l_{p,2} < l_{p,3} < \ldots \} \,; \tag{3.15}$$

$\mathcal{L}_p(\Gamma_1)$ shall now be determined in terms of the $l_{p,n}$'s. To this end one observes that if $\gamma \in \Gamma_2$, then there exists a $k \in \mathbb{Z}$, such that $\gamma^k \in \Gamma_1$. To see this take an arbitrary $\gamma \in \Gamma_2$ and form $\cup_{m \in \mathbb{Z}} \Gamma_1 \gamma^m \subset \Gamma_2$. The union on the l.h.s. cannot be disjoint, since according to (3.14) Γ_1 is of finite index in Γ_2. Therefore there exists a $\gamma_0 \in \Gamma_1\gamma^r \cap \Gamma_1\gamma^s$ for some pair $r \ne s$. Thus γ^{r-s} and γ^{s-r} lie in Γ_1. Choosing $k = |r - s|$ proves the assertion.

Let now γ_p be a primitive hyperbolic element of Γ_2, hence $l(\gamma_p) \in \mathcal{L}_p(\Gamma_2)$. Then either i) $\gamma_p \in \Gamma_1$: γ_p is also primitive hyperbolic in Γ_1 and thus $l(\gamma_p) \in \mathcal{L}_p(\Gamma_1)$, or ii) $\gamma_p \notin \Gamma_1$. Then, by the above remark, there exists a $k \in \mathbb{N}$ with $\gamma_p^k \in \Gamma_1$. The minimal such k takes care for $l(\gamma_p^k) = kl(\gamma_p)$ to be an element of $\mathcal{L}_p(\Gamma_1)$. Altogether, the primitive length spectrum of Γ_1 can be characterized as

$$\mathcal{L}_p(\Gamma_1) = \{ k_1 l_{p,1}, \; k_2 l_{p,2}, \; k_3 l_{p,3}, \ldots \} \tag{3.16}$$

with positive integers k_j. Since the k_j's may take arbitrary values in $\mathbb{N}$, this enumeration of elements of $\mathcal{L}_p(\Gamma_1)$ will in general not be an ordered one.

The determination of the asymptotic behaviour of $\hat{\mathcal{N}}_p^{(1)}(l)$ in terms of that of $\hat{\mathcal{N}}_p^{(2)}(l)$ requires to know how often (asymptotically) a certain value of the k_j's in (3.16) occurs. This question may be answered with the help of the decomposition (3.14) of Γ_2 into cosets of Γ_1. On average, a fraction of $\frac{1}{d}$ of the elements of Γ_2 are also elements of Γ_1. Thus going through the conjugacy classes of hyperbolic elements of Γ_2 and picking one representative from each class yields a fraction of $\frac{1}{d}$ of the latter ones to lie in Γ_1. This statement is meant in the following sense: one chooses an ordering of the conjugacy classes according to the corresponding lengths l of closed geodesics (equivalently, in ascending order of their traces). Then, for $l \to \infty$, the respective fraction of elements of Γ_1 approaches $\frac{1}{d}$. Hence a fraction of $\frac{1}{d}$ of the k_j's equals one. Proceeding further in the same manner yields a fraction of $(1 - \frac{1}{d})\frac{1}{d}$ to be two, and so on.

By (3.15) and (3.16) obviously $\hat{\mathcal{N}}_p^{(1)}(l) \le \hat{\mathcal{N}}_p^{(2)}(l)$. The above remark on the fraction of $k_j = 1$ yields $\hat{\mathcal{N}}_p^{(1)}(l) \ge \frac{1}{d}\hat{\mathcal{N}}_p^{(2)}(l)$ in the limit $l \to \infty$. This relation has to be an inequality rather than an equality because there might occur some $k_j \ge 2$ with $k_j l_{p,j} \le l$. The number of $k_j = k \ge 2$ such that $kl_{p,j} \le l$ is, however, bounded by $\hat{\mathcal{N}}_p^{(2)}(\frac{l}{k})$. But this is asymptotically dominated by the contribution coming from $k = 1$. One therefore observes the asymptotic relation

$$\hat{\mathcal{N}}_p^{(1)}(l) \sim \frac{1}{d} \hat{\mathcal{N}}_p^{(2)}(l) \,, \quad l \to \infty \,. \tag{3.17}$$

Again, a corresponding relation also holds for the counting functions $\hat{\mathcal{N}}^{(1)}(l)$ and $\hat{\mathcal{N}}^{(2)}(l)$ of the full length spectra.

Finally, the asymptotic relation of the counting functions for length spectra of two commensurable Fuchsian groups shall be dealt with, since this will be needed later when treating arithmetic Fuchsian groups. One recalls that two subgroups H_1 and H_2 of a group G are called *commensurable*, if the intersection $H_1 \cap H_2$ is a subgroup of finite index in both H_1 and H_2. Let therefore Γ_a and Γ_b be two commensurable Fuchsian groups of the first kind. $\Gamma_0 := \Gamma_a \cap \Gamma_b$ shall be of index $d_a < \infty$ in Γ_a and of index $d_b < \infty$ in Γ_b. By (3.17) the counting functions are related through

$$\begin{aligned} \hat{\mathcal{N}}_p^{(0)}(l) &\sim \frac{1}{d_a} \hat{\mathcal{N}}_p^{(a)}(l) \,, \\ \hat{\mathcal{N}}_p^{(0)}(l) &\sim \frac{1}{d_b} \hat{\mathcal{N}}_p^{(b)}(l) \,, \end{aligned} \tag{3.18}$$

for $l \to \infty$. From this the asymptotic relation

$$\hat{\mathcal{N}}_p^{(a)}(l) \sim \frac{d_a}{d_b} \hat{\mathcal{N}}_p^{(b)}(l) , \quad l \to \infty , \tag{3.19}$$

easily follows. Recalling the asymptotic behaviour (3.13) of the mean multiplicities of primitive lengths yields

$$< g_p^{(a)}(l) > \sim \frac{d_b}{d_a} < g_p^{(b)}(l) > , \quad l \to \infty . \tag{3.20}$$

Therefore, given two commensurable Fuchsian groups, the respective mean multiplicities of primitive lengths of closed geodesics are asymptotically, for $l \to \infty$, proportional to one another. The factor of proportionality is given by the ratio of the indices with which the two groups contain their intersection as a subgroup.

So far, the discussion of length spectra has been completely general in that it has been valid for all Fuchsian groups of the first kind. A major objective of the present investigation, however, is to stress the importance of distinguishing arithmetic Fuchsian groups from non-arithmetic ones. Concerning length spectra, the difference occurs with the behaviour of the mean multiplicities $< g_p(l) >$ as $l \to \infty$. For arithmetic groups Γ it shall be shown in section 3.4 that $< g_p(l) > \sim c_\Gamma \frac{e^{l/2}}{l}$, $l \to \infty$, where c_Γ is some constant depending on the specific group Γ. This behaviour is exceptional and constitutes a major property by which the arithmetical systems excel among general chaotic dynamical systems. It is indeed known that for general hyperbolic surfaces $g_p(l)$ is always unbounded [57, 78, 36]. The proof of this proceeds algebraically; Buser [36], however, geometrically constructs explicit examples of degenerate closed geodesics, where the multiplicities of the respective lengths cannot be accounted for by symmetries of the surface. He can estimate the multiplicity of a length l to which the construction is applied to be of the order l^a for $a = \frac{\log 2}{\log 5}$ and $l \to \infty$. Aurich [3] computed the lower parts of length spectra for several arbitrarily chosen compact surfaces of genus two and never observed a multiplicity that could not be explained by symmetries in the computed l–range. Thus it appears that high multiplicities (having values beyond the expectation based on symmetries) show up rather scarcely. All this information indicates that for non-arithmetic Fuchsian groups the multiplicities of lengths of closed geodesics do not grow exponentially. The latter seems to be characteristic of surfaces with arithmetic Fuchsian groups.

Without giving the precise definition of arithmetic Fuchsian groups at this stage we would now like to try to illustrate the difference between arithmetic and non-arithmetic groups in order to give an intuitive understanding of the different properties of their respective length spectra. As described in section 3.1 a presentation of a Fuchsian group Γ may be yielded in terms of generator matrices $g_1, \ldots, g_n$, where the g_i's define fractional linear transformations on $\mathcal{H}$ that identify pairs of edges of the fundamental domain $\mathcal{F}$ of Γ. If $\mathcal{F}$ (and thus also Γ) is arbitrary, then $g_1, \ldots, g_n$ will be matrices with "arbitrary" real entries. Consequently the set of matrix entries of all elements of Γ has no obvious algebraic structure other than that the entries are composed of those of the generator matrices. Since concerning length spectra the traces of group elements are the objects of interest, a possible algebraic structure of the set of traces $tr\,\Gamma = \{tr\,\gamma;\ \gamma \in \Gamma\}$ of elements of Γ would influence $\mathcal{L}(\Gamma)$. As an example of an arithmetic group the modular group $\Gamma_{mod} = SL(2, \mathbb{Z})$ should serve. From the theorem proved in [66] one can conclude that $tr\,\Gamma_{mod} = \mathbb{Z}$ (see also [81]). Therefore $tr\,\Gamma_{mod}$ has a nice algebraic structure in that it is just the ring of rational integers. According to (3.4) one can hence enumerate the lengths of closed geodesics on $\Gamma_{mod}\backslash\mathcal{H}$ by $n = 2\cosh(l_n/2)$, $n = 3, 4, 5, \ldots$. Since $\hat{\mathcal{N}}(l_n) = n - 2$, one finds the asymptotic law $\hat{\mathcal{N}}(l_n) \sim e^{l_n/2}$, $n \to \infty$, for the modular group. By (3.13) the

multiplicities of lengths therefore grow asymptotically as

$$< g(l) > \sim 2\,\frac{e^{l/2}}{l} , \quad l \to \infty . \tag{3.21}$$

It was now observed in [32, 27] that such an exponential increase of the multiplicities is a common feature of arithmetic groups and may serve as a characterization of them. The example of the modular group shows most clearly that this law derives from the fact that the condition on the set of traces $tr\,\Gamma_{mod}$ to be the "rigid" set $\mathbb{Z}$ forces the number $\hat{\mathcal{N}}(l)$ of different lengths to grow only like $const.\, e^{l/2}$. One can transform this condition also into one which applies directly to $tr\,\Gamma$. If $\hat{\mathcal{N}}(l) \sim c\, e^{l/2}$, $l \to \infty$, then the number of lengths in an interval $[l, l + 2e^{-l/2}]$ is asymptotically given by c. If one returns to $tr\,\Gamma$, then the above interval corresponds asymptotically to the interval $[t, t+1]$, where $t = 2\cosh(l/2) \sim e^{l/2}$. Thus the number of traces of elements of Γ in the interval $[t, t+1]$ is asymptotically given by

$$\#\{tr\,\Gamma \cap [t, t+1]\} \sim c , \quad t \to \infty . \tag{3.22}$$

This condition shows more explicitly that the condition on $tr\,\Gamma$ enforces the traces to form a "rigid" discrete subset of the real line. The condition $\#\{tr\,\Gamma \cap [t, t+1]\} \leq c(\Gamma)$ for some constant $c(\Gamma)$ has most recently been proven by Sarnak and Luo [80] to be fulfilled for all arithmetic Fuchsian groups Γ and has been called the *bounded clustering property* by them.

Our plan now is to derive (3.21) for general arithmetic groups. Since this needs a lot of techniques from number theory, at first a brief review of relevant notions and facts concerning algebraic number fields and quaternion algebras shall be given. This then allows to state the precise definition of arithmetic Fuchsian groups. Only after this has been done it will be possible to study the traces occurring for arithmetic groups in more detail.

3.3 Arithmetic Fuchsian Groups

This section contains a collection of notions and facts from number theory that are needed to state the definition of arithmetic Fuchsian groups. More background on the algebraic foundations and on number theory may be found in [55, 98, 34]. As references to find additional information on quaternion algebras and arithmetic Fuchsian groups [86, 102, 71] may be consulted.

An extension K of finite degree n of the field of rational numbers $\mathbb{Q}$ is a field that contains $\mathbb{Q}$ as a subfield and, viewed as a vector space over $\mathbb{Q}$ (in the obvious manner), is of finite dimension n. Let $\mathbb{Q}[x]$ denote the ring of polynomials in a variable x with rational coefficients. $\alpha \in K$ will be called *algebraic*, if it is a zero of some polynomial from $\mathbb{Q}[x]$. The *minimal polynomial* of α is the (unique) element in $\mathbb{Q}[x]$ of lowest degree, and with leading coefficient one, that has α as a root. The field K is called an *algebraic number field*, if every $\alpha \in K$ is algebraic. Every extension K of $\mathbb{Q}$ of finite degree is known to be algebraic.

If M is some arbitrary subset of K, $\mathbb{Q}(M)$ is defined to be the smallest subfield of K that contains both M and $\mathbb{Q}$. It is given by all values of all polynomials in the elements of M with rational coefficients, and all possible quotients thereof. $\mathbb{Q}(M)$ is called the *adjunction* of M to $\mathbb{Q}$. One can now show that every algebraic number field K of finite degree over $\mathbb{Q}$ can be realized as an adjunction of a single algebraic number $\alpha \in K$ to $\mathbb{Q}$; therefore $K = \mathbb{Q}(\alpha)$.

Since K is a vector space of dimension n over $\mathbb{Q}$, the $n+1$ algebraic numbers $1, \alpha, \ldots, \alpha^n$ have to be linearly dependent and thus have to obey a relation

$$a_n\alpha^n + \cdots + a_1\alpha + a_0 = 0 \tag{3.23}$$

with rational coefficients a_i and $a_n \neq 0$. Normalizing the leading coefficient to one leaves α as a root of an irreducible polynomial $f_\alpha(x) \in \mathbb{Q}[x]$ of degree n. $f_\alpha(x)$ is the minimal polynomial of α. Since $\{1, \alpha, \ldots, \alpha^{n-1}\}$ may serve as a basis for K over $\mathbb{Q}$, any $x \in K$ may be expanded as a linear combination of powers of α up to the order $n-1$,

$$x = b_{n-1}\alpha^{n-1} + \cdots + b_1\alpha + b_0 , \tag{3.24}$$

with rational coefficients b_i.

The polynomial $f_\alpha(x)$ has n different complex roots $\alpha_1, \ldots, \alpha_n$ ($\alpha_1 = \alpha$). One can thus define n different homomorphisms $\varphi_i : K \longrightarrow \mathbb{C}$, $i = 1, \ldots, n$, that leave $\mathbb{Q}$ invariant, by

$$\varphi_i(x) := b_{n-1}\alpha_i^{n-1} + \cdots + b_1\alpha_i + b_0 , \tag{3.25}$$

$\varphi_1(x) = x$. The φ_i's are called the *conjugations* of K. If all images of K under these homomorphisms are contained in the real numbers, in other words if $f_\alpha(x)$ has only real roots, then K is said to be *totally real*.

On $\mathbb{Q}$ the usual absolute value $\nu_1(x) = |x|$, $x \in \mathbb{Q}$, introduces a topology, which is, however, not complete. The n conjugations φ_i offer n distinct ways to embed K into $\mathbb{R}$. Thus n different absolute values ν_i are given on K by $\nu_i(x) := |\varphi_i(x)|$, and these can be used to complete K to $K_{\nu_i} \cong \mathbb{R}$. The ν_i's are also called the (archimedean) *infinite primes* of K.

All algebraic numbers in K whose minimal polynomials have coefficients in the rational integers $\mathbb{Z}$ form a ring $\mathcal{R}_K$, which is called the *ring of integers of* K. An element $x \in \mathcal{R}_K$ is also called an *algebraic integer*. In K at most n algebraic numbers can be linearly independent over $\mathbb{Q}$. (This is equivalent to saying that these numbers are linearly independent over $\mathbb{Z}$.) Suppose now that $\lambda_1, \ldots, \lambda_m$, $m \leq n$, are linearly independent numbers from K. Then all linear combinations of the λ_i's with integer coefficients form an additive abelian group. This is called a *$\mathbb{Z}$–module of rank m*. A $\mathbb{Z}$–module $o \subset K$ of the maximal possible rank n that at the same time is a subring of K is called an *order* of K. Since we understand a ring to contain a unity, every order $o \subset K$ contains the rational integers $\mathbb{Z}$. Further it is known that there exists a *maximal order* in K that contains all other orders, and that this maximal order is just the ring of algebraic integers $\mathcal{R}_K$. An order o possesses a module-basis of n algebraic numbers $\omega_1, \ldots, \omega_n$ that are linearly independent over $\mathbb{Z}$ and hence also, equivalently, over $\mathbb{Q}$,

$$o = \mathbb{Z}\omega_1 \oplus \cdots \oplus \mathbb{Z}\omega_n . \tag{3.26}$$

The *discriminant* of o is defined to be $D_{K/\mathbb{Q}}(o) := [det(\varphi_j(\omega_i))]^2 \neq 0$. In complete analogy one can also define a discriminant for any $\mathbb{Z}$–module of rank n in K.

Another important notion, to be introduced now, is that of a *quaternion algebra*. In doing so, we will mainly follow [102, 86]. An algebra A over a field K is called *central*, if K is its centre; it is said to be *simple*, if it contains no two-sided ideals besides $\{0\}$ and A itself. A quaternion algebra then is defined to be a central simple algebra A of dimension four over K. In more explicit terms A may be visualized as follows: the elements of a basis $\{1, \alpha, \beta, \gamma\}$ of A over K have to obey the relations $\gamma = \alpha\beta = -\beta\alpha$, $\alpha^2 = a$, $\beta^2 = b$; $a, b \in K\backslash\{0\}$. Any $X \in A$ may then be expanded as

$$X = x_0 + x_1\alpha + x_2\beta + x_3\alpha\beta , \tag{3.27}$$

with $x_0, \ldots, x_3 \in K$. On A there exists an involutory anti-automorphism, called the *conjugation* of A, that maps X to $\bar{X} := x_0 - x_1\alpha - x_2\beta - x_3\alpha\beta$. Thus $\bar{\bar{X}} = X$ and $\overline{X \cdot Y} = \bar{Y} \cdot \bar{X}$. The conjugation enables one to define the *reduced trace* and the *reduced norm* of A,

$$\begin{aligned} tr_A(X) &:= X + \bar{X} = 2x_0 , \\ n_A(X) &:= X \cdot \bar{X} = x_0^2 - x_1^2 a - x_2^2 b + x_3^2 ab . \end{aligned} \tag{3.28}$$

If A is a division algebra, i.e. if every $X \neq 0$ in A possesses an inverse, $n_A(X) = 0$ implies $X = 0$. The inverse is then given by $X^{-1} = \frac{1}{n_A(X)}\bar{X}$.

A $\mathbb{Z}$–module $\mathcal{O} \subset A$ of (the maximal possible) rank $4n$ that also is a subalgebra in A is called an *order* of A. The introduction of a module–basis $\{\tau_1, \ldots, \tau_{4n}\}$ turns the order into

$$\mathcal{O} = \mathbb{Z}\tau_1 \oplus \cdots \oplus \mathbb{Z}\tau_{4n} . \tag{3.29}$$

We further introduce the *group of units of norm one* $\mathcal{O}^1 := \{\varepsilon \in \mathcal{O} \,|\, \varepsilon^{-1} \in \mathcal{O},\ n_A(\varepsilon) = 1\}$.

A well-known example of a (division) quaternion algebra is given by *Hamilton's quaternions*

$$\mathbb{H} := \left\{ \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} ;\ z, w \in \mathbb{C} \right\} . \tag{3.30}$$

$\mathbb{H}$ is a four dimensional $\mathbb{R}$–subalgebra of $M(2, \mathbb{C})$, the algebra of complex 2×2– matrices, characterized by the parameters $a = b = -1$. The subgroup of elements of reduced norm one is just $SU(2, \mathbb{C})$. An even simpler example of a (non-division) quaternion algebra over $\mathbb{R}$ is $M(2, \mathbb{R})$. In fact, $\mathbb{H}$ and $M(2, \mathbb{R})$ are the only quaternion algebras over $\mathbb{R}$.

A classification of quaternion algebras over K can now be achieved by looking at the corresponding algebras over $\mathbb{R}$ with the help of the n completions $K_{\nu_i} \cong \mathbb{R}$. Define $A_i := A \otimes_{\mathbb{Q}} K_{\nu_i} \cong A \otimes_{\mathbb{Q}} \mathbb{R}$, which is a quaternion algebra over $\mathbb{R}$. Hence it is either isomorphic to $\mathbb{H}$ (if it is a division algebra), or to $M(2, \mathbb{R})$ (if it is a non-division algebra). For the definition of arithmetic Fuchsian groups (see [102, 94]) we consider the case $A_1 \cong M(2, \mathbb{R})$ and $A_i \cong \mathbb{H}$ for $i = 2, \ldots, n$. Therefore there exists an isomorphism

$$\rho :\ A \otimes_{\mathbb{Q}} \mathbb{R} \longrightarrow M(2, \mathbb{R}) \oplus \mathbb{H} \oplus \cdots \oplus \mathbb{H} , \tag{3.31}$$

where there occur $n-1$ summands of $\mathbb{H}$. ρ_j will denote the restriction of ρ to A followed by a projection onto the j-th summand in (3.31). The several reduced traces and norms for $X \in A$ in (3.31) are related by

$$\begin{aligned} tr\,\rho_1(X) &= tr_A(X) , \\ det\,\rho_1(X) &= n_A(X) , \\ tr_{\mathbb{H}}\rho_j(X) &= \varphi_j(tr_A(X)) = \varphi_j(tr\,\rho_1(X)) , \\ n_{\mathbb{H}}(\rho_j(X)) &= \varphi_j(n_A(X)) = \varphi_j(det\,\rho_1(X)) , \quad j = 2, \ldots, n . \end{aligned} \tag{3.32}$$

The image of A under ρ_1 in $M(2, \mathbb{R})$ may also be expressed in more explicit terms by using the basis $\{1, \alpha, \beta, \alpha\beta\}$ for A, see (3.27): $\rho_1(1)$ is the 2×2 unit matrix; $\rho_1(\alpha)$ and $\rho_1(\beta)$ may be represented, by using the parameters $a, b > 0$, as

$$\rho_1(\alpha) = \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix} , \quad \rho_1(\beta) = \begin{pmatrix} 0 & \sqrt{b} \\ \sqrt{b} & 0 \end{pmatrix} . \tag{3.33}$$

For $X = x_0 + x_1\alpha + x_2\beta + x_3\alpha\beta \in A$ the matrix $\rho_1(X)$ in this representation takes the form

$$\rho_1(X) = \begin{pmatrix} x_0 + x_1\sqrt{a} & x_2\sqrt{b} + x_3\sqrt{ab} \\ x_2\sqrt{b} - x_3\sqrt{ab} & x_0 - x_1\sqrt{a} \end{pmatrix} . \tag{3.34}$$

We are now seeking for a subset in A whose image under ρ_1 in $M(2, \mathbb{R})$ gives a Fuchsian group Γ. Therefore $\rho_1^{-1}(\Gamma)$ must be a discrete multiplicative subgroup of A. Furthermore, for

$\rho_1(X) = \gamma \in \Gamma$ the condition $det\,\gamma = 1$ must be fulfilled. Thus by (3.32) $n_A(X) = 1$ has to be required. Hence we are led to look at groups of units of norm one $\mathcal{O}^1$ of orders $\mathcal{O} \subset A$. Regarding their images under ρ_1 one finds in [86, 94] the following

PROPOSITION: Let A be a quaternion algebra over the totally real algebraic number field K of degree n. Let $\mathcal{O} \subset A$ be an order and $\mathcal{O}^1$ be its group of units of norm one. Then $\Gamma(A,\mathcal{O}) := \rho_1(\mathcal{O}^1)$ is a Fuchsian group of the first kind. Moreover, $\Gamma(A,\mathcal{O})\backslash\mathcal{H}$ is compact if A is a division algebra. A change of the isomorphism ρ in (3.31) amounts to a conjugation of $\Gamma(A,\mathcal{O})$ in $SL(2,\mathbb{R})$.

REMARK: This proposition can be traced back to a more general theorem of A. Weil that deals with an adelic setting, see e.g. [103, 86]. For the proposition to be true it is essential that the quaternion algebra is such that on the r.h.s. of (3.31) there appears exactly one factor of $M(2,\mathbb{R})$ and $n-1$ factors of $\mathbb{H}$. (3.31) therefore is an integral component of the definition of arithmetic Fuchsian groups.

The proposition now tells us that we have found what we were looking for: a class of arithmetically defined Fuchsian groups.

We are aiming at counting the numbers of distinct primitive lengths to gain information on the mean multiplicities in the length spectra derived from the Fuchsian groups under consideration. For this purpose (3.20) allows to enlarge the class of groups appearing in the proposition a little.

DEFINITION: A Fuchsian group Γ that is a subgroup of finite index in some $\Gamma(A,\mathcal{O})$ will be called a *Fuchsian group derived from the quaternion algebra A*. (For $\Gamma(A,\mathcal{O})$ the shorthand phrase *quaternion group* will also be sometimes used.) A Fuchsian group Γ that is commensurable with some $\Gamma(A,\mathcal{O})$ will be called an *arithmetic Fuchsian group.*

In the preceding sections the modular group Γ_{mod} always served as an example of an arithmetic Fuchsian group. It will now be seen that Γ_{mod} fits into the scheme just introduced. The number field to be considered is $K = \mathbb{Q}$, and hence $n = 1$. In $\mathbb{Q}$ the ring of integers is of course $\mathcal{R}_K = \mathbb{Z}$, which is also the only order in K. The relevant quaternion algebra is the simplest one can think of, namely the one characterized by the two parameters $a = 1 = b$. Thus A is simply the matrix algebra $M(2,\mathbb{Q})$, which clearly is not a division algebra. This is in accordance with the non-compactness of the surface $\Gamma_{mod}\backslash\mathcal{H}$. The order $\mathcal{O} \subset A$ (see (3.29)) that determines Γ_{mod} is characterized by the $\mathbb{Z}$-basis $\{\tau_1,\dots,\tau_4\}$,

$$\tau_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \tau_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.35}$$

Therefore, $\mathcal{O} = M(2,\mathbb{Z})$ and $\mathcal{O}^1 = SL(2,\mathbb{Z})$. The modular group is of course well-studied. A lot of information about this group and the spectral theory on the surface $\Gamma_{mod}\backslash\mathcal{H}$ can be found e.g. in [96].

As a second example let us introduce the regular octagon group Γ_{reg}. This is a strictly hyperbolic Fuchsian group that leads to the most symmetric compact surface $\Gamma_{reg}\backslash\mathcal{H}$ of genus two (see [17, 10, 4]). Γ_{reg} is a subgroup of index two of the quaternion group $\Gamma(A,\mathcal{O})$ that is defined over the number field $K = \mathbb{Q}(\sqrt{2})$ of degree $n = 2$. The ring of integers of this field is $\mathcal{R}_K = \mathbb{Z}[\sqrt{2}] = \{m + n\sqrt{2};\ m,n \in \mathbb{Z}\}$. The module basis $\{\omega_1,\omega_2\} = \{1,\sqrt{2}\}$ for $\mathcal{R}_K$ may

also serve as a basis for K over $\mathbb{Q}$. The only non-trivial conjugation of K is given by

$$\varphi_2(p + q\sqrt{2}) = p - q\sqrt{2}, \quad p,q \in \mathbb{Q}. \tag{3.36}$$

The quaternion algebra A necessary to define $\Gamma(A,\mathcal{O})$ is determined by the two parameters $a = 1 + \sqrt{2}$ and $b = 1$. The order $\mathcal{O} \subset A$ can be characterized by giving the $\mathbb{Z}$-basis $\{\tau_1 \dots, \tau_8\}$ (see (3.29)). In the present case this is $\{\omega_1 \cdot 1, \dots, \omega_2 \cdot \alpha\beta\}$, so that an element $\gamma = \rho_1(X)$ for $X \in \mathcal{O}$ looks like

$$\gamma = \begin{pmatrix} x_0 + x_1\sqrt{1+\sqrt{2}} & x_2 + x_3\sqrt{1+\sqrt{2}} \\ x_2 - x_3\sqrt{1+\sqrt{2}} & x_0 - x_1\sqrt{1+\sqrt{2}} \end{pmatrix}, \tag{3.37}$$

with $x_i = m_i + n_i\sqrt{2}$, $m_i, n_i \in \mathbb{Z}$. The quaternion group $\Gamma(A,\mathcal{O})$ now consists of all matrices γ of the form (3.37) with $det\,\gamma = 1$. The regular octagon group Γ_{reg} is characterized by the fact that m_0 has to be an odd integer. In [76] it is shown that by adjoining the additional (elliptic) matrix $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ to Γ_{reg} one obtains this quaternion group via $\Gamma(A,\mathcal{O}) = \Gamma_{reg} \,\dot\cup\, \Gamma_{reg}S$.

3.4 Multiplicities in Length Spectra for Arithmetic Fuchsian Groups

The problem of this section is to characterize the set of traces $tr\,\Gamma$ of elements of an arithmetic Fuchsian group Γ, and then to determine the asymptotics of the number of distinct lengths corresponding to it. It, however, suffices to concentrate on quaternion groups, since every arithmetic group is by definition commensurable to a quaternion group. Then (3.20) can be used to obtain the desired asymptotics of the mean multiplicities for the given group from that for the related quaternion group.

Let therefore $\Gamma = \Gamma(A,\mathcal{O})$ be a quaternion group over the algebraic number field K of degree n as described in the preceding section. For $X = x_0 + x_1\alpha + x_2\beta + x_3\alpha\beta \in \mathcal{O}$ denote $\rho_1(X) = \gamma \in \Gamma$. By (3.34) one sees that $\frac{1}{2}tr\,\gamma = x_0 \in \mathcal{O}|_K$. Since the coefficients x_0 of elements $X \in \mathcal{O}$ will play a major role in the following, the set of these will be given a name. Let therefore be

$$\mathcal{M} := \{x_0;\ X = x_0 + x_1\alpha + x_2\beta + x_3\alpha\beta \in \mathcal{O}\}. \tag{3.38}$$

Then

$$tr\,\Gamma = tr_A\mathcal{O}^1 \subset tr_A\mathcal{O} = 2\mathcal{M}. \tag{3.39}$$

The inclusion $tr_A\mathcal{O}^1 \subset tr_A\mathcal{O}$ will in general be a proper one and we will return to this problem later. The aim now is to determine the number $\hat{\mathcal{N}}_p(l)$ of distinct primitive lengths on $\Gamma\backslash\mathcal{H}$ for $l \to \infty$. By (3.4) one hence has to count the number of distinct traces in Γ with $2 < tr\,\gamma \le 2R$, $R := \cosh(l/2) \to \infty$.

First we want to describe the set $\mathcal{M}$ a little bit further. To this end we have to introduce some more notation. Let $\{\omega_1,\dots,\omega_n\}$ be a basis for K (as a vector space) over $\mathbb{Q}$. With the help of the basis $\{1,\alpha,\beta,\alpha\beta\}$ of A over K then $\{\chi_1,\dots,\chi_{4n}\} := \{\omega_1 \cdot 1,\dots,\omega_n \cdot \alpha\beta\}$ is a basis of A over $\mathbb{Q}$. On the other hand the module-basis $\{\tau_1,\dots,\tau_{4n}\}$ of $\mathcal{O}$ (see (3.29)) consists of $4n$ linearly independent (over $\mathbb{Z}$ as over $\mathbb{Q}$) elements of A and thus may also serve as a basis for A over $\mathbb{Q}$. The two $\mathbb{Q}$-bases of A are therefore related by

$$\tau_i = \sum_{j=1}^{4n} M_{i,j}\,\chi_j, \tag{3.40}$$

where $(M_{i,j}) \in GL(4n, \mathbb{Q})$. The order $\mathcal{O} \subset A$ then takes the form

$$\mathcal{O} = \mathbb{Z} \sum_{j=1}^{4n} M_{1,j}\chi_j \oplus \cdots \oplus \mathbb{Z} \sum_{j=1}^{4n} M_{4n,j}\chi_j \tag{3.41}$$

after inserting (3.40) into (3.29). As the centre K of A is spanned by $\{\chi_1, \ldots, \chi_n\} \cong \{\omega_1, \ldots, \omega_n\}$, it turns out that

$$\mathcal{M} = \mathbb{Z} \sum_{j=1}^{n} M_{1,j}\omega_j + \cdots + \mathbb{Z} \sum_{j=1}^{n} M_{4n,j}\omega_j \, . \tag{3.42}$$

Obviously, $\mathcal{M}$ is a $\mathbb{Z}$-module in K. Since $(M_{i,j}) \in GL(4n, \mathbb{Q})$, out of the $4n$ algebraic numbers $\sum_{j=1}^{n} M_{i,j}\omega_j$, $i = 1, \ldots, 4n$, n are linearly independent. One can therefore choose the module-basis $\{\mu_1, \ldots, \mu_n\}$ among them,

$$\mathcal{M} = \mathbb{Z}\mu_1 \oplus \cdots \oplus \mathbb{Z}\mu_n \, . \tag{3.43}$$

In general, however, $\mathcal{M}$ is not a subring and hence no order in K, because the multiplication in it need not close. But in [94] one finds that $tr\,\Gamma = 2\mathcal{M}$ is contained in the ring $\mathcal{R}_K$ of integers of K. Defining $\hat{\mu}_i := 2\mu_i$ for $i = 1, \ldots, n$ then yields

$$2\mathcal{M} = \mathbb{Z}\hat{\mu}_1 \oplus \cdots \oplus \mathbb{Z}\hat{\mu}_n \subset \mathcal{R}_K \, . \tag{3.44}$$

By (3.4) this means for the geodesic length spectrum $\mathcal{L}(\Gamma)$ that $2\cosh(l/2) \in 2\mathcal{M} \subset \mathcal{R}_K$ is an algebraic integer for every $l \in \mathcal{L}(\Gamma)$.

So far, $tr_A\mathcal{O} = 2\mathcal{M}$ has been described in an algebraic way. The object of interest, however, is $tr\,\Gamma = tr_A\mathcal{O}^1$. The problem thus is the following. Let an $x_0 \in \mathcal{M}$ be given, which means that there is at least one $X \in \mathcal{O}$ with $X = x_0 + x_1\alpha + x_2\beta + x_3\alpha\beta$. There might be, however, several $X_k \in \mathcal{O}$ all sharing the same first coefficient x_0. In a more formal way this can be described using the bases (3.41) and (3.42). According to (3.42) x_0 can be represented as $x_0 = \sum_{i=1}^{4n}\sum_{j=1}^{n} r_i M_{i,j}\omega_j$, where all $r_i \in \mathbb{Z}$. The choice of $(r_1, \ldots, r_{4n}) \in \mathbb{Z}^{4n}$ is not unique, since $\mathcal{M}$ has only rank n. Thus one can vary $(r_1, \ldots, r_{4n})$ over a certain subset of $\mathbb{Z}^{4n}$ without changing x_0. $X = \sum_{i,j=1}^{4n} r_i M_{i,j}\omega_j$, however, does change under this variation. The so resulting (discrete) set of $X \in \mathcal{O}$ then is the set of X_k's mentioned above. The question now is whether there exists some $X \in \mathcal{O}$ amongst the ones all sharing the same first coefficient x_0 that has a reduced norm $n_A(X) = 1$. If the answer is in the affirmative, then also $2x_0 \in tr_A\mathcal{O}^1 = tr\,\Gamma$. But in order to decide on this question one has to know whether there is a solution to

$$-x_1^2 a - x_2^2 b + x_3^2 ab = 1 - x_0^2 \tag{3.45}$$

for a given $x_0 \in \mathcal{M}$ in the three variables (x_1, x_2, x_3) so that $X = x_0 + x_1\alpha + x_2\beta + x_3\alpha\beta \in \mathcal{O}$. The indefinite quadratic equation (3.45) defined over the given non-trivial domain of variables appears to be difficult to deal with in full generality. In the following this subtle problem shall be avoided, and the determinantal condition $n_A(X) = 1$ for $X \in \mathcal{O}$ shall be replaced by a weaker auxiliary requirement. The price to pay for this will be that a hypothesis will have to be introduced below without which no result could be obtained.

Any $X \in \mathcal{O}^1$ is characterized within $\mathcal{O}$ by the condition $n_A(X) = 1$. By (3.31) and (3.32) this implies that $n_{\mathbb{H}}(\rho_j(X)) = \varphi_j(n_A(X)) = 1$ for $j = 2, \ldots, n$. Therefore $\rho_j(X) \in SU(2, \mathbb{C})$ and hence $tr_{\mathbb{H}}\rho_j(X) = \varphi_j(tr\,\gamma) \in [-2, +2]$ for $j = 2, \ldots, n$. We will now call

$$tr_I\Gamma := \{2x_0;\ x_0 \in \mathcal{M},\ |\varphi_j(x_0)| \le 1,\ j = 2, \ldots, n\} \tag{3.46}$$

the *idealized set of traces* of Γ. Instead of the inclusion $tr_A\mathcal{O}^1 \subseteq tr_A\mathcal{O}$ we are thus now considering $tr_I\Gamma \subseteq tr_A\mathcal{O}$. By the very construction of the idealized traces it is clear that $tr_A\mathcal{O}^1 \subseteq tr_I\Gamma$. From this one obtains the chain of inclusions

$$tr\,\Gamma = tr_A\mathcal{O}^1 \subseteq tr_I\Gamma \subseteq tr_A\mathcal{O} \, . \tag{3.47}$$

To avoid the problems of finding solutions to the quadratic equation (3.45), we are now going to count the number of idealized traces instead of the number of actual traces. Any $2x_0 \in tr_I\Gamma$ that does not occur in $tr\,\Gamma$ is referred to as a *gap* in the length spectrum corresponding to Γ. It is the number of these gaps we have to make a hypothesis on.

In order to count the number of idealized traces up to a certain value, the counting function

$$\mathcal{N}_I(R) := \frac{1}{2} \cdot \#\{x_0 \in \mathcal{M};\ |x_0| \le R,\ |\varphi_j(x_0)| \le 1,\ j = 2, \ldots, n\} \tag{3.48}$$

will be introduced with $R = \cosh(l/2)$. The factor of $\frac{1}{2}$ takes care of the overcounting by admitting both signs for x_0.

The determination of the asymptotics of $\mathcal{N}_I(R)$ for $R \to \infty$ will now be achieved by investigating the number of certain lattice points in some parallelotope. The procedure we are going to follow uses some standard receipt from algebraic number theory, see [34] and [65]. At first K is being mapped to $K_{\nu_1} \times \ldots \times K_{\nu_n} \cong \mathbb{R}^n$ by: $x \in K$, $x \mapsto \mathbf{x} = (x_1, \ldots, x_n) := (\varphi_1(x), \ldots, \varphi_n(x))$. In $\mathbb{R}^n$ we consider, given n linearly independent vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, a lattice

$$L := \mathbb{Z}\mathbf{e}_1 \oplus \cdots \oplus \mathbb{Z}\mathbf{e}_n \tag{3.49}$$

with fundamental cell

$$F := I\mathbf{e}_1 \oplus \cdots \oplus I\mathbf{e}_n \, , \tag{3.50}$$

$I := [0, 1)$. In $\mathbb{R}^n$ we shall consider usual euclidean volumes. F then has a volume of $vol(F) = det(e_{ij})$, where (e_{ij}) denotes the $n \times n$ matrix formed by the n row vectors $\mathbf{e}_j \in \mathbb{R}^n$. We further introduce the parallelotope

$$P_R := \{\mathbf{x} \in \mathbb{R}^n;\ |x_1| \le R,\ |x_j| \le 1,\ j = 2, \ldots, n\} \tag{3.51}$$

of volume $vol(P_R) = 2^n R$. In a first obvious approximation, the number $n_L(R)$ of lattice points in P_R is given by $vol(P_R)/vol(F)$. Corrections to this result are caused by contributions of the surface of P_R; this is of dimension $n-1$ (in $\mathbb{R}^n$), whereas P_R itself is of dimension n. One therefore expects the corrections to be of the order of $vol(P_R)^{(n-1)/n}$. Indeed, in [65] it is shown that

$$n_L(R) = \frac{vol(P_R)}{vol(F)} + c \cdot vol(P_R)^{1-1/n} \, , \tag{3.52}$$

with some constant c. The surface correction is therefore subdominant in the limit $R \to \infty$ and one finds that

$$n_L(R) = \frac{2^n}{det(e_{ij})} \cdot R + O(R^{1-1/n}) \, . \tag{3.53}$$

One can now construct an appropriate lattice L that allows to represent $\mathcal{N}_I(R)$ as the corresponding $\frac{1}{2}n_L(R)$. To this end one notices that the module basis $\{\mu_1, \ldots, \mu_n\}$ of $\mathcal{M} \subset K$ is being mapped to a set of n linearly independent vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, with $\mathbf{e}_j := (\varphi_1(\mu_j), \ldots, \varphi_n(\mu_j))$. The independence may be seen by $[det(e_{ij})]^2 = [det(\varphi_i(\mu_j))]^2 = D_{K/\mathbb{Q}}(\mathcal{M}) \ne 0$. One can thus use these $\mathbf{e}_j$'s to define a lattice L as in (3.49). Its fundamental cell F has volume

$vol(F) = \sqrt{D_{K/\mathbb{Q}}(\mathcal{M})}$. An element $x_0 = k_1\mu_1 + \cdots + k_n\mu_n \in \mathcal{M}$, $k_j \in \mathbb{Z}$, is being mapped to $\mathbf{x}_0 = k_1\mathbf{e}_1 + \cdots + k_n\mathbf{e}_n \in L$, and this relation is clearly bijective. One can thus embed $\mathcal{M}$ as L in $\mathbb{R}^n$, and hence (see (3.48)) $\mathcal{N}_I(R) = \frac{1}{2}n_L(R)$. Using $R = \cosh(l/2) \sim \frac{1}{2}e^{l/2}$, $l \to \infty$, and (3.53), one concludes that

$$\mathcal{N}_I(\cosh(l/2)) \sim 2^{n-2}\,[D_{K/\mathbb{Q}}(\mathcal{M})]^{-1/2}\,e^{l/2}\,, \quad l \to \infty\,. \tag{3.54}$$

As already mentioned, we are not able to pin down the exact number of gaps that might occur for a general Fuchsian group of the type $\Gamma = \Gamma(A, \mathcal{O})$. We expect, however, that the following hypothesis holds true:

HYPOTHESIS: Asymptotically, for $l \to \infty$,

$$\hat{\mathcal{N}}(l) \sim \mathcal{N}_I(\cosh(l/2))\,. \tag{3.55}$$

In other words, it is assumed that the number of gaps grows at most like $O(e^{(\frac{1}{2}-\delta)l})$, $\delta > 0$, $l \to \infty$.

We are now in a position to state our result as the

THEOREM: Let Γ be an arithmetic Fuchsian group, commensurable with the group $\Gamma(A, \mathcal{O})$ derived from the quaternion algebra A over the totally real algebraic number field K of degree n. Denote by d_1 the index of the subgroup $\Gamma_0 := \Gamma \cap \Gamma(A, \mathcal{O})$ in Γ, and by d_2 the respective index of Γ_0 in $\Gamma(A, \mathcal{O})$. Let $D_{K/\mathbb{Q}}(\mathcal{M})$ be the discriminant of the module $\mathcal{M} \subset K$ that contains $\frac{1}{2}tr\,\Gamma(A, \mathcal{O})$. Then, under the hypothesis (3.55), the number $\hat{\mathcal{N}}_p(l)$ of distinct primitive lengths on $\Gamma\backslash\mathcal{H}$ up to l grows asymptotically like

$$\hat{\mathcal{N}}_p(l) \sim 2^{n-2}\,\frac{d_1}{d_2}\,[D_{K/\mathbb{Q}}(\mathcal{M})]^{-1/2} \cdot e^{l/2}\,, \quad l \to \infty\,. \tag{3.56}$$

PROOF: Assume the validity of the hypothesis (3.55) and recall the asymptotic relation $\hat{\mathcal{N}}_p(l) \sim \hat{\mathcal{N}}(l)$, $l \to \infty$, from section 3.2. Therefore also $\hat{\mathcal{N}}_p(l) \sim \mathcal{N}_I(\cosh(l/2))$. Using (3.19) and (3.54) then leads to the assertion.

The main objective of this study was not the counting function $\hat{\mathcal{N}}_p(l)$ but rather the mean multiplicity $< g_p(l) >$. As these two quantities are asymptotically related by (3.13) one can easily obtain from the theorem the following

COROLLARY: The local average of the primitive multiplicities in the cases described in the theorem behaves asymptotically like

$$< g_p(l) > \sim 2^{3-n}\,\frac{d_2}{d_1}\,\sqrt{D_{K/\mathbb{Q}}(\mathcal{M})} \cdot \frac{e^{l/2}}{l}\,, \quad l \to \infty\,. \tag{3.57}$$

To confirm the results of the theorem and the corollary the two examples of arithmetic Fuchsian groups introduced at the end of the preceding section shall be investigated now.

Again, at first the modular group will be treated. For this $\mathcal{O} = M(2, \mathbb{Z})$ and $\mathcal{O}^1 = SL(2, \mathbb{Z})$. Expand $X \in \mathcal{O}$ into the basis (3.35), $X = k_1\tau_1 + \cdots + k_4\tau_4$, $k_i \in \mathbb{Z}$, from which one observes that $\frac{1}{2}tr_A X = x_0 = \frac{1}{2}(k_1 + k_4)$. This yields $\mathcal{M} = \frac{1}{2}\mathbb{Z}$ and $\mu_1 = \frac{1}{2}$, see (3.43). It is known [66, 81] that for the modular group $tr\,\Gamma = \mathbb{Z} = 2\mathcal{M}$ and therefore no gaps in the set of

traces occur. The discriminant of $\mathcal{M}$ now is trivially obtained, and $\sqrt{D_{\mathbb{Q}}(\mathcal{M})} = \mu_1 = \frac{1}{2}$. As the modular group Γ_{mod} is the quaternion group $\Gamma(A, \mathcal{O})$ itself, one concludes using $d_1 = d_2 = 1$,

$$\begin{aligned} \hat{\mathcal{N}}_p(l) &\sim e^{l/2}\,, \\ < g_p(l) > &\sim 2 \cdot \frac{e^{l/2}}{l}\,, \quad l \to \infty\,, \end{aligned} \tag{3.58}$$

which agrees with the previously known result (3.21).

The second example, the regular octagon group Γ_{reg}, can almost as easily been dealt with. For this one has to go back to the quaternion group $\Gamma(A, \mathcal{O})$ described by (3.37). One observes that $x_0 \in \mathcal{M} = \mathbb{Z}[\sqrt{2}]$, for which one can use the basis $\{\mu_1, \mu_2\} = \{1, \sqrt{2}\}$. With the help of the conjugation φ_2 (see (3.36)) one obtains the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ for the lattice $L \subset \mathbb{R}^2$ as $\{(1,1), (\sqrt{2}, -\sqrt{2})\}$. This allows to determine the discriminant of $\mathcal{M}$, leading to $\sqrt{D_{K/\mathbb{Q}}(\mathcal{M})} = |det(e_{ij})| = 2\sqrt{2}$. Since Γ_{reg} is a subgroup of index two in $\Gamma(A, \mathcal{O})$, one has to choose $d_1 = 1$ and $d_2 = 2$ in (3.56) and (3.57). Thus

$$\begin{aligned} \hat{\mathcal{N}}_p(l) &\sim \frac{1}{4\sqrt{2}} \cdot e^{l/2}\,, \\ < g_p(l) > &\sim 8\sqrt{2} \cdot \frac{e^{l/2}}{l}\,, \quad l \to \infty\,. \end{aligned} \tag{3.59}$$

This is exactly the result obtained in [10, 4]. In [10] the fact that $\frac{1}{2}tr\,\gamma = m + n\sqrt{2}$, $m, n \in \mathbb{Z}$, for $\gamma \in \Gamma_{reg}$ was found for the first time. The condition that $|\varphi_2(\frac{1}{2}tr\,\gamma)| = |m - n\sqrt{2}| \le 1$ was then observed empirically. In [4] the numbers $x_0 = m + n\sqrt{2}$ fulfilling $|m - n\sqrt{2}| \le 1$ were called *minimal numbers* and the necessity of this condition for the regular octagon group was shown. Also, by a numerical computation of the primitive length spectrum up to $l = 18$ it was demonstrated that gaps do exist for the regular octagon, but that their existence does not influence the numerically calculated mean multiplicity $< g_p(l) >$.

A final remark on a more constructive approach to determine the asymptotics of $< g_p(l) >$ for a given arithmetic group Γ will be added. First, one has to know the quaternion group $\Gamma(A, \mathcal{O})$ which is commensurable with Γ, and the indices d_1 and d_2 describing $\Gamma \cap \Gamma(A, \mathcal{O})$ as a subgroup in Γ and $\Gamma(A, \mathcal{O})$. In [94] it is shown that one can get the relevant number field K by adjoining $tr\,\Gamma(A, \mathcal{O})$ to $\mathbb{Q}$. Furthermore, the algebra A can be obtained as the linear span of $\Gamma(A, \mathcal{O})$ over K. Analogously, the linear span of $\Gamma(A, \mathcal{O})$ over $\mathcal{R}_K$ yields the order $\mathcal{O} \subset A$. One then has to find the module-basis $\{\tau_1, \ldots, \tau_{4n}\}$ of $\mathcal{O}$. This can be used to obtain the matrix $(M_{i,j})$ appearing in (3.40). Given this one has to identify the module $\mathcal{M}$ containing $\frac{1}{2}tr\,\Gamma(A, \mathcal{O})$ (see (3.42) and (3.43)) in order to determine its discriminant $D_{K/\mathbb{Q}}(\mathcal{M})$. One can now plug all this information into (3.57) to get the answer to the problem. This procedure may, however, be quite formal and in special cases it may be more convenient to try a direct approach to determine $< g_p(l) >$. Nevertheless, the above result is general and sometimes it will only be necessary to know that $< g_p(l) > \sim const.\,\frac{e^{l/2}}{l}$, without specifying the constant. The latter, however, is included in the expressions (4.36) and (4.50) for the model describing the statistical properties of the related quantum energy spectra. For a quantitative description a knowledge of the constants appears to be necessary.

3.5 Pseudosymmetries

Up to now those classical aspects of arithmetical chaos have been investigated that are connected with geodesic length spectra. There is a further peculiarity of the arithmetical systems, which is a property of the classical system also appearing in its quantum version. The phenomenon to be discussed in this section is the occurrence of infinitely many *pseudosymmetries* for hyperbolic surfaces $\Gamma\backslash\mathcal{H}$ with arithmetic Fuchsian groups Γ. As these are closely related to the *Hecke ring* for Γ, which will be represented on the quantum mechanical wave functions by *Hecke operators*, the pseudosymmetries somehow mediate between the classical and quantum aspects of arithmetical chaos. Before going into the details, the description of symmetries of hyperbolic surfaces will be briefly reviewed in order to get an intuitive understanding of pseudosymmetries as some generalizations of symmetries.

A *symmetry* g of a hyperbolic surface $\Gamma\backslash\mathcal{H}$ is an isometry of this surface, and therefore necessarily also an isometry of the hyperbolic plane $\mathcal{H}$. Thus g is a fractional linear transformation on $\mathcal{H}$. The corresponding matrix from $SL(2,\mathbb{R})$ will also be denoted by g. The symmetry group $\Sigma = \{\mathbb{1}, g_1, \ldots, g_{N-1}\}$ of the surface $\Gamma\backslash\mathcal{H}$ is hence a subgroup of $SL(2,\mathbb{R})$. In order that a $g \in SL(2,\mathbb{R})$ is a symmetry it has to commute with the Fuchsian group Γ. This may be seen as follows: $z \in \mathcal{H}$ is being identified with γz for all $\gamma \in \Gamma$, thus also gz with γgz. If g is a symmetry, then the identification of z with $g^{-1}\gamma gz$ has to be the same as of z with $\gamma' z$ for all $\gamma, \gamma' \in \Gamma$. This will be the case, if (and only if) $g^{-1}\Gamma g = \Gamma$. Defining the group $\Gamma' := \Gamma \cup g_1\Gamma \cup \ldots \cup g_{N-1}\Gamma$ this condition means that Γ is a normal subgroup of Γ', and $\Sigma \cong \Gamma'/\Gamma$. The fact that the surface $\Gamma\backslash\mathcal{H}$ possesses symmetries can thus be formulated in an algebraic way. The Fuchsian group Γ is a normal subgroup of some other Fuchsian group Γ' and the symmetry group Σ is the factor group Γ'/Γ. This algebraic setting then allows to deal with symmetries in the context of the Selberg trace formula, see e.g. [100, 101] and appendix B. Since Γ' is also a Fuchsian group one can construct the surface $\Gamma'\backslash\mathcal{H}$. This can be viewed as the result of a desymmetrization procedure. It is, loosely speaking, a fundamental domain for the operation of the symmetry group Σ on the surface $\Gamma\backslash\mathcal{H}$.

A certain generalization of this concept of symmetries is provided by the pseudosymmetries of arithmetic surfaces. The starting point of their discussion will be the algebraic properties satisfied by arithmetic Fuchsian groups. From this the geometric properties of the related surfaces will be studied. The algebraic side of the problem lies at the heart of the construction of arithmetic groups and may be found e.g. in [86, 100, 71]. It resulted from a generalization of Hecke's investigation [49] of automorphic forms for the modular group. This is the reason why in number theory the notion of *modular correspondences* has been introduced for the algebraic setting. To the author's knowledge, it was Sarnak who recently introduced [79] the name *pseudosymmetries* for the geometric setting related to the modular correspondences. Since the geometric construction may be easier to understand the problem intuitively, henceforth the notion of pseudosymmetries will be used throughout. It is the purpose of this section to explain the algebraic as well as the geometric structures accompanying the pseudosymmetries as explicitly as it seems possible. This, however, inevitably requires some algebraic notions that shall be introduced first.

A major role will be played in the following by the *commensurator* $\overline{\Gamma}$ of a Fuchsian group Γ. It is a subgroup of $G := GL^+(2,\mathbb{R}) = \{g \in GL(2,\mathbb{R});\ \det g > 0\}$. Then $\overline{\Gamma} := \{g \in G;\ g^{-1}\Gamma g \sim \Gamma\}$, where "$\sim$" denotes the commensurability of two subgroups of a group. A $g \in \overline{\Gamma}$ hence transforms the Fuchsian group Γ by conjugation into a group $g^{-1}\Gamma g$ that is still commensurable with the original group Γ. Thus $\Gamma'(g) := g^{-1}\Gamma g \cap \Gamma$ is a subgroup of finite index in Γ and in $g^{-1}\Gamma g$. $\overline{\Gamma}$ clearly is a group, and it obviously contains Γ as a subgroup. If g is a symmetry of the surface $\Gamma\backslash\mathcal{H}$, then $g^{-1}\Gamma g = \Gamma$, leading to $\Gamma'(g) = \Gamma$, and thus $g \in \overline{\Gamma}$. Therefore, $\overline{\Gamma}$ contains Γ' (see above) as a subgroup. We will speak of a non-trivial commensurator, if $\overline{\Gamma}$ contains Γ' as a proper subgroup. The objects of $\overline{\Gamma}$ of interest are then the cosets of $\Gamma'\backslash\overline{\Gamma}$. We

will henceforth tacitly assume to mean representatives of these cosets when speaking of the commensurator. These will give rise to pseudosymmetries.

A commensurator $\overline{\Gamma}$ is defined for any Fuchsian group Γ and no use has so far been made of the arithmeticity of the groups of interest. The difference between commensurators of arithmetic and non-arithmetic groups is clarified by a theorem of Margulis [67, 93]:

THEOREM: If Γ is an arithmetic Fuchsian group, then its commensurator $\overline{\Gamma}$ is dense in G. If Γ is non-arithmetic, then $\overline{\Gamma}$ is commensurable with Γ.

REMARK: If $\overline{\Gamma}$ is commensurable with Γ, then $\overline{\Gamma} \cap \Gamma$ is of finite index in $\overline{\Gamma}$. Now, $\overline{\Gamma} \cap \Gamma = \Gamma$, since $\Gamma \subseteq \overline{\Gamma}$. Thus Γ is a subgroup of finite index d in $\overline{\Gamma}$, $\overline{\Gamma} = \Gamma \cup \Gamma\gamma_1 \cup \ldots \cup \Gamma\gamma_{d-1}$. The set of non-trivial elements of the commensurator hence is $\{\mathbf{1}, \gamma_1, \ldots, \gamma_{d-1}\}$, which is finite, and therefore at most finitely many pseudosymmetries exist for non-arithmetic groups. Since in the arithmetic case $\overline{\Gamma}$ is dense in G infinitely many pseudosymmetries are then present. This criterion thus may serve as a characterization of arithmeticity of Fuchsian groups. Although in the non-arithmetic case finitely many pseudosymmetries might exist, to the author's knowledge no explicit example of a non-trivial pseudosymmetry is known for this case.

By the definition of the commensurator, for every $g \in \overline{\Gamma}$ the group $\Gamma'(g)$ is a subgroup of finite index n in Γ, thus

$$\Gamma = \Gamma'(g)\gamma_1 \cup \ldots \cup \Gamma'(g)\gamma_n\ , \quad \gamma_1 = \mathbf{1}\ . \tag{3.60}$$

The pseudosymmetry related to this g is then said to be of *order* n. A symmetry is in this notation a pseudosymmetry of order $n = 1$, since if $g \in \Sigma$, then $\Gamma = \Gamma'(g)$. For the following it appears to be useful not to deal with $g \in \overline{\Gamma}$, but rather with the double cosets $\Gamma g\Gamma$. By (3.60) these decompose as

$$\Gamma g\Gamma = \Gamma g\gamma_1 \cup \ldots \cup \Gamma g\gamma_n = \Gamma\alpha_1 \cup \ldots \cup \Gamma\alpha_n\ , \tag{3.61}$$

where the definition $\alpha_i := g\gamma_i$ has been used. The double cosets $\Gamma g\Gamma$ for $g \in \overline{\Gamma}$ are the quantities to construct the Hecke ring for Γ from. Its definition and discussion will be postponed to section 4.1 where the quantum aspects of arithmetical chaos will be investigated, since it directly leads to the definition of Hecke operators that are relevant for the quantum mechanical problem.

The decomposition (3.60) of Γ into cosets of $\Gamma'(g)$ now enables one to interpret the pseudosymmetry related to g geometrically. We require g to be a pseudosymmetry of order $n \geq 2$ in order to deal with a proper generalization of a symmetry. Then $\Gamma'(g)$ is a proper subgroup of Γ, and the surface $\Gamma'(g)\backslash\mathcal{H}$ is an n-sheeted covering of the original surface $\Gamma\backslash\mathcal{H}$. To illustrate the situation one can draw the following commutative diagram (see e.g. [40, 86]):

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{\ id\ } & \mathcal{H} \\ \varphi_g \downarrow & & \downarrow \varphi \\ \Gamma'(g)\backslash\mathcal{H} & \xrightarrow{\ f\ } & \Gamma\backslash\mathcal{H} \end{array}$$

φ, φ_g and f are the natural projections. For $z \in \mathcal{H}$ denote by $\Gamma_z := \{\gamma \in \Gamma; \gamma z = z\}$ the subgroup of Γ that stabilizes z; the corresponding subgroup of $\Gamma'(g)$ then is $\Gamma'_z := \Gamma_z \cap \Gamma'(g)$. Γ_z and Γ'_z consist of elliptic elements (or parabolic elements if one admits cusps as fixed points). For $z \in \mathcal{H}$ denote its image under the projection on the surface $\Gamma\backslash\mathcal{H}$ by $p = \varphi(z)$. The preimage of p under f on $\Gamma'(g)\backslash\mathcal{H}$ then consists of the $h \leq n$ points $f^{-1}(p) = \{q_1, \ldots, q_h\}$. If e_j denotes the ramification number of f over q_j, then $\sum_{j=1}^{h} e_j = n$. On $\mathcal{H}$ one then chooses points $w_1, \ldots, w_h$ such that $q_j = \varphi_g(w_j)$, i.e. $p = f(\varphi_g(w_j))$, $j = 1, \ldots, h$. Ramification numbers e_j different from one can only occur at elliptic points of $\Gamma\backslash\mathcal{H}$, since e_j is the index of Γ'_{w_j} as a subgroup of Γ_{w_j}. These two groups are non-trivial only at elliptic points. The fractional linear transformations $\{\sigma_1, \ldots, \sigma_h\}$, defined by $\sigma_j z = w_j$, mediate the mappings that interchange the sheets of the covering $\Gamma'(g)\backslash\mathcal{H} \to \Gamma\backslash\mathcal{H}$, when projected down onto the surfaces by φ_g and φ. Since the above diagram is commutative one finds for all $\gamma \in \Gamma$ that $f^{-1}(p) = f^{-1}(\varphi(z)) = f^{-1}(\varphi(\gamma z)) = \varphi_g(\gamma z)$. Thus there exists a unique index j such that $\varphi_g(\gamma z) = q_j$; hence $\varphi_g(\gamma z) = q_j = \varphi_g(w_j) = \varphi_g(\sigma_j z)$. Since γz and $\sigma_j z$ project to the same point on $\Gamma'(g)\backslash\mathcal{H}$, there is some $\delta \in \Gamma'(g)$ with $\gamma z = \delta\sigma_j z$. Therefore $\gamma^{-1}\delta\sigma_j \in \Gamma_z$, implying by inverting the l.h.s., that $\gamma \in \Gamma'(g)\sigma_j\Gamma_z$. One hence obtains that

$$\Gamma = \Gamma'(g)\sigma_1\Gamma_z \,\dot\cup\, \ldots \,\dot\cup\, \Gamma'(g)\sigma_h\Gamma_z \,. \tag{3.62}$$

It is not difficult to show that this decomposition of Γ is disjoint. Choosing z not to be an elliptic point, hence $\Gamma_z = \{\mathbb{1}\}$ and $h = n$, and comparing (3.62) with (3.60) leads to the conclusion that $\sigma_j = \gamma'_j\gamma_j$, for some $\gamma'_j \in \Gamma'(g)$, $j = 1, \ldots, n$. The transformations $\{\gamma_1, \ldots, \gamma_n\}$ appearing in (3.60) can therefore be given an interpretation as mediating the interchanging of the sheets of the covering outside branch points. If Γ is strictly hyperbolic, i.e. if $\Gamma\backslash\mathcal{H}$ is a compact surface of genus $g \geq 2$ without elliptic points, then $\Gamma_z = \{\mathbf{1}\}$ for all $z \in \mathcal{H}$ and hence all ramification numbers are one. In this case $\Gamma'(g)\backslash\mathcal{H} \to \Gamma\backslash\mathcal{H}$ is an unramified n-sheeted covering.

In conclusion one can give the following geometric picture of a non-trivial pseudosymmetry of order $n \geq 2$ (see also [79]): it leads to an n-sheeted covering of the surface $\Gamma\backslash\mathcal{H}$ that is unramified, if Γ contains no elliptic elements. Otherwise it is ramified over the elliptic fixed points of Γ. The stabilizing group Γ_z of an elliptic fixed point z then determines the ramification number at this point. In case the operation under consideration is a symmetry, it is a pseudosymmetry of order $n = 1$, and thus no non-trivial covering occurs. The generalization of symmetries to non-trivial pseudosymmetries in this picture consists in the spreading out of the coverings over the base surface. But still, by definition, these always have a finite numbers of sheets.

Another interpretation of a pseudosymmetry related to a non-trivial $g \in \overline{\Gamma}$ can be given in terms of the effect of g on the closed geodesics on $\Gamma\backslash\mathcal{H}$. The operation of g on $\mathcal{H}$ is equivalent to its operation on Γ by conjugation. Thus its effect on a closed geodesic can be described by the mapping $\gamma \mapsto g^{-1}\gamma g$ for the hyperbolic $\gamma \in \Gamma$ related to the geodesic. Since $g^{-1}\Gamma g$ is commensurable with Γ, the fraction of $\gamma \in \Gamma$ that are being mapped onto Γ is given by the index of $\Gamma'(g) = g^{-1}\Gamma g \cap \Gamma$ in Γ. By (3.60) this is just the order n of the pseudosymmetry. Thus a finite fraction of $\frac{1}{n}$ of the closed geodesics on $\Gamma\backslash\mathcal{H}$ are mapped by g again to closed geodesics on the same surface. Notice that this reasoning is similar to the one leading to (3.17).

We are now going to consider the modular group $\Gamma_{mod} = SL(2,\mathbb{Z})$ as an example to illustrate the above construction explicitly. As already mentioned, this was historically also the first case to be studied, and where modular correspondences have been introduced. A nice presentation of several facts about the modular group can be found in [96]; also [86, 71] are useful to be consulted. The first task in this context is to obtain the commensurator $\overline{\Gamma}$ of the modular group. This can be found to be $\overline{\Gamma} = GL^+(2,\mathbb{Q})$, a fact that is relatively easy to prove. (See [86, 71] for details.) Notice that $GL^+(2,\mathbb{Q})$ is dense in $GL^+(2,\mathbb{R})$, as it is predicted by Margulis' theorem. Let now be $M_n(\mathbb{Z}) := \{g \in M(2,\mathbb{Z});\ det\, g = n\}$, $n \in \mathbb{N}$. This set is not

a group, as the multiplication does not close in it. However, it may be decomposed disjointly as (see [96])

$$M_n(\mathbb{Z}) = \bigcup_{\substack{ad=n\\ 0\leq b<d}} \Gamma_{mod}\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} . \tag{3.63}$$

Now define the semi-group

$$\Delta := \{g \in M(2,\mathbb{Z});\ det\, g > 0\} = \bigcup_{n\in\mathbb{N}} M_n(\mathbb{Z}) \,, \tag{3.64}$$

and take a $g \in \overline{\Gamma}$. Then there exists a $q \in \mathbb{Q}$ such that $g' := qg \in \Delta$. Since $g^{-1}\Gamma g = g'^{-1}\Gamma g'$, ones attention may be restricted from $\overline{\Gamma}$ to Δ. In [71] one can now find the following result: let $g \in \Delta$, then there exist $l, m \in \mathbb{N}$, $l|m$ (this notation means that l is a divisor of m), such that $\Gamma_{mod} g \Gamma_{mod} = \Gamma_{mod}\left(\begin{smallmatrix} l & 0 \\ 0 & m \end{smallmatrix}\right)\Gamma_{mod}$. It is then further shown in [71] that $\Gamma_{mod}\left(\begin{smallmatrix} l & 0 \\ 0 & m \end{smallmatrix}\right)\Gamma_{mod} = \cup\, \Gamma_{mod}\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, $ad = lm$, $0 \leq b < d$, $(a,b,d) = l$. (The latter notation means that the largest common divisor of a, b and d is l.) Thus the $\{\alpha_i\}$ in (3.61) are given by $\{\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right);\ ad = lm,\ 0 \leq b < d,\ (a,b,d) = l\}$. Conventionally one looks at all $g \in \Delta$ with $det\, g = n$, $n \in \mathbb{N}$, simultaneously and finds that

$$\bigcup_{\substack{g\in\Delta\\ det\, g=n}} \Gamma_{mod} g \Gamma_{mod} = M_n(\mathbb{Z}) = \bigcup_{\substack{ad=n\\ 0\leq b<d}} \Gamma_{mod}\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} . \tag{3.65}$$

The simplest non-trivial case is $n = 2$, where one has to choose $g = \left(\begin{smallmatrix} l & 0 \\ 0 & m \end{smallmatrix}\right)$ with $l = 1$, $m = 2$. Then

$$\Gamma_{mod}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\Gamma_{mod} = \Gamma_{mod}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \cup \Gamma_{mod}\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \cup \Gamma_{mod}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} . \tag{3.66}$$

Since the r.h.s. of (3.66) consists of three cosets, the pseudosymmetry related to this $g \in \Delta$ is of order three. The corresponding covering $\Gamma'(g)\backslash\mathcal{H} \to \Gamma_{mod}\backslash\mathcal{H}$ will be found by studying the effect of g on an arbitrary $\gamma \in \Gamma_{mod}$. Let therefore be $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_{mod}$. Then $g^{-1}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) g = \left(\begin{smallmatrix} a & 2b \\ c/2 & d \end{smallmatrix}\right)$. Thus

$$\Gamma'(g) = g^{-1}\Gamma_{mod} g \cap \Gamma_{mod} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_{mod};\ b, c \text{ even} \right\} \tag{3.67}$$

is the group that defines the desired covering over the modular surface $\Gamma_{mod}\backslash\mathcal{H}$.

The modular group certainly provides the simplest example to consider. In [86, 71] also congruence subgroups of Γ_{mod} are treated explicitly. Miyake even goes one step further in that he deals with arithmetic groups that are unit groups of quaternion algebras over $\mathbb{Q}$. Already then he has to argue adelicly, which complicates the discussion considerably. To the author's knowledge there does not exist any explicit treatment of arithmetic groups derived from quaternion algebras over number fields K of degree $n \geq 2$. In such cases a first major obstacle is to identify the commensurator group $\overline{\Gamma}$. An explicit knowledge of the relations (3.60) and (3.61) for a given arithmetic group is mandatory to construct the Hecke ring and the Hecke operators for that group explicitly. From the case of the modular group one learned that this knowledge helped a lot for the numerical determination of the energy eigenvalues and the eigenfunctions, see [91]. To apply the methods used for the modular group also to e.g. the regular octagon group, however, seems at the moment too hard a problem.

3.6 Statistical Properties of Geodesic Length Spectra

This final section of the present chapter contains a topic that might serve as a link to the following chapter dealing with the quantum mechanics of arithmetical chaos. The reason for this is that the statistical properties of length spectra on hyperbolic surfaces are investigated using the Selberg trace formula and the Selberg zeta function, which provide a duality relation between classical and quantum aspects. It is then possible to use results on quantum energy spectra to gain information on classical length spectra. Also, some analytic methods employed here using properties of the Selberg zeta function are quite similar to those that will be applied in chapter 4.

In the same way as the spectral staircase $N(E)$ plays a major role in the investigation of quantum energy spectra, the classical counting function $\mathcal{N}_p(l)$ is an important tool to study geodesic length spectra. Since the PGT (3.8) states that the respective counting functions $\mathcal{N}_p(l)$ share the same asymptotic behaviour for all hyperbolic surfaces, it is the remainder to the asymptotic value that is the quantity of interest. It plays a similar role as $N_{fl}(E)$ does for quantum energy spectra. The aim of this section now is to gain information on this remainder; and by what has been said so far about geodesic length spectra of hyperbolic surfaces it seems to be clear that arithmetic Fuchsian groups have to be distinguished from non-arithmetic ones. For the analytic part of the study an analogy to the Riemann zeta function $\zeta(s)$ appears to be constructive, since an important tool to obtain information on the remainder term to the PGT will be the *Selberg zeta function* $Z(s)$ [84, 51, 100], which is in many respects similar to the Riemann zeta function. The theory of the latter is briefly reviewed in appendix A. A detailed analysis of $\zeta(s)$ then leads to the *prime number theorem* (PNT) [97, 59], which states that the number of primes $\pi(x)$ not exceeding the value x is asymptotically given by $\pi(x) \sim li(x) \sim \frac{x}{\log x}$, $x \to \infty$. Identifying the n-th prime p_n with e^{l_n} then clearly shows a close resemblance of the PNT with the PGT (3.8). It shall now be explained how far the analogy between primes and primitive closed geodesics on a hyperbolic surface can go.

As already mentioned, the Selberg zeta function $Z(s)$ is needed in order to proof the PGT and to estimate the remainder term. The notation already indicates that the Selberg zeta function is the dynamical zeta function (2.9) for geodesic flows on hyperbolic surfaces. In the same way as the dynamical zeta function is derived from Gutzwiller's trace formula with a special test function (2.7), the Selberg zeta function is obtained from the *Selberg trace formula* [84, 51, 100]. The latter is an exact analogue of the smeared version (1.4) of Gutzwiller's trace formula, i.e. it is an exact identity and not only a semiclassical approximation.

For ease of notation from now on only strictly hyperbolic Fuchsian groups will be considered in this section. The general case of Fuchsian groups of the first kind yields nothing new regarding the present problem, since the contributions of the hyperbolic elements are the relevant ones concerning length spectra. But of course, elliptic and parabolic elements can also be treated, see [84, 51, 100] for details.

The quantization of the geodesic flow on a hyperbolic surface $\Gamma\backslash\mathcal{H}$ is determined by the stationary Schrödinger equation

$$-\Delta\psi(z) = E\psi(z) . \tag{3.68}$$

The *hyperbolic Laplacian* is given in terms of the coordinates of $\mathcal{H}$ by

$$\Delta = y^2(\partial_x^2 + \partial_y^2) , \tag{3.69}$$

and the wave functions are required to be invariant under the operation of Γ on $\mathcal{H}$, $\psi(\gamma z) = \psi(z)$ for all $\gamma \in \Gamma$, in order to yield functions on the orbit space $\Gamma\backslash\mathcal{H}$. $-\Delta$ can then be defined as

a self-adjoint operator on $L^2(\Gamma\backslash\mathcal{H})$ with a purely discrete spectrum $0 = E_0 < E_1 \leq E_2 \leq \ldots$, $E_n = p_n^2 + \frac{1}{4}$. The scalar product on $L^2(\Gamma\backslash\mathcal{H})$ is derived from the hyperbolic metric as

$$<\psi,\varphi> = \int_{\Gamma\backslash\mathcal{H}} \frac{dx\,dy}{y^2}\,\overline{\psi(z)}\varphi(z) , \tag{3.70}$$

for $\psi, \varphi \in L^2(\Gamma\backslash\mathcal{H})$. Quantum energies E_k that are related to complex momenta p_k, $0 < E_k < \frac{1}{4}$, are called *small eigenvalues*. Their existence depends on the geometry of $\Gamma\backslash\mathcal{H}$, and it is known that only finitely many can exist on a single surface, see [36] for a review. On compact surfaces of genus $g = 2$ at most one might occur [83]. Small eigenvalues play a special role in connection with the PGT and have to be treated separately.

The Selberg trace formula now reads [84, 51, 100]

$$\sum_{n=0}^{\infty} h(p_n) = \frac{\text{area}(\mathcal{F})}{4\pi}\int_{-\infty}^{+\infty} dp\; p\,h(p)\tanh(\pi p) + \sum_{\{\gamma\}_p}\sum_{k=0}^{\infty}\frac{l(\gamma)\;g(kl(\gamma))}{2\sinh(kl(\gamma)/2)} . \tag{3.71}$$

The outer sum on the r.h.s. of (3.71) runs over all Γ–conjugacy classes of primitive hyperbolic $\gamma \in \Gamma$, thus equivalently, over all primitive closed geodesics on $\Gamma\backslash\mathcal{H}$. Comparing (3.71) with Gutzwiller's trace formula (1.4) shows that all Lyapunov exponents are identical, $\lambda_\gamma = 1$ for all γ, and hence are also identical to the metric entropy $\overline{\lambda} = 1$. An admissible test function $h(p)$ has to be even, and to be holomorphic in the strip $|Im\,p| \leq \frac{1}{2}+\varepsilon, \varepsilon > 0$. Also, $h(p) = O(|p|^{-2-\epsilon})$ for $|p| \to \infty$. $g(x) = \int_{-\infty}^{+\infty}\frac{dp}{2\pi}e^{ipx}h(p)$ denotes the Fourier-transform of the test function. The Selberg zeta function $Z(s)$ arises from (3.71) as the dynamical zeta function has been obtained from (2.7), see (2.7)–(2.9). Its Euler product converges for $Re\,s > \tau = 1$,

$$Z(s) = \prod_{\{\gamma\}_p}\prod_{n=0}^{\infty}\left(1 - e^{-(s+n)l(\gamma)}\right) , \tag{3.72}$$

where the variable $s = \frac{1}{2} - ip$ is related to the energy variable through $E = s(1-s)$. Choosing the test function $h(p) = \frac{1}{p^2+(s-\frac{1}{2})^2} - \frac{1}{p^2+(\sigma-\frac{1}{2})^2}$ for $Re\,s$, $Re\,\sigma > 1$ yields a regularized trace of the resolvent operator for $-\Delta$, which is the appropriate analogue of (2.7). From this relation one can obtain the analytic properties of $Z(s)$. It is an entire holomorphic function with *trivial zeros* at s=0 of multiplicity $\frac{\text{area}(\mathcal{F})}{2\pi}+1$, at $s = 1$ of multiplicity one, and at $s = -k$, $k \in I\!N$, of multiplicities $\frac{\text{area}(\mathcal{F})}{2\pi}(k+1)$. Its *non-trivial zeros* are related to the eigenvalues of $-\Delta$ through $s_n = \frac{1}{2} \pm ip_n$, $E_n = s_n(1-s_n)$; their multiplicities are given by the respective multiplicities of the eigenvalues. Small eigenvalues therefore correspond to zeros of $Z(s)$ in the interval $(0,1)$. Leaving aside the latter ones $Z(s)$ thus fulfills an analogue of the Riemann hypothesis (RH) for $\zeta(s)$ in that $Re\,s_n = \frac{1}{2}$.

In the case of the Riemann zeta function the magnitude of the remainder term in the PNT is determined by the non-trivial zero with largest real part σ_0, see appendix A. Since for the Selberg zeta function the RH is known to be true, and hence $\sigma_0 = \frac{1}{2}$, once the contributions of small eigenvalues have been extracted explicitly, one would expect the remainder term $Q_R(l)$ in the PGT (see also (3.85)) to grow like

$$|Q_R(l)| = e^{\frac{1}{2}l}\cdot\omega(l) , \tag{3.73}$$

where $\omega(l)$ is a combination of powers and logarithms of l. Up to now it was, however, not possible to prove this. The analogy with $\zeta(s)$ will in the following be pushed as far as possible, adopting the strategy employed in appendix A for $\zeta(s)$ to construct a Dirichlet series whose abscissa of conditional convergence should act as a "detector" for σ_0. It will then become clear where the obstacle comes from that prevents one from proving the assertion (3.73) about the magnitude of the remainder term.

The Selberg zeta function has a simple zero at $s=1$ due to the eigenvalue $E_0=0$. Hence in the vicinity of $s=1$ it behaves like $Z(s)=Z'(1)(s-1)+O((s-1)^2)$. One could now think of using a Dirichlet series for $Z(s)$ itself, since it has no pole at $s=1$, as $\zeta(s)$ has, and thus there seems to be no obvious obstruction in pushing the domain of convergence of the Dirichlet series further to the left in the s-plane. However, $Z(s)$ is entire holomorphic so that even on the critical line $Re\, s=\frac{1}{2}$ there are no poles serving as such obstructions. Therefore the use of $Z(s)$ would not yield the desired effect. To circumvent this problem one would like to use $Z(s)^{-1}$ instead, which has the desired poles at the non-trivial zeros of $Z(s)$, and then subtract the poles at $s=1$ and at the $s_k\in(\frac{1}{2},1)$, $k=1,\dots,M$, corresponding to small eigenvalues. The problem one then immediately faces is that because of the product over n in the Euler product representation (3.72) of $Z(s)$ there is no convenient Dirichlet series for the inverse of the Selberg zeta function. The way out of this problem will be to discard the n-product and to define a new *Ruelle-type zeta function* $R(s):=\frac{Z(s)}{Z(s+1)}$, which has for $Re\, s>1$ the Euler product representation

$$R(s)=\prod_{\{\gamma\}_p}(1-e^{-sl(\gamma)})\,. \tag{3.74}$$

This is easily obtained by inserting (3.72) into the definition of $R(s)$. (3.74) now is the exact analogue to the inverse of the Euler product for $\zeta(s)$. Using the analytic properties of $Z(s)$ one can derive those of $R(s)$. The latter is a meromorphic function of $s\in\mathbb{C}$ that is holomorphic for $Re\, s>0$. In this half-plane it has the same zeros as $Z(s)$. Small eigenvalues lead to zeros at $s_1,\dots,s_M$ in $(\frac{1}{2},1)$ and at $1-s_1,\dots,1-s_M$ in $(0,\frac{1}{2})$; the eigenvalue $E_0=0$ produces a zero at $s_0=1$. If the small eigenvalues are not degenerate, $R(s)$ behaves in the vicinity of s_k, $k=0,\dots,M$, like $R(s)=\frac{Z'(s_k)}{Z(s_k+1)}(s-s_k)+O((s-s_k)^2)$. Its logarithmic derivative hence has a simple pole with residue one at the s_k's. In order to obtain a meromorphic function with poles at the non-trivial zeros of $Z(s)$ on the critical line that is holomorphic for $Re\, s>\frac{1}{2}$ one can define

$$f_R(s):=\frac{R'(s)}{R(s)}-\sum_{k=0}^{M}\frac{Z'(s_k)}{Z(s_k+1)}\frac{1}{R(s)}\,. \tag{3.75}$$

To simplify the notation it will henceforth be assumed that there do not occur small eigenvalues, hence $M=0$. The only pole that has to be subtracted then is the one at $s_0=1$.

The next step now consists of finding a Dirichlet series representation for $f_R(s)$. From the Euler product (3.74) for $R(s)$ one finds that for $Re\, s>1$

$$\frac{R'(s)}{R(s)}=\sum_{\{\gamma\}_p}l(\gamma)\frac{e^{-sl(\gamma)}}{1-e^{-sl(\gamma)}}=\sum_{\{\gamma\}_p}\sum_{k=1}^{\infty}l(\gamma)\,e^{-skl(\gamma)}\,. \tag{3.76}$$

In view of Beurling's theory of generalized prime numbers (see e.g. [18]) one can identify primitive hyperbolic conjugacy classes $\{\gamma\}_p$ with primes p [15]. Comparing (3.76) with the Dirichlet series for $-\frac{\zeta'(s)}{\zeta(s)}$ then leads to the definition of an analogue of the von Mangoldt function $\Lambda(n)$, see appendix A. What is lacking so far is an analogue of the positive integers $\mathbb{N}$. Recently it has become customary to introduce such analogues as *pseudo-orbits* [69, 22]. These are the generalized integers in Beurling's theory and comprise of formal combinations of powers of primitive conjugacy classes,

$$\rho:=\{\gamma_1^{k_1}\}_p\oplus\cdots\oplus\{\gamma_n^{k_n}\}_p\,. \tag{3.77}$$

On the surface $\Gamma\backslash\mathcal{H}$ ρ corresponds to a formal combination of the primitive closed geodesics related to $\gamma_1,\dots,\gamma_n$ that are traversed $k_1,\dots,k_n$ times, respectively. These are the objects that were named pseudo-orbits [22]. The corresponding "lengths"

$$L_\rho=k_1l(\gamma_1)+\cdots+k_nl(\gamma_n) \tag{3.78}$$

are then called *pseudo-lengths*. Having defined analogues of integers it is now possible to introduce a von Mangoldt function for $R(s)$,

$$\Lambda_R(\rho):=\begin{cases} l(\gamma) & ,\ \rho=\{\gamma^k\}_p \\ 0 & ,\ \text{otherwise}\end{cases}\,. \tag{3.79}$$

With the help of the above notions one can first introduce a Dirichlet series for $R(s)^{-1}$,

$$\frac{1}{R(s)}=\prod_{\{\gamma\}_p}\sum_{k=0}^{\infty}e^{-skl(\gamma)}=\sum_{\rho}e^{-sL_\rho}\,, \tag{3.80}$$

and then one for $f_R(s)$,

$$f_R(s)=\sum_\rho A_\rho\,e^{-sL_\rho}\,,\qquad A_\rho=\Lambda_R(\rho)-\frac{Z'(1)}{Z(2)}\,. \tag{3.81}$$

In appendix A the Chebyshev functions $\theta(x)$ and $\psi(x)$ were used to relate the analytic properties of $\zeta(s)$ to the PNT. Trying to carry this over to the case of the Ruelle-type zeta function and the PGT one is led to define analogues of the Chebyshev functions as

$$\begin{aligned}\theta_R(L) &:= \sum_{l(\gamma)\le L} l(\gamma)\,,\\ \psi_R(L) &:= \sum_{L_\rho\le L}\Lambda_R(\rho)=\sum_{k\ge1}\sum_{kl(\gamma)\le L}l(\gamma)\\ &= \sum_{k\ge1}\theta_R(L/k)\,.\end{aligned} \tag{3.82}$$

(In the above notations the $l(\gamma)$'s and the L_ρ's are counted with their respective multiplicities.) As with the classical Chebyshev functions, see appendix A, an estimate of the remainder in $\psi_R(L)=\theta_R(L)+\mathcal{R}_R(L)$ gives $\mathcal{R}_R(L)=O(e^{\frac{1}{2}L}L^2)$, $L\to\infty$. Since the analogue of the prime counting function $\pi(x)$ is the counting function $\mathcal{N}_p(l)$ for primitive closed geodesics, exactly the same reasoning as in appendix A leads to

$$\mathcal{N}_p(l)=\frac{\psi_R(l)}{l}+\int_{l_1}^{l}\frac{dl'}{l'^2}\,\psi_R(l')+O(e^{\frac{1}{2}l})\,, \tag{3.83}$$

where l_1 denotes the length of the shortest closed geodesic on $\Gamma\backslash\mathcal{H}$. Using the integral (A.11) then allows to express $\psi_R(L)$ through $\frac{R'(s)}{R(s)}$ by ($b>1$, $L\neq L_\rho$)

$$\psi_R(L)=\sum_\rho\Lambda_R(\rho)\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty}\frac{ds}{s}\,e^{s(L-L_\rho)}=\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty}\frac{ds}{s}\,e^{sL}\,\frac{R'(s)}{R(s)}\,. \tag{3.84}$$

Since the analytic properties of the integrand on the very right of (3.84) are known one would expect to obtain an analogue of the explicit formula (A.13) for the classical Chebyshev function

$\psi(x)$. This would involve a sum $\sum_{s_n} \frac{e^{s_n L}}{s_n}$ over the non-trivial zeros of $Z(s)$ on the critical line. At this point, however, the present case differs from that of the Riemann zeta function in that the sum over the s_n's diverges and thus no explicit formula in the desired manner exists for $\psi_R(L)$. The reason for this difference lies in the stronger growth of $N(E)$ as compared to the counting function $N_\zeta(p) := \{s_n = \beta_n + i\gamma_n;\ 0 < \gamma_n \leq p\}$. From (3.71) one can rederive Weyl's law (1.3) to yield $N(E(p)) \sim \frac{area(\mathcal{F})}{4\pi} p^2$, $p \to \infty$, whereas it is known [97] that $N_\zeta(p) \sim \frac{p}{2\pi}\log p$, $p \to \infty$. There are hence "too many" terms in the sum over the s_n's that prevent it from converging.

Hejhal [51] proceeds in defining an integrated version $\psi_{R,1}(L) := \int_0^L dl\, e^l\, \psi_R(l)$ of the Chebyshev function $\psi_R(L)$. Inspecting (3.84) one observes that one has to deal with the sum $\sum_{s_n} \frac{e^{s_n L}}{s_n(s_n+1)}$ instead, which is conditionally convergent. A tedious analysis then leads to the PGT (*Theorem 6.19* in [51])

$$\begin{aligned} \psi_R(L) &= e^L + \sum_{k=1}^{M} e^{s_k L} + P_R(L)\,, \qquad P_R(L) = O(e^{\frac{3}{4}L} L^{\frac{1}{2}})\,, \\ \mathcal{N}_p(l) &= Ei(l) + \sum_{k=1}^{M} Ei(s_k l) + Q_R(l)\,, \qquad Q_R(l) = O(e^{\frac{3}{4}l} l^{-\frac{1}{2}})\,, \end{aligned} \tag{3.85}$$

where the contributions of small eigenvalues have been reintroduced explicitly.

Another way to obtain the PGT is to restrict the contour of integration in (3.84) to a finite interval and then to estimate the resulting sum. Employing methods that can be found in [97], pp. 60, and in [60] one can show that ($b > 1$, $0 < T < \infty$, $L \neq L_\rho$)

$$\psi_R(L) = \frac{1}{2\pi i}\int_{b-iT}^{b+iT} \frac{ds}{s}\, e^{sL}\, \frac{R'(s)}{R(s)} + O(L^2 T^{-1} e^L) + O(T^{-1}(b-1)^{-1} e^{bL})\,. \tag{3.86}$$

The proof for the remainder terms on the r.h.s. is the same as for the lemma of section 4.5, where $Z(s)$ replaces $\frac{R'(s)}{R(s)}$. We therefore postpone its explicit discussion to section 4.5. Deforming the contour of integration in (3.86) and using estimates for $\frac{R'(s)}{R(s)}$ on the contour Iwaniec succeeded to estimate $Q_R(l)$ for the modular group as [60]

$$Q_{R,mod}(l) = O(e^{(\frac{35}{48}+\varepsilon)l}) \quad \forall \varepsilon > 0\,, \tag{3.87}$$

which is slightly better than the general result (3.85). This is also the best upper bound for the remainder term in a PGT available. It seems that it is very hard to "break the barrier" at $e^{\frac{3}{4}l}$ for the upper bound of $Q_R(l)$.

This difficulty has a consequence for estimating the abscissa of conditional convergence for the Dirichlet series representation (3.81) of $f_R(s)$. Since $f_R(s)$ has been designed in complete analogy to the function $f(s)$ introduced in appendix A to study the PNT, the analysis applied to the latter function will now be repeated for the former one. In appendix A also the convergence properties of general Dirichlet series have been reviewed, according to which the series (3.81) converges for $Re\, s > \sigma_c$ and converges absolutely for $Re\, s > \sigma_a$, $\sigma_a \geq \sigma_c$. Arranging the pseudo-orbits in ascending order of their lengths, $L_1 \leq L_2 \leq L_3 \leq \ldots$, one finds that

$$\begin{aligned} \sigma_a &= \limsup_{N\to\infty} \frac{1}{L_N} \log \sum_{n=1}^{N} \left| \Lambda_R(n) - \frac{Z'(1)}{Z(2)} \right|\,, \\ \sigma_c &= \limsup_{N\to\infty} \frac{1}{L_N} \log \left| \sum_{n=1}^{N} \left(\Lambda_R(n) - \frac{Z'(1)}{Z(2)} \right) \right| \end{aligned} \tag{3.88}$$

$$= \limsup_{N\to\infty} \frac{1}{L_N} \log \left| \psi_R(L_N) - \mathcal{N}^{(P)}(L_N)\frac{Z'(1)}{Z(2)} \right|\,,$$

where $\mathcal{N}^{(P)}(L) := \#\{\rho;\ L_\rho \leq L\}$ denotes the counting function for pseudo-orbits. In [15] it was shown that

$$\mathcal{N}^{(P)}(L) = \frac{Z(2)}{Z'(1)}\, e^L + Q^{(P)}(L)\,, \qquad Q^{(P)}(L) = O(e^{L - c_1 L^\alpha})\,, \tag{3.89}$$

with some constants $c_1 > 0$ and $0 < \alpha < \frac{1}{3}$. As for $f(s)$ (see appendix A) one also concludes here, using (3.89), that $\sum_{n=1}^{N} |\Lambda_R(n) - \frac{Z'(1)}{Z(2)}| \sim const. \cdot e^{L_N}$. The Dirichlet series (3.81) therefore converges absolutely for $Re\, s > \sigma_a = 1$. Inserting $\psi_R(L) = e^L + P_R(L)$ one observes that

$$\sigma_c = \limsup_{N\to\infty} \frac{1}{L_N} \log \left| P_R(L_N) - \frac{Z'(1)}{Z(2)} Q^{(P)}(L_N) \right|\,. \tag{3.90}$$

The only conclusion one can now draw from (3.90), and from the PGT (3.85), $P_R(L) = O(e^{\frac{3}{4}L}L^{\frac{1}{2}})$, is that $\sigma_c \leq 1$. The analytic properties of $f_R(s)$, however, suggest that $\sigma_c = \frac{1}{2}$, since $f_R(s)$ is a holomorphic function for $Re\, s > \frac{1}{2}$ and has poles on the critical line. The analogous function $f(s)$ built from the Riemann zeta function indeed shows this behaviour: its Dirichlet series converges (conditionally) on the maximal half-plane $Re\, s > \sigma_0$ where $f(s)$ is still holomorphic.

The weakness of the upper bound $\sigma_c \leq 1$ hinges on the estimate $Q^{(P)}(L) = O(e^{L - c_1 L^\alpha})$, which was obtained in [15] from the theory of generalized prime numbers, see e.g. [18]. However, this upper bound only requires the rather weak estimate $Q_R(l) = O(e^{l - cl^\beta})$ for some constants $c > 0$ and $0 < \beta < 1$. Since $Q_R(l)$ is known to be smaller than required, see (3.85), the true magnitude of $Q^{(P)}(L)$ could be much smaller than just being below e^L. To support this idea a function similar to $f_R(s)$ will be introduced. Starting with the Dirichlet series (3.81) for $R(s)^{-1}$ one obtains for $Re\, s > 1$ through an integration by parts

$$\begin{aligned} \frac{1}{R(s)} = \int_0^\infty d\mathcal{N}^{(P)}(L)\, e^{-sL} &= \mathcal{N}^{(P)}(L)\, e^{-sL}\Big|_0^\infty + s\int_0^\infty dL\, \mathcal{N}^{(P)}(L)\, e^{-sL} \\ &= \frac{Z(2)}{Z'(1)}\frac{s}{s-1} + s\int_0^\infty dL\, Q^{(P)}(L)\, e^{-sL}\,. \end{aligned} \tag{3.91}$$

This yields an integral representation

$$\frac{1}{R(s)} - \frac{Z(2)}{Z'(1)}\frac{1}{s-1} = \frac{Z(2)}{Z'(1)} + s\int_0^\infty dL\, Q^{(P)}(L)\, e^{-sL}\,, \tag{3.92}$$

of a function that is holomorphic for $Re\, s > \frac{1}{2}$ and has poles on the critical line. The estimate for $Q^{(P)}(L)$, however, only permits to use the r.h.s. of (3.92) for $Re\, s \geq 1$. Again, the domain of holomorphy of the l.h.s. suggests that the integral might exist for $Re\, s > \frac{1}{2}$, yielding the estimate $Q^{(P)}(L) = O(e^{\frac{1}{2}L})$. Assuming now that indeed $\sigma_c = \frac{1}{2}$ for the Dirichlet series of $f_R(s)$, then by (3.90) one concludes that $P_R(L) = O(e^{(\frac{1}{2}+\varepsilon)L})$ $\forall \varepsilon > 0$, leading to the estimate $Q_R(l) = O(e^{(\frac{1}{2}+\varepsilon)l})$ $\forall \varepsilon > 0$. The desired result (3.73) thus would follow if one assumed that the two representations (3.81) and (3.92) converged in the maximal half-planes where the functions they define are still holomorphic.

Certainly, (3.92) gives the lower bound $Q^{(P)}(L) = \Omega(e^{\frac{1}{2}L})$. This estimate already accounts for the lower bound $\sigma_c \geq \frac{1}{2}$, that arises from the fact that $f_R(s)$ has poles on the critical line, by inserting $Q^{(P)}(L)$ into (3.90). For $P_R(L)$ Hejhal could prove the lower bound $P_R(L) = \Omega_\pm(e^{\frac{1}{2}L}(\log L)^{\frac{1}{2}})$ [51], implying for the remainder term in the PGT $Q_R(l) = \Omega_\pm(e^{\frac{1}{2}l}l^{-1}(\log l)^{\frac{1}{2}})$. But unfortunately no further rigorous conclusion can be drawn from (3.90) and (3.92). The analytic theory seems to be stuck at this point. The question one might ask now is whether the problem is a technical one or whether there lurks some yet undiscovered phenomenon behind it.

One idea that could come to one's mind is that $|Q_R(l)| = e^{\frac{1}{2}l}\omega(l)$ is indeed true for generic, i.e. non-arithmetic, Fuchsian groups. The arithmetic case should then be treated separately and might violate the expected behaviour of $Q_R(l)$. It seems to be quite natural to distinguish arithmetic from non-arithmetic groups, after having having discussed the exceptional structure of length spectra in the arithmetic case. $\mathcal{N}_p(l)$ is a staircase function with steps of width $\Delta l_n = l_{p,n+1} - l_{p,n}$ and of height $g_p(l_{p,n})$ at $l = l_{p,n}$. It is thus the interplay of fluctuations of lengths and multiplicities that results in fluctuations of the staircase function $\mathcal{N}_p(l)$. This is in contrast to the fluctuation properties of the spectral staircase $N(E)$, for which only fluctuations in the quantum energies E_n are responsible (since we require the systems under consideration to be completely desymmetrized and thus being void of degeneracies in their energy spectra). The interferences of the two contributions to $Q_R(l)$ describing the fluctuations of $\mathcal{N}_p(l)$ are involved, but different multiplicities of lengths in the arithmetic and the non-arithmetic cases clearly lead to different kinds of fluctuations. For arithmetic groups the mean step heights are $< g_p(l) > \sim c_\Gamma \frac{e^{l/2}}{l}$, $l \to \infty$. These alone give a contribution of the order $l^{-1}e^{l/2}$ to $Q_R(l)$, since the mean behaviour $Ei(l)$ cannot follow the step structure of $\mathcal{N}_p(l)$. In addition to its mean behaviour fluctuations of $g_p(l)$ can give further contributions to $Q_R(l)$, let alone the fluctuations of the lengths themselves. For non-arithmetic groups the mean multiplicities, i.e. the mean step heights, do not give an exponential contribution to $Q_R(l)$. Since in any case the lower bound $Q_R(l) = \Omega_\pm(e^{\frac{1}{2}l}l^{-\frac{1}{2}}(\log l)^{\frac{1}{2}})$ requires exponentially large oscillations about $Ei(l)$ (in the positive and the negative direction), in the non-arithmetic case these have to come from length fluctuations.

In the following arithmetic and non-arithmetic Fuchsian groups will therefore be treated separately. At first for the generic, non-arithmetic case an approach of *inverse quantum chaology* will be employed. The latter notion stands for drawing conclusions on the classical properties of a chaotic system from its quantum energy spectrum using the (Selberg or Gutzwiller) trace formula. As mentioned in chapter 2 Berry's theory of spectral rigidity predicts for the energy dependence of the saturation value $\Delta_\infty(E)$ of the rigidity a logarithmic behaviour, $\Delta_\infty(E) \sim \frac{1}{2\pi^2}\log E$, $E \to \infty$, if the classical system is chaotic and time-reversal invariant. It has also been mentioned in chapter 2 that this result cannot be applied to arithmetic systems because of their exponentially degenerate length spectra. Since $\Delta_\infty(E)$ is related to $N_{fl}(E)$ via (2.24), one can conclude that $|N_{fl}(E)| \sim \frac{1}{\sqrt{2\pi}}\sqrt{\log E}$, $E \to \infty$. Hejhal now proved a theorem that yields an estimate for the remainder term in the PGT depending on an upper bound for $N_{fl}(E)$, therefore being truely a result in the spirit of inverse quantum chaology. In the mathematical literature on the Selberg trace formula it has become customary to introduce the notation $S(p) := N_{fl}(E(p)) = \frac{1}{\pi}\arg Z(\frac{1}{2}+ip)$. Hejhal's *Theorem 14.18* [51] then states that for $0 < \delta < \infty$ the estimate $|S(p)| = O((\log p)^\delta)$ implies $P_R(L) = O(e^{\frac{1}{2}L}L^{2+\delta})$. In our case, $\delta = \frac{1}{2}$, the upper bound $Q_R(l) = O(e^{\frac{1}{2}l}l^{\frac{3}{2}})$ follows. Comparing this with the lower bound $Q_R(l) = \Omega_\pm(e^{\frac{1}{2}l}l^{-1}(\log l)^{\frac{1}{2}})$ one concludes that $|Q_R(l)| = e^{\frac{1}{2}l}\omega(l)$, where the asymptotics of $\omega(l)$ lie somewhere between $\frac{\sqrt{\log l}}{l}$ and $l^{\frac{3}{2}}$. It should be stressed that this conclusion has not been drawn rigorously because it hinges on Berry's non-rigorous theory for $\Delta_3(L;E)$. It is not clear either what effect the

unboundedness of the multiplicities in the length spectra even for non-arithmetic groups will have on the applicability of Berry's theory. However, together with everything else discussed above, this gives a further hint supporting the expectation that the remainder term in the PGT grows like (3.73) for generic Fuchsian groups.

Considering arithmetic groups the inverse quantum chaology reasoning does not yield the same result as in the generic case since the upper bound for $S(p)$ assumed in *Theorem 14.18* of [51] is violated here. For a certain class of arithmetic groups (derived from quaternion algebras defined over $\mathbb{Q}$) Hejhal obtains in his *Theorem 18.8* and *Remark 18.14* of [51] the lower bound $S(p) = \Omega_\pm(\sqrt{p}/\log p)$. (Originally this was an unpublished result of Selberg.) Since the proof only requires the exponential increase of the multiplicities of lengths present for all arithmetic groups, the Ω–result for $S(p)$ applies to any arithmetic group. This lower bound is quite close to the general upper bound $S(p) = O(p/\log p)$, see [51]. In order to deal with the arithmetic case one then has to use Hejhal's *Theorem 15.13* [51] which states that if $|S(p)| = O(p^\alpha)$, $0 < \alpha < 1$, then $P_R(L) = O(e^{\frac{1+2\alpha}{2+2\alpha}L}L^{\frac{1}{1+\alpha}})$. The general upper bound for $S(p)$ requires to take the limit $\alpha \to 1$, yielding $Q_R = O(e^{\frac{3}{4}l}l^{-\frac{1}{2}})$, which is exactly like in the PGT (3.85). In section 4.4, however, it will be argued that $|S(p)| \sim \sqrt{\frac{d}{\pi}}\sqrt{\frac{p}{\log p}}$, $p \to \infty$, so that $\alpha = \frac{1}{2}$ can be chosen. This yields the result $Q_R(l) = O(e^{\frac{2}{3}l}l^{-\frac{1}{3}})$. Thus, using inverse quantum chaology in conjunction with reasonable assumptions on the fluctuations in the respective quantum energy spectra, the upper bounds on $Q_R(l)$ could be improved. But only in the non-arithmetic case the expectation (3.73) could be supported. In the arithmetic case it was merely possible to bring the bound down to $e^{\frac{2}{3}l}l^{-\frac{1}{3}}$.

The methods of inverse quantum chaology did not work as effective as desired when applied to arithmetic Fuchsian groups. It will thus now be tried to support the assumption (3.73) on the magnitude of the remainder term in the PGT by numerical investigations of some arithmetic groups. Again, there are two possible approaches one could choose: a direct calculation of $Q_R(l)$ from length spectra, or an indirect one in the spirit of inverse quantum chaology. The latter way was employed by Aurich and Steiner in [14]. To explain this approach one has to go back to the starting point (3.86) of a modified explicit formula for $\psi_R(L)$. Iwaniec succeeded in showing from (3.86) [60] that for $1 \le T \le e^{\frac{1}{2}L}L^{-2}$

$$\psi_R(L) = e^L + \sum_{|p_n|\le T} \frac{e^{s_n L}}{s_n} + O(T^{-1}L^2e^L) . \tag{3.93}$$

The (finite) sum runs over pairs of non-trivial zeros $s_n = \frac{1}{2} \pm ip_n$ of $Z(s)$. The optimal choice for T to resolve a given L then is $T = L\,e^{\frac{1}{4}L}$ [60]. Inserting (3.93) into (3.83) thus yields

$$\mathcal{N}_p(l) = Ei(l) + \sum_{|p_n|\le T} Ei(s_n l) + O(T^{-1}le^l) . \tag{3.94}$$

The optimal choice for T results in a remainder term of $O(e^{\frac{3}{4}l})$ in (3.94). Aurich and Steiner, however, derived the formula

$$\mathcal{N}_p(l) = Ei(l) - \frac{1}{2}Ei(l/2) + \sum_{|p_n|\le T} Ei(s_n l) + \dots \tag{3.95}$$

by employing the Selberg trace formula with a special test function. They hence added an additional contribution of $-\frac{1}{2}Ei(l/2)$ to the r.h.s. of (3.94). In the modified explicit formula (3.93) this extra term does not appear explicitly because it is hidden in the remainder. In the numerical evaluation of (3.95) this, however, appears to be needed. For the (arithmetic) regular octagon group Aurich and Steiner calculated the first 200 eigenvalues, leading to a cut-off at $T = 14.2$. Relating this to l through the optimal choice $T = l\,e^{\frac{1}{4}l}$ yields $l \approx 4.6$. Indeed, the numerical calculations presented in [14] show a good approximation of $\mathcal{N}_p(l)$ in the interval $3 \le l \le 6$. And furthermore, the r.h.s. of (3.95) reveals a reasonable approximation to the actual staircase even at larger values of l, which, however, cannot follow the step structure properly. But the approximation does not "leave" the staircase. This can be taken as an indication that the omitted remainder on the r.h.s. of (3.95) does not exceed the mean step height $< g_p(l) > \sim 8\sqrt{2}\frac{e^{l/2}}{l}$. Thus it seems that the complete remainder on the r.h.s. of (3.94) is of the order of magnitude of $\frac{e^{l/2}}{l}$.

In order to investigate the fluctuation properties of arithmetic length spectra numerically in detail, first fluctuations of the multiplicities will be studied. The first example chosen is given by Artin's billiard [68]. This is a system that is derived from a desymmetrization of the modular surface $\Gamma_{mod}\backslash\mathcal{H}$. The latter possesses an orientation reversing symmetry commuting with the hyperbolic Laplacian. Dividing this symmetry out leads to a billiard problem in a triangle on the hyperbolic plane $\mathcal{H}$ first discussed by Artin [2]. For a detailed presentation of the desymmetrization procedure see [99]. The regular octagon group Γ_{reg} then serves as the second example. A third arithmetic group that is included in the numerical studies presented here is the *Gutzwiller octagon group* Γ_G. This is a group commensurable with Γ_{reg} leading to a compact surface $\Gamma_G\backslash\mathcal{H}$ of genus two. The intersection of Γ_{reg} and Γ_G is a subgroup of index two in Γ_{reg} and of index five in Γ_G. The mean multiplicity of lengths of closed geodesics on $\Gamma_G\backslash\mathcal{H}$ is thus asymptotically smaller by a factor of $\frac{2}{5}$ (see (3.20)) than that for the regular octagon group. For a detailed description of Γ_G see [73]. The numerical data of the lengths and multiplicities in the three examples have been kindly placed at disposal by the authors of [68, 4, 73].

As a numerical expression to calculate the mean multiplicities

$$\frac{1}{2N}\sum_{k=-N}^{+N} g_p(l_{p,n+k}) = \frac{1}{2N}\left[\mathcal{N}_p(l_{n+N}) - \mathcal{N}_p(l_{n-N-1})\right] \tag{3.96}$$

is taken. Asymptotically, for $l_n \to \infty$, this expression approaches $< g_p(l_n) >$, compare (3.13). Analogously, $< g_p(l_n)^2 >$ is approximated by $\frac{1}{2N}\sum_{k=-N}^{+N} g_p(l_{p,n+k})^2$. Employing Chebyshev's inequality, stating that $(\sum_{n=1}^{N} a_n)^2 \le N\sum_{n=1}^{N} a_n^2$ if all $a_n \ge 0$, one merely ends up with the lower bound

$$\sigma_g^2(l) := < g_p(l)^2 > - < g_p(l) >^2 \ge 0\,. \tag{3.97}$$

To find out about the asymptotic l-dependence of the fluctuations $\sigma_g^2(l)$ numerical calculations were performed in the three systems mentioned above. In fig.1 $\sigma_g^2(l)\,l^2\,e^{-l}$ is plotted for the three systems mentioned above. The respective primitive length spectra are completely known up to $l_{max} = 19.360$ for Artin's billiard, $l_{max} = 18.092$ for the regular octagon, and $l_{max} = 17.680$ for the Gutzwiller octagon. The averaging has been performed over $2N = 100$ lengths in all three examples. One notes that asymptotically $\sigma_g^2(l)\,l^2\,e^{-l}$ seems to fluctuate about a constant value so that the asymptotic behaviour of the fluctuations appears to be $|\sigma_g(l)| \sim const.\ \frac{e^{l/2}}{l}$, $l \to \infty$. It thus turns out that the fluctuations of the multiplicities about their mean values do not give stronger contributions to $Q_R(l)$ than the mean multiplicities themselves.

One might now be interested in seeing the behaviour of $Q_R(l)$ directly. Since the contributions from the mean of the multiplicities and from their fluctuations are of the order of magnitude of $\frac{e^{l/2}}{l}$, fig.2 presents a numerical calculation of $Q'(l) := |Q_R(l)|\,l\,e^{-l/2}$. As it is

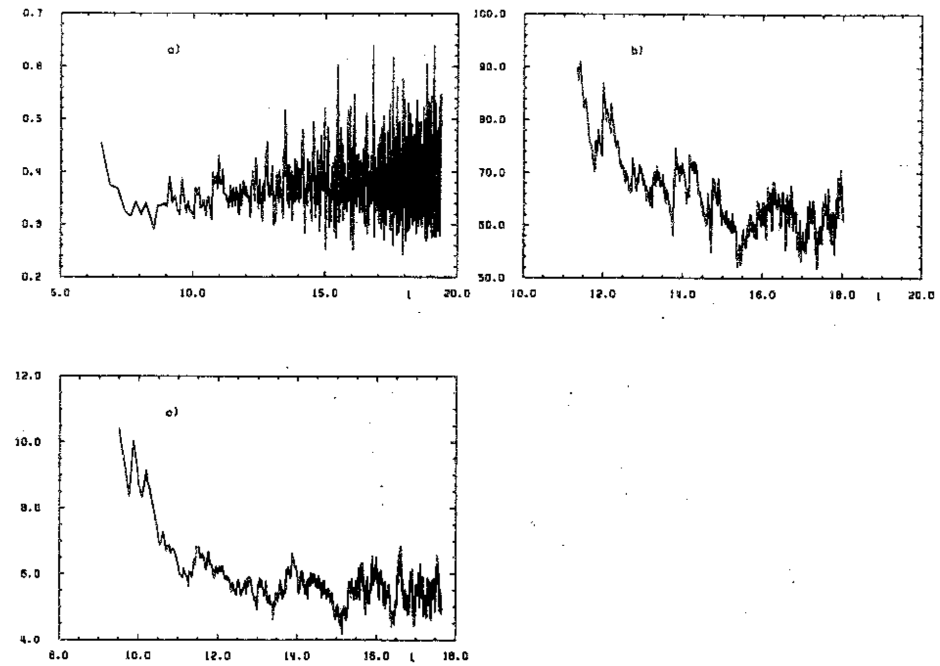

Figure 1: For a) Artin's billiard, b) the regular octagon and c) the Gutzwiller octagon the quantity $\sigma_g^2(l)\,l^2\,e^{-l}$ is shown.

observed from fig.2 that $Q'(l)$ fluctuates in some bounded strip one can conclude that $Q_R(l)$ appears to be of the order of magnitude of $\frac{e^{l/2}}{l}$ in the computed range of l-values. It should, however, be noted that a slightly different power of l in the asymptotics of $Q_R(l)$ can hardly be excluded by the numerical results. Therefore, the expectation (3.73) is supported by numerical evidence to hold also for the arithmetic Fuchsian groups studied here. Numerical calculations for the modular group as well as for a non-arithmetic group also supporting (3.73) were performed in [29].

In summary, there are several reasons, both from an analytical as well as from a numerical point of view, that the remainder term $Q_R(l)$ in the PGT is asymptotically for $l \to \infty$ given by $|Q_R(l)| = e^{\frac{1}{2}}\omega(l)$, where $\omega(l)$ is a combination of powers and logarithms of l, both for arithmetic and non-arithmetic Fuchsian groups. But since the mechanisms of fluctuations are different for these two classes of systems, the methods employed to support the expected behaviour of $Q_R(l)$ had to be different ones. It seems that the non-arithmetic case behaves like other "generic" chaotic systems (see also [87]). But for those the analytic tools used here are not applicable because the Gutzwiller trace formula is not an exact identity and hence no such detailed information about the analytic properties of the associated dynamical zeta functions

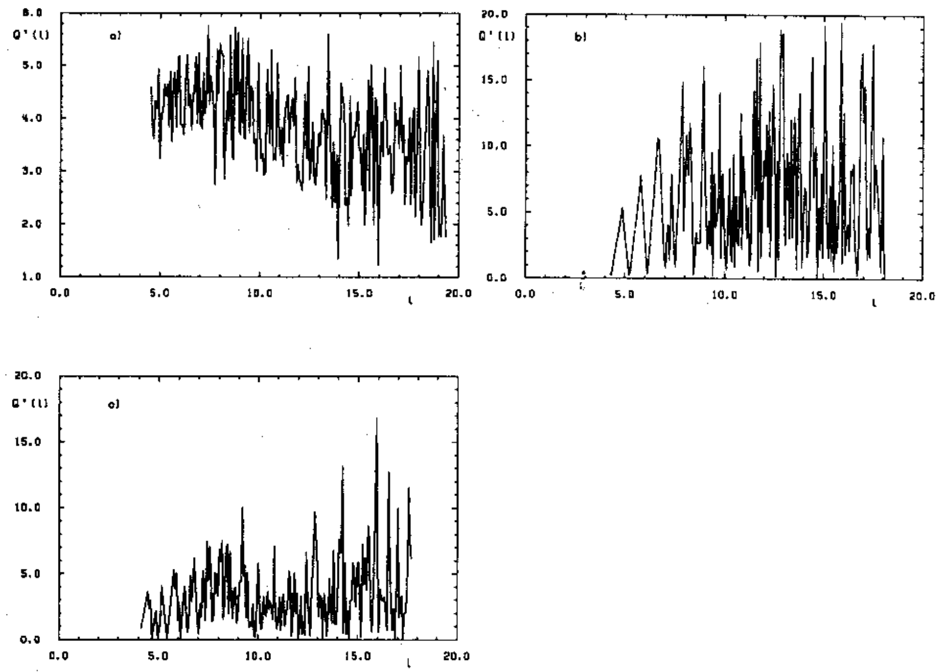

Figure 2: The quantity $Q'(l) = |Q_R(l)|\, l\, e^{-\frac{l}{2}}$ is shown for a) Artin's billiard, b) the regular octagon and c) the Gutzwiller octagon.

are available like for the Selberg zeta function. Inspecting (1.4), however, there seems to be no fundamental difference between the general case and the case studied here. It is only that the powerful machinery of the Selberg trace formula is available for analytic investigations. The only peculiarity of the arithmetic systems that played a role was their exceptional structure of their respective length spectra including the exponential degeneracies, a property not shared by generic systems. One would therefore expect a behaviour like the one for non-arithmetic Fuchsian groups also for general, generic chaotic systems.

4 Quantum Aspects of Arithmetical Chaos

The item of this chapter is a discussion of the quantized versions of geodesic flows on hyperbolic surfaces derived from arithmetic Fuchsian groups. The main objective thereby is to understand the statistical properties of arithmetic energy spectra in contrast to those of non-arithmetic systems. The Selberg trace formula will play a decisive role because it allows to trace back the exceptional spectral statistics of the arithmetical systems to the peculiarities of their respective classical limits, namely to the exponentially growing degeneracies in their classical length spectra.

The first section presents a discussion of pseudosymmetries in the quantum mechanical context and introduces Hecke operators. The discussion and interpretation of these leads to a discovery of constraints on the arithmetical quantum energy spectra that are taken as indications for exceptional statistical properties for the latter. Then the empirical observations about arithmetic quantum energy spectra are reviewed and the role of the form factor for the spectral statistics is discussed. Sections 4.3 and 4.4 present a model to describe the level spacings distributions and the number variance, respectively. The final section of this chapter is devoted to an investigation of the convergence properties of the Selberg zeta function both for arithmetic and non-arithmetic Fuchsian groups.

4.1 Hecke Operators

In section 3.5 pseudosymmetries on arithmetic surfaces have been introduced as generalizations of symmetries. In quantum mechanics the latter ones manifest themselves as being represented unitarily on the wave functions. The *Hecke operators* are generalizations of the representation operators of symmetries to pseudosymmetries. For arithmetic groups they form an infinite algebra of self-adjoint operators commuting with the Hamiltonian $H = -\Delta$.

As in section 3.5 we will now open the discussion of the realization of pseudosymmetries in quantum systems by briefly reviewing the case of symmetries. Let therefore be Γ a Fuchsian group of the first kind that is a normal subgroup of index N in Γ'. $\Gamma/\Gamma' \cong \Sigma = \{\mathbf{1}, g_1, \ldots, g_{N-1}\}$ then is the symmetry group of the surface $\Gamma\backslash\mathcal{H}$. Σ shall be unitarily represented on the (finite dimensional) vector space V_χ by $\chi \in \mathrm{End}(V_\chi)$. Since χ may be decomposed into irreducible components, henceforth only unitary irreducible representation will be discussed; Σ^* then denotes the set of these (the unitary dual of Σ). The relevant quantum mechanical Hilbert space is the space of V_χ-valued square-integrable functions $L^2(\Gamma\backslash\mathcal{H}) \otimes V_\chi$, which may be realized as follows. Let $\psi : \mathcal{H} \to V_\chi$ transform under Γ' via the unitary representation χ that is defined on Γ' by extending it trivially onto Γ, i.e. χ is viewed as a representation of Γ' with $\Gamma \subseteq \ker(\chi)$. Expressed in explicit terms: if $\gamma' \in \Gamma'$, then $\gamma' \in g_i\Gamma$ for some $g_i \in \Sigma$, and, with some $\gamma \in \Gamma$; $\psi(\gamma' z) = \psi(g_i \gamma z) = \chi(g_i)\psi(z)$. One is hence considering the spectral problem of the hyperbolic Laplacian on Γ'-automorphic functions transforming under unitary representations of Γ' that act trivially on Γ such that they yield irreducible representations of $\Sigma \cong \Gamma'/\Gamma$. The representation operators T_{g_i} for $g_i \in \Sigma$ act as

$$T_{g_i}\psi(z) := \psi(\gamma' z) = \chi(g_i)\psi(z)\,. \tag{4.1}$$

On the level of the Selberg trace formula and the Selberg zeta function the desymmetrization procedure can be carried through by decomposing the representation of Γ' that is induced from the trivial representation of $\Gamma \subset \Gamma'$ into irreducible components, see [100, 101]. The Selberg zeta function $Z_\Gamma(s)$ for the group Γ then factorizes as

$$Z_\Gamma(s) = \prod_{\chi\in\Sigma^*} Z_{\Gamma'}(s,\chi)^{\dim V_\chi} , \tag{4.2}$$

where for $Re\, s > 1$

$$Z_{\Gamma'}(s,\chi) = \prod_{\{\gamma'\}_p} \prod_{n=0}^{\infty} \det\left(1 - \chi(\gamma')\, e^{-(s+n)l(\gamma')}\right) \tag{4.3}$$

is the Selberg zeta function referring to Γ' and incorporating the representation χ. (The determinant is defined on the representation space V_χ.) (4.2) clearly shows that an eigenvalue $E_n^\chi = (p_n^\chi)^2 + \frac{1}{4}$ of the hyperbolic Laplacian related to the symmetry class χ occurs with multiplicity dim V_χ, since this is the multiplicity of the corresponding zero $s_n = \frac{1}{2} - ip_n^\chi$ of $Z_\Gamma(s)$. The explicit treatment of an example for such a desymmetrization will be presented in appendix B.

According to the algebraic setting of pseudosymmetries reviewed in section 3.5 the latter ones are related to (non-trivial) elements of the commensurator $\overline{\Gamma}$ of the arithmetic Fuchsian group Γ under consideration. For $g \in \overline{\Gamma}$ the double cosets $\Gamma g \Gamma$ are the basic objects the Hecke ring is constructed from, see [86, 71] as general references. Let now be Δ a semi-group, $\Gamma \subset \Delta \subseteq \overline{\Gamma}$, and form the free $\mathbb{Z}$-module $\mathcal{R}(\Gamma,\Delta)$ generated by all $\Gamma g\Gamma$ for $g \in \Delta$, i.e.

$$\mathcal{R}(\Gamma,\Delta) := \left\{ \sum_{g\in\Delta} c(g)\,\Gamma g \Gamma;\ c(g) \in \mathbb{Z},\ c(g) \neq 0 \text{ for finitely many } g \right\} . \tag{4.4}$$

On $\mathcal{R}(\Gamma,\Delta)$ a multiplication will be introduced. To this end consider the decompositions (3.61)

$$\begin{aligned} \Gamma g_1 \Gamma &= \Gamma\alpha_1 \cup \ldots \cup \Gamma\alpha_k , \\ \Gamma g_2 \Gamma &= \Gamma\beta_1 \cup \ldots \cup \Gamma\beta_l , \end{aligned} \tag{4.5}$$

corresponding to two elements $g_1, g_2 \in \Delta$. For every $g \in \Delta$ define the integer

$$m(g) := \#\{(i,j);\ \Gamma\alpha_i\beta_j = \Gamma g\} . \tag{4.6}$$

The product of the cosets $\Gamma g_1 \Gamma$ and $\Gamma g_2 \Gamma$ is then defined as

$$\Gamma g_1 \Gamma \cdot \Gamma g_2 \Gamma := \sum_{g\in\Delta} m(g) \Gamma g \Gamma . \tag{4.7}$$

One can now show [86, 71] that this product does not depend on the choice of the representatives $\{\alpha_i\}$ and $\{\beta_j\}$ in (4.5), and that it is associative. Furthermore, $\mathbb{1} \in \Delta$ yields the identity element $\Gamma\mathbb{1}\Gamma$ for this multiplication. Extending the law of multiplication (4.7) linearly to all of $\mathcal{R}(\Gamma,\Delta)$ turns this module into an associative ring with identity, which is called the *Hecke ring of* Γ *with respect to* Δ. Choosing $\Delta = \overline{\Gamma}$, $\mathcal{R}(\Gamma) := \mathcal{R}(\Gamma,\overline{\Gamma})$ is named the *Hecke ring* of Γ. In general, the Hecke ring need not be commutative, but there exists a sufficient criterion for its commutativity, see [86, 71]: if there exists a one-to-one mapping $\iota : \Delta \to \Delta$ such that $(g_1 g_2)^\iota = g_2^\iota g_1^\iota$, $\Gamma g^\iota \Gamma = \Gamma g \Gamma$ for all $g \in \Delta$, and $\Gamma^\iota = \Gamma$, then the Hecke ring $\mathcal{R}(\Gamma,\Delta)$ is commutative. Several examples of arithmetic groups have been dealt with in the literature that actually do result in commutative Hecke rings, the most prominent ones, as always, being the modular group and its congruence subgroups, see [86, 96, 71]. Also the Hecke rings of arithmetic groups that are obtained from unit groups of maximal orders [85] or of orders of the Eichler type of level N [71] in indefinite quaternion algebras over $\mathbb{Q}$ are known to be commutative.

The Hecke ring $\mathcal{R}(\Gamma,\Delta)$ can be represented on various linear spaces; see [49, 86, 71] for examples not treated here. The one we are interested in is the Hilbert space $L^2(\Gamma\backslash\mathcal{H})$. Let therefore be $\psi \in L^2(\Gamma\backslash\mathcal{H})$ and define for $g \in \Delta$

$$T(g)\,\psi(z) := \sum_{i=1}^{n} \psi(\alpha_i z) , \tag{4.8}$$

where the fractional linear transformations $\alpha_1, \ldots, \alpha_n$ are yielded from the decomposition (3.61). The bounded linear operator $T(g)$ on $L^2(\Gamma\backslash\mathcal{H})$ is called a *Hecke operator*, see [100] for proofs of its properties. Since $T(g)$ acting on the function ψ results in a linear combination of ψ taken at points translated by operations of $GL^+(2,\mathbb{R})$, the Hecke operator commutes with the hyperbolic Laplacian. In [100] one also finds that $T(g^{-1}) = T(g)^*$, where "$*$" denotes the adjoint with respect to the scalar product of $L^2(\Gamma\backslash\mathcal{H})$. Thus, $\tilde{T}(g) := T(g) + T(g^{-1})$ is self-adjoint. Moreover, the $T(g)$, $g \in \Delta$, form a representation of $\mathcal{R}(\Gamma,\Delta)$ on $L^2(\Gamma\backslash\mathcal{H})$, i.e. for $g_1, g_2 \in \Delta$, $\Gamma g_1 \Gamma \cdot \Gamma g_2 \Gamma = \sum_{g\in\Delta} m(g) \Gamma g \Gamma$, one can show that $T(g_1)T(g_2)\,\psi(z) = \sum_{g\in\Delta} m(g) T(g)\,\psi(z)$.

In case $g \in \overline{\Gamma}$ is a symmetry, i.e. a pseudosymmetry of order $n = 1$, it is an element of the group Γ' that contains Γ as a normal subgroup such that Γ'/Γ is isomorphic to the symmetry group. By (3.60) then only $\gamma_1 = \mathbb{1}$ occurs in the decomposition of Γ into cosets of $\Gamma'(g)$ $(= \Gamma)$. Thus the Hecke operator related to g only involves $\alpha_1 = g$, see (4.8). One now observes that this Hecke operator is nothing else than the representation operator of the symmetry g, see (4.1),

$$T(g)\,\psi(z) = \psi(gz) = \chi(g)\psi(z) . \tag{4.9}$$

Hecke operators therefore yield proper generalizations of symmetry operators, which is in accordance with the assertion that pseudosymmetries are generalizations of symmetries.

At this point one might seek for a geometrical interpretation of Hecke operators. In section 3.5 pseudosymmetries of order n were demonstrated to result in n-sheeted coverings of the surface $\Gamma\backslash\mathcal{H}$. The $\gamma_1, \ldots, \gamma_n$ in the decomposition (3.60) of Γ into cosets of $\Gamma'(g)$ were found to define the mappings that interchange the n sheets of the covering. Exactly these γ_i's now occur in the definition (4.8) of the Hecke operator $T(g)$ via $\alpha_i = g\gamma_i$, which therefore in some sense averages the wave function $\psi(z)$ defined on $\Gamma\backslash\mathcal{H}$ over the n sheets of the covering $\Gamma'(g)\backslash\mathcal{H} \to \Gamma\backslash\mathcal{H}$. Looking at the diagram in section 3.5 one observes that the $\varphi_g(\gamma_i z)$ are the n points on $\Gamma'(g)\backslash\mathcal{H}$ that lie over $\varphi(z) \in \Gamma\backslash\mathcal{H}$. Transforming these n points by g, $T(g)$ averages ψ over the resulting n images.

As in section 3.5 the modular group $\Gamma_{mod} = SL(2,\mathbb{Z})$ will serve as an example to illustrate the general procedure just discussed, see also [32]. An exhaustive treatment of this particular case may be found in e.g. [86, 96, 71]. The semi-group Δ needed to define the Hecke ring $\mathcal{R}(\Gamma,\Delta)$ has been introduced in (3.64) as $\Delta = \{g \in M(2,\mathbb{Z});\ \det g > 0\}$. According to the convention (3.65) to look at all $g \in \Delta$ with $\det g = n$, $n \in \mathbb{N}$, simultaneously, the Hecke operators for Γ_{mod} are defined as

$$T_n := \frac{1}{\sqrt{n}} \sum_{\substack{g\in\Delta \\ \det g = n}} T(g) , \qquad n \in \mathbb{N} , \tag{4.10}$$

after normalizing them as it is commonly done in the literature, see e.g. [96]. The decomposition (3.65) then yields the explicit form

$$T_n\psi(z) = \frac{1}{\sqrt{n}} \sum_{\substack{ad=n \\ 0\le b<d}} \psi\left(\frac{az+b}{d}\right) , \tag{4.11}$$

for $\psi(z)$ fulfilling $\psi(\gamma z) = \psi(z)$, $\gamma \in \Gamma_{mod}$. One can prove several properties for the Hecke operators (4.11), see e.g. [96]: they are self-adjoint on $L^2(\Gamma\backslash\mathcal{H})$ and form a commutative algebra. Their law of multiplication can be drawn from (4.6) and (4.7) as

$$T_n T_m = \sum_{d|(n,m)} T_{\frac{nm}{d^2}} , \qquad n, m \in \mathbb{N} . \tag{4.12}$$

Since the T_n's, for $n \in \mathbb{N}$, and the hyperbolic Laplacian all commute with one another, they can be simultaneously diagonalized. The square-integrable eigenfunctions of the Laplacian for Γ_{mod}, the so-called *Maaß cusp forms*, can be expanded on $\mathcal{H}$ as

$$\psi(z) = N \sum_{k\neq 0} c(k) \sqrt{y} K_{ip}(2\pi|k|y)\, e^{2\pi i k x} , \tag{4.13}$$

see [96]. $K_\nu(t)$ denotes the modified Bessel function, and p is the momentum related to the eigenvalue E, $-\Delta\psi(z) = E\psi(z)$, by $E = p^2 + \frac{1}{4}$. The $c(k)$'s are the Fourier expansion coefficients, and N is a normalization factor. The surface $\Gamma_{mod}\backslash\mathcal{H}$ possesses one symmetry that can be realized in the coordinates employed to derive (4.13) as $z \mapsto -\bar{z}$. Accordingly, the eigenfunctions of the type (4.13) can be distinguished as having positive or negative parity under this symmetry operation, $\psi_\pm(z) = \pm\psi_\pm(-\bar{z})$. Thus the expansion of $\psi_+(z)$ contains $\cos(2\pi kx)$ and the one of $\psi_-(z)$ contains $\sin(2\pi kx)$ replacing $e^{2\pi ikx}$ in (4.13). In addition, the sum then extends only over positive k's. The coefficients $c(k)$ will now be chosen in order to yield simultaneous eigenfunctions of all the Hecke operators T_n, $n \in \mathbb{N}$, $T_n\psi(z) = t_n\psi(z)$, $t_n \in \mathbb{R}$. Choosing the normalization N such that $c(1) = 1$, which is possible since it is known that $c(1) \neq 0$, results in the identification $c(k) = t_k$, see [96]. One thus observes the interesting property that the coefficients of the Fourier expansions (4.13) of eigenfunctions of the Laplacian can be chosen to be the eigenvalues of the Hecke operators. This realization immediately yields from (4.12) the constraints

$$c(n)\, c(m) = \sum_{d|(n,m)} c\left(\frac{nm}{d^2}\right) , \tag{4.14}$$

which are named *Hecke relations*, for the Fourier coefficients of the eigenfunctions. There are further properties of these coefficients and of their statistics either known, conjectured, or numerically investigated, see [51, 96, 53, 91] for further references.

It is also possible to derive a trace formula for Hecke operators in close analogy to the Selberg trace formula. The former already occurs in Selberg's original article [84] and has been further explored in [51]. The first numerical computations with this trace formula that yield the lower Fourier coefficients for some cusp forms (4.13) of definite parity have appeared in [28].

The Hecke relations allow to express the Fourier coefficients $c(k)$ as polynomials in the coefficients $c(p)$ for primes $p \leq k$. Hence the coefficients are not independent, as it would be required by RMT for wave functions of classically chaotic systems [35]. There it is asserted that the expansion coefficients of wave functions in a generic basis are Gaussian random variables. The correlations induced by (4.14) are, however, known to be weak in the sense that [54]

$$C_N(l) := \frac{1}{N} \sum_{k\leq N} c(k+l)\, c(k) = O(N^{-\frac{1}{3}+\epsilon}) , \tag{4.15}$$

for all $\epsilon > 0$ and every $l \in \mathbb{N}$, leading to $\lim_{N\to\infty} C_N(l) = 0$. Although the expansion (4.13) uses a special basis, the Hecke relations manifest themselves in the wave functions irrespective of any basis because the values of $\psi(z)$, being an eigenfunction of every T_n, are related by

T_n at different points (see (4.11)) for every $n \in \mathbb{N}$. Berry's *random wave conjecture* [19] now asserts that the values of wave functions for classically chaotic systems become, in the semiclassical limit, Gaussian random variables. The existence of Hecke operators inducing relations on the values of wave functions seems to indicate a violation of this randomness assumption. Hejhal and Rackner [54], however, computed numerically several eigenfunctions of the hyperbolic Laplacian on $\Gamma_{mod}\backslash\mathcal{H}$ and found good agreement with the random wave conjecture. It therefore seems that the correlations in the eigenfunctions of arithmetic systems induced by the presence of the infinitely many pseudosymmetries do not suffice to violate the conjectured random character for wave functions in generic classically chaotic systems. This observation is in contrast to the case of symmetries, where it is known that in order to obtain a generic behaviour one has to desymmetrize the systems first.

Although the presence of the Hecke operators seemingly does not cause an exceptional behaviour of the eigenfunctions, the pseudosymmetries they have been derived from could have an influence on the distribution of the eigenvalues. Therefore a final remark concerning the spectrum of the hyperbolic Laplacian on an arithmetic surface shall conclude this section. The infinitely many non-trivial pseudosymmetries of the surface $\Gamma\backslash\mathcal{H}$ lead to infinitely many constraints for the eigenvalue spectrum $\sigma_p(\Gamma) = \{E_0 < E_1 \leq E_2 \leq \ldots\}$ of $-\Delta$ on $\Gamma\backslash\mathcal{H}$. For every $g \in \tilde{\Gamma}$ the Fuchsian group $\Gamma'(g)$ is a subgroup of finite index in Γ to which there is related the eigenvalue problem of $-\Delta$ on $\Gamma'(g)\backslash\mathcal{H}$. Denoting these spectra by $\sigma_p(\Gamma'(g))$, the inclusions $\Gamma'(g) \subset \Gamma$ of Fuchsian groups result in inclusions $\sigma_p(\Gamma) \subset \sigma_p(\Gamma'(g))$ of the respective spectra. In order to proof these inclusions one recalls that $E \in \sigma_p(\Gamma)$ is equivalent to the existence of a Maaß waveform $\psi : \mathcal{H} \to \mathbb{C}$ fulfilling

1. $-\Delta\psi(z) = E\psi(z)$,
2. $\psi(\gamma z) = \psi(z)$ for all $\gamma \in \Gamma$,
3. $\int_{\Gamma\backslash\mathcal{H}} \frac{dx\,dy}{y^2}\, |\psi(z)|^2 < \infty$.

The assertion that then also $E \in \sigma_p(\Gamma'(g))$ is proved once one can show that $\psi(z)$ obeys 1. – 3. with $\Gamma'(g)$ replacing Γ for all $g \in \tilde{\Gamma}$: 1. is trivial, and 2. is also obviously true, since $\Gamma'(g) \subset \Gamma$. The requirement that $\Gamma'(g)$ is of finite index n in Γ results in $\text{area}(\Gamma'(g)\backslash\mathcal{H}) = n \cdot \text{area}(\Gamma\backslash\mathcal{H}) < \infty$, so that the integral in 3. gets multiplied by n when replacing Γ through $\Gamma'(g)$ and thus remains finite. In conclusion, the discrete spectrum $\sigma_p(\Gamma)$ for an arithmetic group Γ is a subspectrum of infinitely many spectra $\sigma_p(\Gamma'(g))$,

$$\sigma_p(\Gamma) \subseteq \bigcap_{g\in\tilde{\Gamma}} \sigma_p(\Gamma'(g)) . \tag{4.16}$$

The groups $\Gamma'(g)$ are themselves arithmetic, since they are commensurable with the (arithmetic) group Γ. Arithmetic spectra are thus constrained by obeying infinitely many inclusions. The question now is whether these constraints are strong enough in order to yield exceptional statistical properties of discrete energy spectra related to arithmetic groups. The following sections of the present chapter are devoted to trying to answering this question in the affirmative. The methods employed there, however, do not rely on these constraints but rather use the classical properties of arithmetical systems as starting points, since it seems to be difficult to formulate the constraints in a way that enables one to use them in a quantitative analysis.

4.2 Spectral Statistics and the Form Factor

Following the general belief on the existence of a universal classification for the statistical properties of discrete quantum energy spectra according to the characters of the corresponding classical systems it is expected that time-reversal invariant systems with chaotic classical limits possess spectra that can be described by the GOE random matrix ensemble up to a maximal scale L_{max}. Berry's theory for the spectral rigidity supporting this assumption on scales $1 \ll L \ll L_{max}$, however, presupposes that the multiplicities of lengths of periodic orbits asymptotically approach two for long orbits. The discussion of geodesic flows on hyperbolic surfaces with arithmetic Fuchsian groups Γ in chapter 3, however, revealed that for these systems the mean multiplicities of lengths grow exponentially, $< g_p(l) > \sim c_\Gamma \frac{e^{l/2}}{l}$, $l \to \infty$. This property will certainly influence the spectral rigidity and therefore the medium- and long-range correlations in the respective quantum energy spectra. Conjecturally, then also the short-range correlations, in particular the level spacings distributions, will be affected, since the universal behaviour of the spectral statistics for generic systems holds, according to empirical observations, on all scales up to L_{max}. The question then remains what spectral statistics the arithmetical systems share?

To get an idea what the answer might look like one can recall what is known about the function $S(p) = \frac{1}{\pi} \arg Z(\frac{1}{2} + ip) = N_{fl}(E(p))$ for arithmetic as well as for non-arithmetic Fuchsian groups, and compare this with the corresponding results for classically integrable systems. For general cocompact Fuchsian groups the best known asymptotic upper bound is $N_{fl}(E) = O(\sqrt{E}/\log E)$, whereas the best lower bound is only $N_{fl}(E) = \Omega_\pm(\sqrt{\frac{\log E}{\log\log E}})$, see [51] and especially *Theorem 7.10* in [50]. One hence observes a rather large gap between the upper and the lower bound. This can be understood once one consults the lower bound valid for arithmetic groups. In section 3.6 this lower bound was already employed and it was remarked there that although in [51] this was only proved for a certain class of arithmetic groups, the result extended to every arithmetic group, since it were the exponential degeneracies in the respective geodesic length spectra that were responsible for this lower bound of $N_{fl}(E) = \Omega_\pm(E^{\frac{1}{4}}/\log E)$. There hence remains for arithmetic groups only a much more modest gap to the upper bound. Berry's result on the spectral rigidity for generic systems, $\Delta_\infty(E) \sim \frac{1}{2\pi^2} \log E$, now suggests that in the non-arithmetic case $|N_{fl}(E)| \sim \frac{1}{\sqrt{2}\pi}\sqrt{\log E}$, thus being close to the general lower bound. It therefore appears that the non-arithmetic groups obstruct the lower bound to be improved considerably, whereas the arithmetic groups are responsible for the upper bound. Certainly, the lower bound for arithmetic groups shows that $\Delta_\infty(E) \geq const.\frac{\sqrt{E}}{(\log E)^2}$ asymptotically for $E \to \infty$, which clearly violates Berry's result, reflecting the fact that the presuppositions to apply the latter are not met. Actually, the discussion in section 4.4 yields that $\Delta_\infty(E) \sim \frac{2\bar{d}}{\pi} \frac{\sqrt{E}}{\log E}$, which is only by a factor of $\log E$ larger than the lower bound. Therefore the arithmetic case reminds more of the saturation value of the rigidity for classically integrable systems, $\Delta_\infty(E) \sim const. \sqrt{E}$.

The integrable case also indicates that it is rather the lower bound for $N_{fl}(E)$ that comes closer to its true magnitude than the upper one. Rigorous results for $N_{fl}(E)$ are known for the quantization of the geodesic flow on a torus $T = (\mathbb{R}/2\pi\mathbb{Z})^2$. The associated spectral problem is that of minus the euclidean Laplacian $\Delta_E = \partial_x^2 + \partial_y^2$ acting on doubly periodic wave functions on $\mathbb{R}^2$, $\psi(x,y) = \psi(x+2\pi, y) = \psi(x, y+2\pi)$. The spectrum is then given by $E_{nm} = n^2 + m^2$, $n, m \in \mathbb{Z}$. Since $N(E)$ is the number of points of $\mathbb{Z}^2$ inside a circle of radius $\propto \sqrt{E}$, the estimation of $N(E)$ is the classical *circle problem*, see [50] for a review. It is known that $N_{fl}(E) = O(E^{\frac{1}{3}-\delta})$ for some small values of δ, and $N_{fl}(E) = \Omega_\pm(E^{\frac{1}{4}})$. The rigidity result

is therefore in accordance with the lower bound for $N_{fl}(E)$.

Everything discussed so far has been concerned with $\Delta_3(L;E)$ for $L \to \infty$, that is with correlations in the spectra on very large scales. It seems that in this regime the arithmetical systems behave more like classically integrable ones than like generic classically chaotic ones. It would now be interesting to learn whether or not this similarity to the integrable case pertains also to smaller scales, especially for the level spacings. The first numerical calculations of quantum energies for arithmetical systems are due to Schmit, who considered a special symmetry class for the spectral problem related to the regular octagon group. In [31, 17] he obtained a level spacings distribution that revealed a level attraction somewhat weaker than for a Poissonian $P(s)$, but the -at that time expected- Wigner surmise was clearly ruled out. In addition it seemed that the computed $P(s)$ would the more approach a Poissonian distribution the higher in energy one went. Aurich and Steiner [11] then calculated eigenvalues in all symmetry classes for the regular octagon group corresponding to one dimensional representations of the symmetry group for $\Gamma_{reg}\backslash\mathcal{H}$, and obtained the same findings as Schmit did. Hejhal was the first to compute a considerable number of eigenvalues for the modular group [52] and it was observed [92] that the corresponding level spacings behaved like the ones for the regular octagon group. Later, Schmit [82] could calculate more eigenvalues for the odd symmetry class on $\Gamma_{mod}\backslash\mathcal{H}$ and further confirmed the results for the modular group. At that time, however, it remained unclear how the observed violation of the RMT hypothesis could come about and what class of systems would share alike properties. Steil then computed [91, 32] 3167 eigenvalues for the even symmetry class and 3475 eigenvalues for the odd symmetry class comprising the complete spectrum up to $p = 300$, i.e. in energy up to $E = 90\,000.25$, revealing that $P(s)$ can be rather well described for high energies, corresponding to $250 \leq p \leq 300$, by a Poissonian distribution. In [32, 27] then the explanation was given that the arithmetic properties of the Fuchsian groups Γ_{reg} and Γ_{mod} involved were responsible for the exceptional statistical properties observed, and that consequently all arithmetical systems would share alike spectral statistics. In contrast, non-arithmetic systems were considered in [12, 14, 82, 27] and found to be in good agreement with a RMT behaviour of their energy spectra. Up to now, however, no rigorous argument or quantitative heuristics could be given that explains the observed phenomena.

It is our aim to present in the following two sections a simple model that should account for the observed spectral properties of the arithmetical systems. The key quantity studied there in order to determine the level spacings distribution $P(s)$ and the number variance $\Sigma^2(L;x)$ is the *spectral form factor* $K(\tau;x)$ [20] for the unfolded spectrum $\{x_i\}$. It is defined as a Fourier transform of the pair correlation of the energy fluctuations and will be introduced below. The *pair correlation function* $g(t)$ is the two-point correlation function of the spectral density $d(x)$,

$$g(t) := < d(x - \frac{t}{2})\, d(x + \frac{t}{2}) > , \tag{4.17}$$

where $< \dots >$ denotes a semiclassical averaging, as in section 2. $g(t)$ then is the density function for the probability of finding an energy level in the interval $(0,T)$, given one at $t = 0$. This can be used to construct the level spacings distribution $P(s)$ as

$$P(s) = g(s) \exp\left[-\int_0^s dt\; g(t) \right] , \tag{4.18}$$

see e.g. Porter's contribution in [77]. Defining $G(s) := \int_0^s dt\, g(t)$, one observes that $P(s) = -\frac{d}{ds}\, e^{-G(s)}$, and therefore

$$\int_0^T ds\, P(s) = e^{-G(0)} - e^{-G(T)} , \qquad T > 0 . \tag{4.19}$$

Hence, as long as $g(s)$ is integrable on any finite interval $[0,T]$, but with $\int_0^\infty dt\, g(t) = +\infty$, $P(s)$ is a normalized probability density, $\int_0^\infty ds\, P(s) = 1$.

The interesting contribution to $d(x) = 1 + d_{fl}(x)$ now comes from its fluctuating part $d_{fl}(x)$. Inserting this splitting into (4.17) for the pair correlation function, one is left among others with two terms of the form $< d_{fl}(x \pm \frac{t}{2}) >$. It will now be argued that these vanish in the semiclassical limit. To support this idea one can go back to the regularization (2.20) for $d_{fl}(E)$. In section 2 it was demonstrated that averaging over an interval of length L, $< d_{fl}(E) >$ vanishes like L^{-1}. Choosing now $\Delta x = x^a$ for the length of the interval involved in the semiclassical averaging, which meets the prerequisit $1 \ll \Delta x \ll x$ for a small enough power $1 > a > 0$, $< d_{fl}(x \pm \frac{t}{2}) >$ behaves like x^{-a} and thus vanishes in the semiclassical limit $x \to \infty$. Hence, in this limit,

$$g(t) \sim 1 + < d_{fl}(x - \frac{t}{2})\, d_{fl}(x + \frac{t}{2}) > . \qquad (4.20)$$

$K(\tau; x)$ is now defined as the Fourier transform of the correlation function on the r.h.s. of (4.20),

$$K(\tau; x) := \int_{-\infty}^{+\infty} dt\; e^{2\pi i t \tau} < d_{fl}(x - \frac{t}{2})\, d_{fl}(x + \frac{t}{2}) > . \qquad (4.21)$$

Since this definition involves only $d_{fl}(x)$, which is related via (2.20) to the periodic orbits of the classical system, the form factor is especially suited to be used in periodic-orbit theory. It only remains to reexpress the pair correlation function $g(t)$ in terms of $K(\tau; x)$ in order to be prepared for a periodic-orbit analysis of the level spacings distribution $P(s)$. To this end one inserts $\delta(x) = \int_{-\infty}^{+\infty} dz\; e^{2\pi i x z}$ into the r.h.s. of the identity

$$< d_{fl}(x - \frac{t}{2})\, d_{fl}(x + \frac{t}{2}) > = \frac{1}{2t} \int_{-\infty}^{+\infty} dy\; y < d_{fl}(x - \frac{y}{2})\, d_{fl}(x + \frac{y}{2}) > [\delta(y - t) - \delta(y + t)] , \qquad (4.22)$$

yielding after some simple manipulations

$$\begin{aligned} < d_{fl}(x - \frac{t}{2})\, d_{fl}(x + \frac{t}{2}) > \;&=\; -\frac{1}{2\pi t} \int_{-\infty}^{+\infty} d\tau\; \sin(2\pi t \tau) \cdot \\ &\cdot \frac{\partial}{\partial \tau} \int_{-\infty}^{+\infty} dy\; e^{2\pi i \tau y} < d_{fl}(x - \frac{y}{2})\, d_{fl}(x + \frac{y}{2}) > . \end{aligned} \qquad (4.23)$$

Using (4.20) and (4.21) then leads to

$$g(t) \sim 1 - \frac{1}{\pi t} \int_0^\infty d\tau\; \sin(2\pi t \tau) \frac{\partial}{\partial \tau} K(\tau; x) , \qquad (4.24)$$

which is a fundamental expression to be used in the following section.

In order to derive a periodic-orbit expression for the form factor one has to rescale the unfolded spectrum by $d_{fl}(x) = \frac{dE}{dx}\frac{d}{dE} N_{fl}(E) = \frac{1}{\bar{d}(E)} d_{fl}(E)$. One can then use (2.20) and insert it into the definition of $K(\tau; x)$ expressed in terms of E,

$$K(\tau; E) = \frac{1}{\bar{d}(E)} \int_{-\infty}^{+\infty} d\lambda\; e^{2\pi i \tau \bar{d}(E) \lambda} < d_{fl}(E - \frac{\lambda}{2})\, d_{fl}(E + \frac{\lambda}{2}) > . \qquad (4.25)$$

Introducing the momentum variable p for convenience, (2.20) for the hyperbolic Laplacian on a surface $\Gamma\backslash\mathcal{H}$ reads

$$\begin{aligned} d_{\varepsilon,fl}(p) \;&=\; \frac{1}{\pi} \sum_{\{l_n\}} \sum_{k=1}^{\infty} A_{n,k} \cos(pkl_n)\, e^{-\frac{\varepsilon^2}{4} k^2 l_n^2} , \\ A_{n,k} \;&:=\; \frac{l_n}{2p} \frac{\chi_n^k\, g_p(l_n)}{\left| e^{\frac{k}{2} l_n} - \sigma_n^k e^{-\frac{k}{2} l_n} \right|} . \end{aligned} \qquad (4.26)$$

The outer sum runs over all distinct lengths l_n of primitive closed geodesics, whereby their respective multiplicities have been incorporated in the amplitude factors $A_{n,k}$. The possibility to include *inverse hyperbolic orbits* has been left open by allowing for $\sigma_n = -1$. As may be drawn from the Selberg trace formula (3.71) ordinary closed geodesics corresponding to hyperbolic $\gamma \in \Gamma$ are *hyperbolic orbits*, i.e. possessing $\sigma_\gamma = 1$. The consideration of inverse hyperbolic orbits is necessary in order to be able to treat Artin's billiard, which is obtained from $\Gamma_{mod}\backslash\mathcal{H}$ by dividing out the orientation reversing symmetry $z \mapsto -\bar{z}$, thus yielding a billiard system on the hyperbolic plane $\mathcal{H}$. Choosing the standard fundamental domain $\mathcal{F}_{mod} = \{z \in \mathcal{H};\ |z| \geq 1,\ -\frac{1}{2} \leq x \leq \frac{1}{2}\}$ for the modular group, Artin's billiard takes place on $\mathcal{F}_A = \{z \in \mathcal{F}_{mod};\ x \geq 0\}$. Closed geodesics on $\Gamma_{mod}\backslash\mathcal{H}$ that are invariant under the reflection $z \mapsto -\bar{z}$ then result in inverse hyperbolic orbits on the billiard domain, which are reflected an odd number of times on $\partial\mathcal{F}_A$, see [68] for further information. In addition, representations $\chi: \Gamma \longrightarrow \{\pm 1\}$ for the Fuchsian groups Γ are admitted in (4.26). For Artin's billiard it turns out that $\chi_n = \sigma_n$. In (4.26) it has further been assumed that all $g_p(l_n)$ geodesics γ that share the same length l_n have identical $\chi_\gamma = \chi_n$ and $\sigma_\gamma = \sigma_n$, respectively. In Artin's billiard this requirement is indeed met [68].

Using the regularized fluctuating part of the spectral density (4.26) in (4.25), and performing the limit $\varepsilon \to 0$ at the end of the calculation, leads to

$$K(\tau; E) \sim \frac{1}{(2\pi \bar{d}(E))^2} < \sum_{\{l_n\},\{l_m\}} \sum_{r,s}{}' A_{n,r} A_{m,s}\, e^{ip(rl_n - sl_m)}\, \delta(\tau - \frac{rl_n + sl_m}{8\pi p \bar{d}(E)}) > , \qquad (4.27)$$

see [20] and Berry's contribution in [41]. The inner sums extend over all non-zero integers, the prime indicating the omission of $r, s = 0$.

The r.h.s. of (4.27) consists of two double sums, each running over the primitive closed geodesics and their repetitions. The diagonal contribution of these two double sums is

$$K_D(\tau; E) = \frac{1}{(2\pi \bar{d}(E))^2} \sum_{\{l_n\}} \sum_{k=1}^{\infty} A_{n,k}^2\, \delta(\tau - \frac{kl_n}{4\pi p \bar{d}(E)}) . \qquad (4.28)$$

Naively one would expect the non-diagonal terms in (4.27) to be washed out by the semiclassical averaging. This to work would require large enough "random" phases $e^{ip(rl_n - sl_m)}$. However, by going to longer orbits different lengths l_n lie closer and closer due to their exponentially increasing density. This effect can be drawn from the exponential growth of $\hat{N}_p(l)$. Thus, for long orbits, the random phase argument fails, and the diagonal approximation (4.28) is only reasonable up to some τ^*. For small values of τ, $\tau \leq \tau^*$, only short orbits contribute to (4.27), and these are well separated in length. In [20, 41] Berry demonstrates that for $\tau \to \infty$ the form factor saturates, $K(\tau; E) \to 1$. He further claims that even for $\tau > 1$ one obtains $K(\tau; E) \approx 1$. (Notice that by (4.28) already $\tau = 1$ corresponds in the semiclassical limit to long orbits, $l = 4\pi p \bar{d}(E)$, and thus $\tau^* \ll 1$.) It should be remarked, however, that very recently Aurich and Sieber [8] found a violation of this saturation for the form factor in cases where there exists an eigenvalue $E_0 = 0$ for the hyperbolic Laplacian. They obtain an exponentially increasing $K(\tau; E)$ for $\tau \to \infty$ instead. In such cases the contribution to (4.27) coming from the zero mode has to be subtracted using the trace formula, see [8].

For small values of τ, $\tau \leq \tau^*$, the diagonal approximation (4.28) indicates that δ-spikes determined by individual lengths l_n will characterize the form factor. In this regime $K(\tau; E)$ therefore behaves non-universally. The intermediate range $\tau^* < \tau \leq 1$ is then governed by the contribution of rather long orbits. According to Berry [20, 41] a sum rule of Hannay and Ozorio de Almeida [47], which exploits the uniform exploration of phase space by long orbits, yields $K(\tau;E) \approx \overline{g}\tau$ for $\tau^* < \tau \leq 1$. Thereby it is assumed that the multiplicities of lengths of primitive orbits approach $\overline{g} = const.$ asymptotically for long orbits. For generic time-reversal invariant systems thus $\overline{g} = 2$, whereas generic systems without time-reversal symmetry show $\overline{g} = 1$. Obviously this sum rule cannot be applied to arithmetic systems because the exponentially increasing multiplicities of lengths prohibit $< g_p(l) >$ to approach a constant. One would, however, expect from the sum rule that $K(\tau;E)$ grows much faster for the arithmetic systems than generically, see also [27]. This observation lies at the heart of the model that will be presented in the next two sections.

4.3 A Model for the Level Spacings Distribution

It will now be attempted to set up a model describing the statistical properties of the eigenvalues of hyperbolic Laplacians on surfaces with arithmetic Fuchsian groups. The main tool to be employed will be the spectral form factor $K(\tau;E)$. Then, as explained in the preceding section, the relations (4.18) and (4.24) allow to determine the level spacings distribution $P(s)$ once $K(\tau;E)$ is known in sufficient detail. Following Berry's reasoning as reviewed in section 4.2, the form factor can be substituted by its diagonal approximation $K_D(\tau;E)$ (4.28) for $\tau \leq \tau^* \ll 1$. From now on τ^* will be fixed at some value so that $K_D(\tau;E)$ approximates the complete form factor sufficiently well. Then only $K_D(\tau;E)$ will be used for $\tau \leq \tau^*$.

The semiclassical limit $E \to \infty$ can for fixed τ also be viewed as the limit of long orbits, $l \to \infty$, as can be drawn e.g. from (4.28). Thus the sum in (4.28) will be evaluated in the asymptotic regime $l \to \infty$. In particular the multiplicities $g_p(l_n)$ appearing in the amplitude factors $A_{n,k}$ (4.26) will be replaced by their asymptotic values $c_\Gamma \frac{e^{l_n/2}}{l_n}$. Therefore one obtains

$$A_{n,k} \sim \frac{c_\Gamma}{2p} \chi_n^k e^{\frac{1}{2}(1-k)l_n} [1 + O(e^{-kl_n})] , \qquad l_n \to \infty . \tag{4.29}$$

As always when dealing with periodic-orbit sums one observes also here that the $(k = 1)$-contribution to the sum over repetitions of primitive orbits is the leading one for $l_n \to \infty$. Thus, asymptotically in the semiclassical limit, one finds that

$$K_D(\tau;E) \sim \frac{c_\Gamma^2}{(4\pi p\overline{d}(E))^2} \sum_{\{l_n\}} \delta(\tau - \frac{l_n}{4\pi p\overline{d}(E)}) . \tag{4.30}$$

To get rid of the hardly tractable Dirac-δ's one integrates (4.30),

$$\int_0^\tau dt\ K_D(t;E) \sim \frac{c_\Gamma^2}{(4\pi p\overline{d}(E))^2} \hat{\mathcal{N}}_p(4\pi p\overline{d}(E)\tau) . \tag{4.31}$$

Introducing the asymptotic behaviour $\hat{\mathcal{N}}_p(l) \sim \frac{2}{c_\Gamma} e^{l/2}$, $l \to \infty$, and differentiating the result with respect to τ yields

$$K_D(\tau;E) \sim \frac{c_\Gamma}{4\pi p\overline{d}(E)} e^{2\pi p\overline{d}(E)\tau} , \tag{4.32}$$

compare also [27]. Relation (4.32) shows that the diagonal approximation $K_D(\tau;E)$ grows exponentially and already at a value of $\tau_0 := \frac{1}{2\pi p\overline{d}(E)} \log(\frac{4\pi}{c_\Gamma} p\overline{d}(E))$ it has reached the value one. τ_0 being a function of E decreases in the semiclassical limit and above some energy $E = \hat{E}$ it is smaller than the fixed value τ^*. Hence $K_D(\tau;E)$ can be taken as an approximation for the complete form factor $K(\tau;E)$ in the whole range $[0,\tau_0]$ once the energy is chosen sufficiently

large, $E \geq \hat{E}$. In conclusion, the exponential behaviour (4.32) semiclassically describes the complete form factor in the mean up to τ_0.

According to Berry's investigation of the form factor [20, 41] this approaches one for $\tau \to \infty$. Since even for $\tau = 1$ mainly long orbits contribute, he argues that $K(\tau;E) \approx 1$ for $\tau \geq 1$, and then takes $K(\tau;E) \equiv 1$ to model the actual form factor in this domain, leading to his result for the spectral rigidity. In the sequel we will proceed analogously and define a model form factor

$$K_M(\tau;E) := \begin{cases} \frac{c_\Gamma}{4\pi p\overline{d}(E)} e^{2\pi p\overline{d}(E)\tau} & , \quad \tau \leq \tau_0 \\ 1 & , \quad \tau > \tau_0 \end{cases} . \tag{4.33}$$

In addition to Berry's reasoning $K_M(\tau;E)$ is defined to be one also on the interval $[\tau_0, 1]$. This may be justified by the finding that $K_D(\tau;E)$ already reaches the value one at τ_0, whereas in the generic case Berry considers the form factor has to be interpolated on the interval $[\tau^*, 1]$ by the result obtained from the sum rule of [47], as described in section 4.2. Put in sloppy terms, the exponential increase (4.32) somehow exempts one from the need to discuss the interval $[\tau^*, 1]$ separately. Inspecting (4.30) one notes that the definition (4.33) of $K_M(\tau;E)$ results in cutting-off the periodic-orbit sum at $l_{max} = 4\pi p\overline{d}(E)\tau_0 = 2\log(\frac{4\pi}{c_\Gamma} p\overline{d}(E))$. In the semiclassical limit the cut-off is therefore being removed automatically. Certainly, the actual form factor will oscillate about the mean value described by (4.33). This fine structure is left out in the model and thus one cannot really expect the resulting model distribution $P_M(s)$ to describe the actual level spacings in full detail. However, the model should reproduce the facts at least qualitatively.

Inserting (4.33) into (4.24) yields the model pair correlation function

$$\begin{aligned} g_M(t) &= 1 - \frac{c_\Gamma}{2\pi t} \int_0^{\tau_0} d\tau\ \sin(2\pi t\tau)\, e^{2\pi p\overline{d}(E)\tau} \\ &= 1 - \frac{1}{\pi t} \frac{p\overline{d}(E)}{(p\overline{d}(E))^2 + t^2} \left\{ p\overline{d}(E) \sin(2\pi t\tau_0) - t\cos(2\pi t\tau_0) + \frac{c_\Gamma}{4\pi} \frac{t}{p\overline{d}(E)} \right\} \\ &= 1 - \frac{1}{\pi} \frac{p\overline{d}(E)}{(p\overline{d}(E))^2 + t^2} \left\{ \log(\frac{4\pi}{c_\Gamma} p\overline{d}(E)) - 1 + \frac{c_\Gamma}{4\pi} \frac{1}{p\overline{d}(E)} + O((\frac{\log p\overline{d}(E)}{p\overline{d}(E)})^2) \right\} . \end{aligned} \tag{4.34}$$

Integrating this result, $G_M(s) = \int_0^s dt\, g_M(t)$, yields

$$G_M(s) \sim s - \frac{1}{\pi} \left\{ \log(\frac{4\pi}{c_\Gamma} p\overline{d}(E)) - 1 + \frac{c_\Gamma}{4\pi} \frac{1}{p\overline{d}(E)} \right\} \arctan(\frac{s}{p\overline{d}(E)}) . \tag{4.35}$$

The model level spacings distribution is then obtained as $P_M(s) = g_M(s)\, e^{-G_M(s)}$,

$$\begin{aligned} P_M(s) \sim & \left\{ 1 - \frac{1}{\pi} \frac{p\overline{d}(E)}{(p\overline{d}(E))^2 + s^2} \left[\log(\frac{4\pi}{c_\Gamma} p\overline{d}(E)) - 1 + \frac{c_\Gamma}{4\pi} \frac{1}{p\overline{d}(E)} \right] \right\} \cdot \\ & \cdot \exp\left\{ -s + \frac{1}{\pi} \left[\log(\frac{4\pi}{c_\Gamma} p\overline{d}(E)) - 1 + \frac{c_\Gamma}{4\pi} \frac{1}{p\overline{d}(E)} \right] \arctan(\frac{s}{p\overline{d}(E)}) \right\} . \end{aligned} \tag{4.36}$$

Consulting (4.19), and $G_M(s) \to \infty$ for $s \to \infty$, one obtains that $P_M(s)$ is a normalized distribution. Moreover, for $E \to \infty$ one finds that $g_M(s) \to 1$ and $G_M(s) \to s$, leading to the observation that in the semiclassical limit $P_M(s)$ approaches a Poissonian distribution. Since $P_M(s)$ is designed to reproduce the actual level spacings distribution for $E \to \infty$, also the latter is expected to converge to a Poissonian.

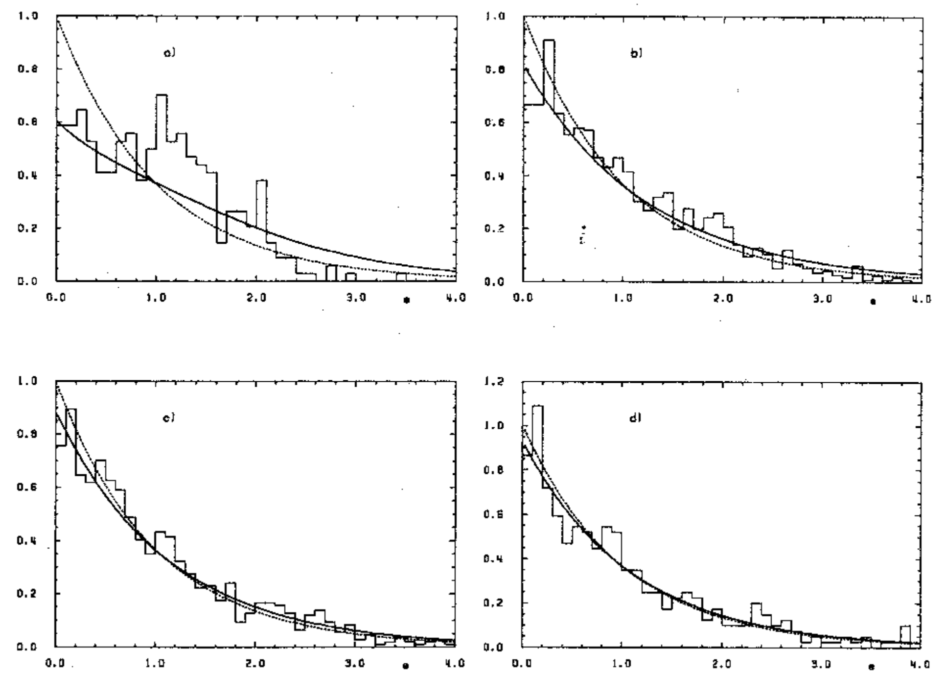

Figure 3: The model $P_M(s)$ applied to the odd symmetry class of Artin's billiard is shown as the full curves compared to the actual level spacings distributions in the intervals a) $0 \le p \le 100$, b) $100 \le p \le 200$, c) $250 \le p \le 300$, and d) $500 \le p \le 510$. The dotted curves show Poissonian distributions.

The model shows a level attraction, which is for finite E weaker than that one of a pure Poissonian spectrum, since for $s \to 0$ one finds

$$\begin{aligned} P_M(s) &\sim g_M(0) + \left[g_M'(0) - g_M(0)^2\right] s \\ &\sim 1 - \frac{1}{\pi p\bar{d}(E)}\left[\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E)) - 1 + \frac{c_\Gamma}{4\pi}\frac{1}{p\bar{d}(E)}\right] \\ &\quad - \left\{1 - \frac{1}{\pi p\bar{d}(E)}\left[\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E)) - 1 + \frac{c_\Gamma}{4\pi}\frac{1}{p\bar{d}(E)}\right]\right\}^2 s \,. \end{aligned} \tag{4.37}$$

Thus $P_M(0) < 1$ for finite E, and also $0 > P_M'(0) > -1$. In the semiclassical limit the strength of the level attraction then approaches the Poissonian result $P(s) \sim 1 - s$. The model hence qualitatively reproduces the numerical findings in [31, 17, 11, 32, 27] correctly.

Using the modular group as an example it will now be studied how well the model describes the actual level spacings distribution of an arithmetical system quantitatively. The energy

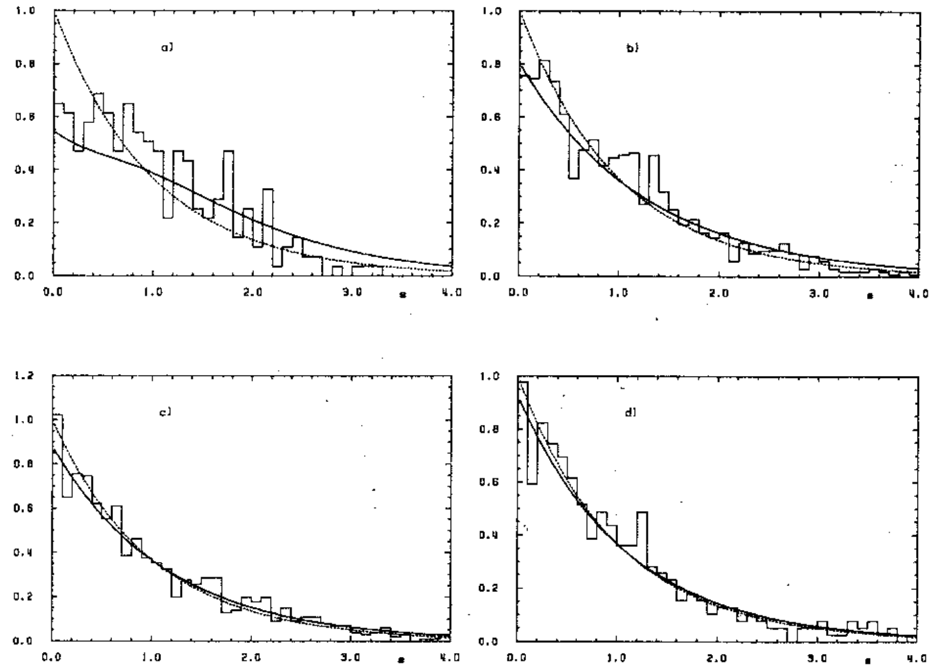

Figure 4: The same as in fig.3, but for the even symmetry class.

eigenvalues for the two symmetry classes occurring for the modular group have been kindly placed at disposal by Gunther Steil, see also [91]. The mean spectral density for the odd symmetry class of Artin's billiard reads [68]

$$\bar{d}_-(E) \sim \frac{1}{24} - \frac{1}{8\pi}\frac{\log E}{\sqrt{E}} - \frac{3\log 2}{8\pi}\frac{1}{\sqrt{E}} \,, \tag{4.38}$$

and for the even symmetry class

$$\bar{d}_+(E) \sim \frac{1}{24} - \frac{3}{8\pi}\frac{\log E}{\sqrt{E}} - \frac{\log 2 - 4\log\pi}{8\pi}\frac{1}{\sqrt{E}} \,. \tag{4.39}$$

As a final input the constant $c_\Gamma = 1$ of $< g_p(l) > \sim c_\Gamma \frac{e^{l/2}}{l}$ is needed. Fig.3 presents histograms of the level spacings in the odd symmetry class for the four momentum-intervals $0 \le p \le 100$, $100 \le p \le 200$, $250 \le p \le 300$, and $500 \le p \le 510$, comprising of 341, 1157, 1093, and 409 levels, respectively. The momenta p that have been used in the model (4.36), which is shown as the full curves, have been chosen in the middle of each interval, i.e. $p = 50$, $p = 150$, $p = 275$, and $p = 505$, respectively. One now observes from fig.3 that the model $P_M(s)$ approximates the actual distributions reasonably well, and that the quality of the approximation grows with energy. Especially, the strength $P(0)$ of the level attraction is reproduced rather well, even in the low energy regime. Apparently the histograms as well as the model the more approach a Poissonian distribution the higher in energy one goes. Fig.4 contains the same information as fig.3, but for the even symmetry class. Here the respective momentum-intervals contain 277, 1040, 1026, and 395 levels, respectively. From fig.4 one observes the same level of agreement of the model with the data as for the odd symmetry class.

4.4 A Model for the Number Variance

The discussion in the preceding section has revealed that the statistical properties of quantum energy spectra in arithmetical chaos can on small scales and in the semiclassical limit be described by those of a Poissonian spectrum. The present section now provides a continuation of that study to medium- and long-range correlations in arithmetical quantum energy spectra. In chapter 2 the spectral rigidity $\Delta_3(L;x)$ has been introduced as a means to investigate spectral statistics on scales $L > 1$ that take several levels into account. For the following, however, it proves useful to study a different quantity that principally provides the same information on spectral correlations; this is the *number variance* $\Sigma^2(L;x)$, defined as the variance of the distribution of the numbers $n(L;x) = N(x+L) - N(x)$ of levels in intervals $[x, x+L]$,

$$\Sigma^2(L;x) := \langle [n(L;x) - L]^2 \rangle , \tag{4.40}$$

where $\langle \dots \rangle$ as usual denotes a semiclassical averaging over x. A truely Poissonian spectrum excels by a linear number variance, $\Sigma^2(L;x) = L$, whereas the GOE in RMT possesses a $\Sigma^2(L;x)$ that is asymptotically given by $\Sigma^2(L;x) \sim L$ for $L \to 0$, and by $\Sigma^2(L;x) \sim \frac{2}{\pi^2}[\log(2\pi L) + \gamma + 1 - \frac{\pi^2}{8}]$ for $L \to \infty$; here γ denotes Euler's constant. Further details can be found e.g. in Bohigas' contribution in [41].

In [21, 41] Berry presents a semiclassical treatment of the number variance very much in the spirit of his considerations of the spectral rigidity [20]. In the semiclassical limit he expresses the number variance through the form factor by

$$\Sigma^2(L;E) \sim \frac{2}{\pi^2} \int_0^\infty \frac{d\tau}{\tau^2} \sin^2(\pi L \tau)\, K(\tau;E) . \tag{4.41}$$

Using the conclusions Berry draws for the functional form of $K(\tau;E)$ one obtains in the intermediate L-range $1 \ll L \ll L_{max}$ that the number variance is given by the GOE result, whereas for $L \gg L_{max}$ it oscillates non-universally about a saturation value $\Sigma^2_\infty(E)$. These oscillations as well as the value of $\Sigma^2_\infty(E)$ are determined by the contributions of the short periodic orbits to the form factor. As can be drawn from the relation

$$\Delta_3(\frac{L}{2};E) = \frac{2}{L^4} \int_0^L dr\ \Sigma^2(r;E)\, [L^3 - 2L^2 r + r^3] \tag{4.42}$$

of the spectral rigidity to the number variance, the saturation values of both quantities satisfy $\Sigma^2_\infty(E) = 2\Delta_\infty(E)$, see [12]. Hence the number variance can also be used to find out about the asymptotic energy dependence of $\langle N_{fl}(E)^2 \rangle = \Delta_\infty(E) = \frac{1}{2}\Sigma^2_\infty(E)$ for $E \to \infty$.

The relation (4.41) now easily allows for an application of the model from the preceding section also to the number variance. It turns out, however, that it is possible to improve the form factor for the model a little in that the complete diagonal approximation is taken into account for $\tau \le \tau_0$. Therefore

$$\hat{K}_M(\tau;E) := \begin{cases} K_D(\tau;E) & , \ \tau \le \tau_0 \\ 1 & , \ \tau > \tau_0 \end{cases} , \tag{4.43}$$

with the diagonal term $K_D(\tau;E)$ taken from (4.28), will be used in (4.41) to yield a model $\Sigma^2_M(L;E)$ for the number variance in arithmetical chaos. $\Sigma^2_M(L;E)$ then consists of two contributions $\Sigma^2_{M,1}(L;E)$ and $\Sigma^2_{M,2}(L;E)$, the first one being derived from the integration in τ along $[0,\tau_0]$, and the second one resulting from the respective integration along the remaining interval $[\tau_0, \infty]$. These integrals can be performed exactly, yielding

$$\begin{aligned} \Sigma^2_{M,1}(L;E) &= \frac{8p^2}{\pi^2} \sum_{\{l_n\}} \sum_{\substack{k \ge 1 \\ k l_n \le l_{max}}} \frac{A^2_{n,k}}{(k l_n)^2} \sin^2\left(\frac{k l_n L}{4p\bar{d}(E)}\right) , \\ \Sigma^2_{M,2}(L;E) &= \frac{1}{\pi^2 \tau_0} - \frac{\cos(2\pi L \tau_0)}{\pi^2 \tau_0} - \frac{2L}{\pi} \mathrm{Si}\,(2\pi L \tau_0) + L . \end{aligned} \tag{4.44}$$

$l_{max} := 4\pi p\bar{d}(E)\tau_0$, and $\mathrm{Si}\,(x) = \int_0^x dt\, \frac{\sin t}{t}$ denotes the sine integral, see e.g. [39]. The amplitude factor $A_{n,k}$ is given by (4.26). Because of the exponential vanishing of $A_{n,k}$ for $k \ge 2$ and $l_n \to \infty$ (4.29) only the ($k=1$)–contribution from the summation over k in the periodic orbit sums will be considered for the further analytic investigations; for the numerics, however, the complete double summation of (4.44) will be used.

In the limit $L \to 0$ one can expand the sine in (4.44) and then use the asymptotic value (4.29) for $A_{n,k}$, yielding

$$\Sigma^2_{M,1}(L;E) \sim \frac{c_\Gamma^2}{8\pi^2 p^2 \bar{d}(E)^2} L^2 \hat{\mathcal{N}}_p(4\pi p\bar{d}(E)\tau_0) + O(L^4) . \tag{4.45}$$

Employing the (semiclassical) asymptotics $\hat{\mathcal{N}}_p(l) \sim \frac{2}{c_\Gamma} e^{l/2}$, $l \to \infty$, and expanding $\Sigma^2_{M,2}(L;E)$ for $L \to 0$ leaves one with

$$\Sigma^2_M(L;E) \sim L - \frac{1}{\pi p\bar{d}(E)} \left[\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E)) - 1\right] L^2 + O(L^3) . \tag{4.46}$$

Thereby the definition $\tau_0 = \frac{1}{2\pi p\bar{d}(E)} \log(\frac{4\pi}{c_\Gamma} p\bar{d}(E))$ has been employed. Thus, for $L \to 0$, the model $\Sigma^2_M(L;E)$ follows in lowest order a Poissonian number variance. For finite values of L and large enough energies, however, it is smaller than the latter.

Fig.5 presents as the dotted curves a numerical evaluation of (4.44) for the odd symmetry class of Artin's billiard, and compares these to the number variance obtained from the quantum energies computed by Steil [91]. The full curves in 5a) and c) refer to a sample of eigenvalues between the 1000^{th} and the 2000^{th} one, whereas 5b) and d) are obtained from the 2400^{th} up to the 3400^{th} eigenvalue. The respective momentum intervals are $164.92 \le p \le 229.70$ and $250.86 \le p \le 296.83$. The model (4.44) has been evaluated using momentum values from the middle of each interval, namely $p = 200$ and $p = 270$, respectively. The dashed and the dashed-dotted curves provide a comparison with Poissonian and GOE number variances, respectively. One observes from fig.5 that the model reproduces the small-L behaviour as well as the saturation of the actual number variance reasonably well. It fails, however, to describe the oscillations properly; but due to the simplicity of the assumed form factor (4.43) no better agreement of the model with reality could actually be expected. A further conclusion that can be drawn from fig.5 is that $\Sigma^2(L;E)$ leaves the Poissonian form factor already for rather small values of L. The small-L asymptotics (4.46) of $\Sigma^2_M(L;E)$, however, reveals that the (negative) coefficient of L^2 vanishes for $E \to \infty$. Thus the small-L behaviour of the model is Poissonian-like on the larger L-intervals the higher in energy one goes. This Poissonian behaviour of the number variance on small scales is in accordance with the findings about

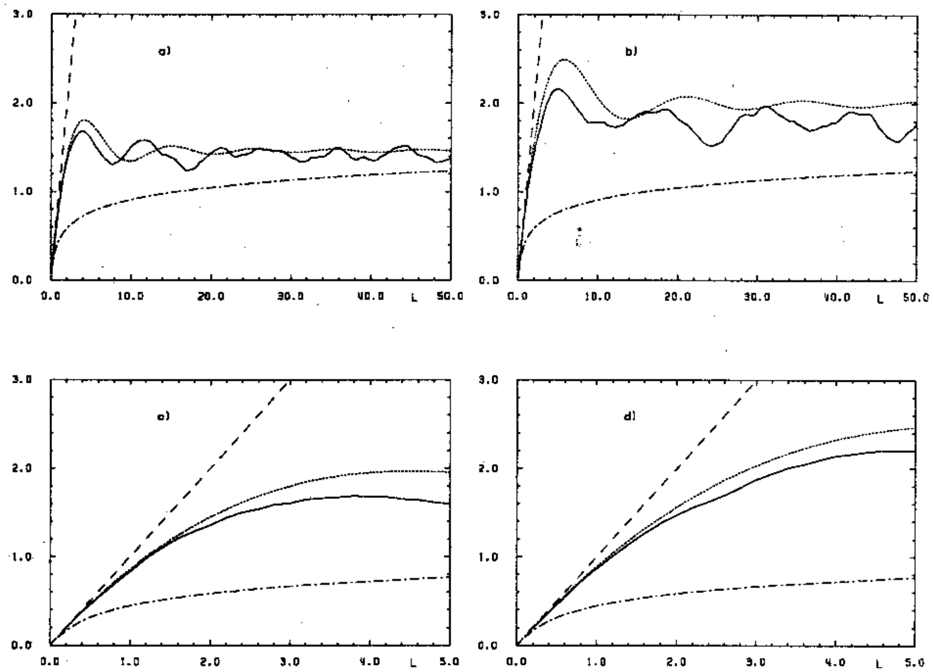

Figure 5: The number variances obtained from the eigenvalues are given as the full curves in comparison with the model, which is shown as the dotted curves. The Poissonian (dashed curves) and GOE (dashed-dotted curves) results are added. a) and c) refer to the interval $164.92 \le p \le 229.70$, b) and d), however, to $250.86 \le p \le 296.83$.

the level spacings distributions in the preceding section. Moreover, the rate of approaching a Poissonian distribution observed there is the same as the respective rate here, compare (4.46) with (4.36) and (4.37).

A further question that e.g. arises in the context of inverse quantum chaology as employed in sections 3.6 and 4.5 is the one for the energy dependence of the saturation value $\Delta_\infty(E) = \frac{1}{2}\Sigma^2_\infty(E)$. In the following therefore the limit $L \to \infty$ will be studied for the number variance $\Sigma^2_M(L;E)$ derived from the model introduced above. The second contribution $\Sigma^2_{M,2}(L;E)$ immediately yields

$$\Sigma^2_{M,2}(L;E) = \frac{1}{\pi^2 \tau_0} + O(L^{-2}) , \qquad L \to \infty , \tag{4.47}$$

when using the asymptotic behaviour $\mathrm{Si}\,(x) = \frac{\pi}{2} - \frac{\cos x}{x} + O(\frac{1}{x^2})$ for $x \to \infty$. The periodic-orbit term $\Sigma^2_{M,1}(L;E)$ contributes oscillations to the large–L asymptotics of the model number variance caused by the $\sin^2$'s. For fixed E the sum represents a superposition of finitely many

oscillations of incommensurable wave lengths. In order to obtain the average value about which this superposition of oscillations fluctuates, one replaces each $\sin^2$ by its mean value $\frac{1}{2}$. Therefore

$$\begin{aligned} < \Sigma^2_{M,1}(L;E) > \; &\sim \; \frac{c_\Gamma^2}{\pi^2} \sum_{l_n \le l_{max}} \frac{1}{l_n^2} \\ &= \; \frac{c_\Gamma^2}{\pi^2} \int_{l_1}^{l_{max}} \frac{d\hat{\mathcal{N}}_p(l)}{l^2} \, . \end{aligned} \tag{4.48}$$

Since $l_{max} = 4\pi p \bar{d}(E)\tau_0 = 2\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E)) \to \infty$ in the semiclassical limit, one can introduce the asymptotics $\hat{\mathcal{N}}_p(l) \sim \frac{2}{c_\Gamma} e^{l/2}$, $l \to \infty$, on the r.h.s. of (4.48),

$$\begin{aligned} < \Sigma^2_{M,1}(L;E) > \; &\sim \; \frac{c_\Gamma}{2\pi^2} \int_{\frac{l_1}{2}}^{\frac{l_{max}}{2}} \frac{dt}{t^2}\, e^{-t} \\ &= \; \frac{c_\Gamma}{2\pi^2}\, Ei\left(\log\left(\frac{4\pi}{c_\Gamma} p\bar{d}(E)\right)\right) - \frac{2}{\pi} \frac{p\bar{d}(E)}{\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E))} + C(l_1) \, . \end{aligned} \tag{4.49}$$

$C(l_1) := \frac{c_\Gamma}{2\pi^2}[2\frac{e^{l_1/2}}{l_1} - Ei(\frac{l_1}{2})]$ is an energy independent constant determined by the shortest primitive length l_1 on $\Gamma\backslash\mathcal{H}$. The contribution (4.47) to the saturation value coming from $\Sigma^2_{M,2}(L;E)$ now exactly cancels the second term on the r.h.s. of (4.49),

$$\begin{aligned} \Sigma^2_{M,\infty}(E) \; &\sim \; \frac{c_\Gamma}{2\pi^2}\, Ei\left(\log\left(\frac{4\pi}{c_\Gamma} p\bar{d}(E)\right)\right) + C(l_1) \\ &\sim \; \frac{2}{\pi} \frac{p\bar{d}(E)}{\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E))} \, , \qquad E \to \infty \, . \end{aligned} \tag{4.50}$$

The energy dependence of $\Delta_\infty(E)$ derived from the model thus is

$$\Delta_{M,\infty}(E) \sim \frac{2\bar{d}}{\pi} \frac{\sqrt{E}}{\log E} \, , \qquad E \to \infty \, . \tag{4.51}$$

This result should be compared with the rigorous lower bound $N_{fl}(E(p)) = S(p) = \Omega_\pm(\frac{\sqrt{p}}{\log p})$ [51] for arithmetic groups. Via (2.24) the latter yields

$$\Delta_\infty(E) = \Omega\left(\frac{\sqrt{E}}{(\log E)^2}\right) \, . \tag{4.52}$$

This being a lower bound is well in accordance with the result (4.51) obtained from the model. Given the latter describes the actual saturation value of $\Sigma^2(L;E)$ correctly in the semiclassical limit, this means that the lower bound (4.52) is off the true magnitude only by a factor of $\log E$. The upper bound $S(p) = O(\frac{p}{\log p})$, yielding $\Delta_\infty(E) = O(\frac{E}{(\log E)^2})$, is therefore much less sharp than (4.52).

In order to test the model, the r.h.s. of the first line of (4.50) has been evaluated numerically and multiplied by $\frac{1}{2}$ for the odd symmetry class of Artin's billiard. Fig.6 presents the result as the full curve. Steil has in [91] approximated $\Delta_\infty(E)$ by $\Delta_3(L;E)$ at $L = 37.5$. Since the rigidity is monotonically increasing as a function of L, these values yield lower bounds on $\Delta_\infty(E)$ and are marked as dots in fig.6. One notes from fig.6 that the model appears to reproduce the

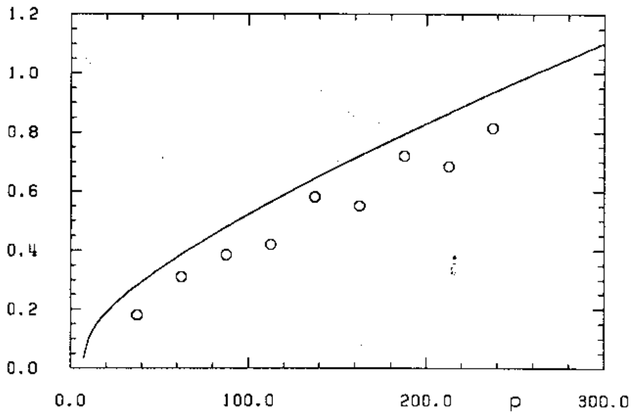

Figure 6: The saturation value $\frac{1}{2}\Sigma^2_{M,\infty}(p)$ of the model number variance for the odd symmetry class of Artin's billiard is shown as the full curve. The dots mark lower bounds on $\frac{1}{2}\Sigma^2_{\infty}(p)$ obtained in [91] from the eigenvalues.

functional form of $\Delta_\infty(E)$ correctly, but that the constant $C(l_1) = 0.075$ seems to come out a little bit to large. In view of the fact, however, that the dots only indicate lower bounds, the model for $\Delta_\infty(E)$ might describe the true saturation values better than it is observed from fig.6.

Finally the above observation should be compared to very recent rigorous results of Sarnak and Luo [80]. They define the number variance as

$$\Sigma^2_{SL}(L;x) := \frac{1}{x}\int_x^{2x} dx' \; [N(x'+L) - N(x') - L]^2 \qquad (4.53)$$

for $x \to \infty$, and obtain the following

THEOREM: In the range $\frac{\sqrt{x}}{\log x} \ll L \ll L_{max}$, $L_{max} \propto \sqrt{x}$, the estimates

$$\begin{aligned} \Sigma^2_{SL}(L;x) &= \Omega\left(\frac{L}{\log L}\right) \\ \Sigma^2_{SL}(L;x) &= \Omega\left(\frac{\sqrt{x}}{\log x}\right) \end{aligned} \qquad (4.54)$$

hold for all arithmetic Fuchsian groups.

The L–range these bounds refer to is the upper part of Berry's universal regime $1 \ll L \ll L_{max}$. The first line bounding the L–dependence is, if taken as sharp, in accordance with the

numerical observations as well as with our model in that in any case $\Sigma^2(L;E) \leq L$ on the interval alluded to above. A direct comparison of the model (4.46) and the first line of (4.54) is not possible because (4.46) is only valid for $L \to 0$. Since $\Sigma^2_M(L;E)$ is certainly positive, an approximation using the r.h.s. of (4.46) up to $O(L^2)$ is restricted to

$$L \leq \frac{\pi p \bar{d}(E)}{\log(\frac{4\pi}{c_\Gamma} p\bar{d}(E)) - 1} \; . \qquad (4.55)$$

Thus there is no overlap with the range of validity of the Theorem. Those two results should rather be viewed as being complementary. The second line of the Theorem estimating the E–dependence at finite L is in accordance with the above findings about $\Sigma^2_{M,\infty}(E)$ and, of course, with the rigorous estimate (4.52), although a direct comparison is also, strictly speaking, prohibited by the restriction $L \ll L_{max}$ in the Theorem.

Summarizing the results on the number variance of arithmetical systems obtained in this section one notices that $\Sigma^2(L;E)$ is for small values of L reasonably well approximated by a Poissonian behaviour. The L–range on which this agreement takes place grows in the semiclassical limit. For larger values of L $\Sigma^2(L;E)$ deviates slowly from being linear as first described by the model (4.46) and then by the Theorem of Sarnak and Luo. At some $L_{max} \propto \sqrt{E}$ the number variance saturates and oscillates non-universally beyond that value. The oscillations and their mean value are determined by the short closed geodesics on the surface $\Gamma\backslash\mathcal{H}$. The number of distinct lengths contributing, however, is $\hat{N}_p(l_{max}) \sim \frac{8\pi}{c_\Gamma^2} p\bar{d}(E)$ and therefore tends to infinity for $E \to \infty$. The saturation value $< N_{fl}(E)^2 >= \Delta_\infty(E) = \frac{1}{2}\Sigma^2_\infty(E) \sim \frac{2\bar{d}}{\pi}\frac{\sqrt{E}}{\log E}$, $E \to \infty$, grows slightly less (by a factor of $\frac{1}{\log E}$) than the corresponding value for classically integrable systems. It is, however, certainly well beyond the one for generic classically chaotic systems, $\Delta_\infty(E) \gg \frac{1}{2\pi^2}\log E$.

Finally a remark on the class of systems that are being described by the findings of the present chapter will be added. The reason why the model works is provided by the exponentially increasing multiplicities of lengths of closed geodesics for arithmetic Fuchsian groups, since the exponential behaviour (4.32) of the form factor is caused by the compensation of the exponential damping present in the amplitude factor $A_{n,k}$ (4.26) through $g_p(l_n)$. In order this to work it has been assumed that all $g_p(l_n)$ closed geodesics of the same length l_n shared alike factors $\chi(\gamma)$. Otherwise the sum over closed geodesics could not have been rewritten as a sum over distinct primitive lengths, see (4.26). Assuming the simplest case of $\mathbb{Z}_2$–valued representations $\chi : \Gamma \to \{\pm 1\}$, one can group the geodesics in classes of alike signs and define $g_p(l_n) = g_n^+ + g_n^-$, where $g_n^\pm$ denotes the number of geodesics of length l_n with $\chi(\gamma) = \pm 1$, respectively. The diagonal term (4.28) then reads

$$K_D(\tau;E) = \frac{1}{(4\pi p\bar{d}(E))^2}\sum_{\{l_n\}}\sum_{k=1}^{\infty} l_n^2\, e^{-kl_n}\,[g_n^+ + (-1)^k g_n^-]^2\,\delta(\tau - \frac{kl_n}{4\pi p\bar{d}(E)})\,[1 + O(e^{-kl_n})] \; . \quad (4.56)$$

The leading ($k = 1$)–contribution thus contains the difference $[g_n^+ - g_n^-]^2$ of the multiplicities referring to $\chi(\gamma) = +1$ and $\chi(\gamma) = -1$. Only if this difference grows like $|g_n^+ - g_n^-| \sim const.\ \frac{e^{l_n/2}}{l_n}$, $l_n \to \infty$, the model of sections 4.3 and 4.4 is applicable. Once g_n^+ and g_n^- are of the same order of magnitude the leading contribution comes from $k = 2$. Since then $[g_n^+ + (-1)^2 g_n^-]^2 = g_p(l_n)^2 \sim c_\Gamma^2 \frac{e^{l_n}}{l_n^2}$, this term is of a similar form as the analogous one for $k = 1$ in the non-arithmetic case. Thus the statistical properties are expected to be generic, i.e. the level spacings should be close to the GOE behaviour and the medium- and long-range correlations should be described by Berry's theory.

As an example for a non-trivial representation χ take an arithmetic Fuchsian group Γ_1 leading to a symmetric surface $\Gamma_1\backslash\mathcal{H}$. Then Γ_1 is a normal subgroup of index N in another, also arithmetic, group Γ_2. According to (3.20) thus $< g_p^{(2)}(l) >\sim \frac{1}{N} < g_p^{(1)}(l) >$, $l \to \infty$. The symmetry group Γ_2/Γ_1 is represented via $\chi : \Gamma_2 \to \mathrm{End}\,(V_\chi)$ with $\ker\chi \supseteq \Gamma_1$. Since therefore $\chi(\gamma) = +1$ for $\gamma \in \Gamma_1$, one concludes that $g_n^+ \geq g_p^{(1)}(l_n)$. Again only $\mathbb{Z}_2$–valued symmetry classes shall be considered for simplicity. Then $\chi(\gamma) = -1$ is only possible for $\gamma \in \Gamma_2$, hence $< g_n^- > \leq < g_p^{(2)}(l_n) >\sim \frac{1}{N} < g_p^{(1)}(l_n) >$, yielding $< g_n^+ > - < g_n^- > \geq (1 - \frac{1}{N}) < g_p^{(1)}(l_n) >\sim const.\ \frac{e^{l_n/2}}{l_n}$ for $l_n \to \infty$. Therefore the model is still applicable to this case.

The more general situation of arbitrary representations of a symmetry group can be treated analogously. In view of the possible cancellations of multiplicities in the $(k = 1)$–term of (4.56) the case of a $\mathbb{Z}_2$–valued representation, however, is the worst possible. Hence the Laplacian on a symmetric arithmetic surface is always expected to share statistical properties as discussed in the preceding sections. Artin's billiard may serve as an example that has already been studied above. Although the symmetry on the modular surface is orientation reversing and thus Γ_2 is not a subgroup of $SL(2, \mathbb{R})$ but rather of the full group of isometries of $\mathcal{H}$, the above reasoning extends also to Artin's billiard, since inverse hyperbolic orbits occurring due to orientation reversing symmetries can as well be dealt with in the present framework, see [99, 100, 68].

There may of course exist more general representations of arithmetic Fuchsian groups Γ than the ones being derived from symmetries. The latter excel be their triviality on a subgroup of finite index, which leads to the observation just made that a large enough fraction of closed geodesics is equipped with positive χ's. Examples for the former may be provided by the presence of Aharonov-Bohm type magnetic fluxes on arithmetic surfaces $\Gamma\backslash\mathcal{H}$. Depending on the strength of such a flux the spectral properties of the respective Laplacian are expected to deviate from the findings of the present chapter. Once the phases $\chi(\gamma)$ that arise when a wave function $\psi(z)$ is carried along the geodesics related to the $\gamma \in \Gamma$ and enclosing the Aharonov-Bohm flux line "mix" sufficiently among those geodesics being degenerate in length, one could even retain generic spectral statistics like those for non-arithmetic groups.

4.5 Convergence Properties of the Selberg Zeta Function

The final topic of the present chapter now again deals with general Fuchsian groups of the first kind. However, it is observed in the course of the following discussion that arithmetic groups play a special role. As in section 3.6 methods of inverse quantum chaology are applied and once again it turns out that the strong fluctuations present for arithmetic quantum energy spectra violate the prerequisits to apply the formalism developed below to the arithmetic case. Referring to the heuristic reasoning presented first in [6] one can, however, understand the reason for the obstruction occurring for arithmetic groups.

The item of this section lies at the foundation of one of the major objectives of quantum chaology, namely the derivation of certain quantization rules that allow to determine the quantum energies of a classically chaotic system in a semiclassical approximation. Recently such quantization rules involving dynamical zeta functions have been introduced and successfully applied to a variety of different chaotic systems, see e.g. [89, 68, 95, 26, 63, 23, 13].

All these methods make strong use of the fact that the semiclassical quantum energies are directly related to the zeros of the dynamical zeta function on the critical line. The problem one immediately faces when trying to compute the non-trivial zeros explicitly is that the Euler product (2.9) defining the dynamical zeta function in general does not converge on the critical line. This phenomenon is also referred to as the existence of an *entropy barrier*, since it is

the topological entropy τ that determines the half-plane of convergence $Re\, s > \tau$ for the Euler product. Concerning geodesic flows on hyperbolic surfaces and the Selberg zeta function it was mentioned earlier that the topological entropy universally is $\tau = 1$, and that the critical line is located at $Re\, s = \frac{1}{2}$. The entropy barrier to be overcome hence has a width of $\frac{1}{2}$.

In order to find a quantization rule one therefore has to develop a consistent procedure to calculate the non-trivial zeros of the zeta function other than searching for the zeros of the Euler product. It was McKean [69] who apparently first noticed the possibility to rewrite the Euler product (3.72) of the Selberg zeta function as a Dirichlet series, but made no use of this. Berry and Keating [22] then were the first to introduce this Dirichlet series in order to obtain a quantization rule from it, leading to their ***Riemann-Siegel look alike formula.*** Sieber and Steiner [89], Matthies and Steiner [68] and Aurich and Steiner [13] then investigated the convergence properties of the Dirichlet series in several examples and used it to calculate non-trivial zeros. After that in [6] a statistical model was developed that predicts the domain of conditional convergence for the Dirichlet series of the Ruelle-type zeta function (3.74), which as well can be used to set up a quantization rule. This model can also be extended to the Selberg zeta function, see e.g. [5].

The above remarks stress the importance of investigating the convergence properties of the Dirichlet series representing the Selberg zeta function. Very much alike the discussion in section 3.6 methods from analytic number theory will be used below in conjunction with inverse quantum chaology to derive statements about the abscissa of convergence for the Dirichlet series. Again, as in section 3.6, for simplicity only cocompact Fuchsian groups Γ will be considered. But as also has been stated earlier, the results do not depend on this restriction, since it is only the contributions of the hyperbolic $\gamma \in \Gamma$ that are relevant.

Before entering the detailed studies just announced the model introduced in [6] will be reviewed. In order to convert the Euler product (3.72) into a Dirichlet series one has to transform the product over $n \in \mathbb{N}_0$ with the help of Euler's identity

$$\prod_{n=0}^{\infty}(1 - y\,x^n) = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m\, y^m\, x^{\frac{m}{2}(m-1)}}{\prod_{k=1}^{m}(1 - x^k)} \tag{4.57}$$

first; $|x| < 1$ and $y \in \mathbb{C}$ suffice for the product and the sum to converge. One then obtains that for $Re\, s > 1$

$$Z(s) = \sum_{\rho} A_\rho\, e^{-sL_\rho}\ , \tag{4.58}$$

where the sum extends over all pseudo-orbits ρ, see (3.77), and the L_ρ's denote the pseudo-lengths (3.78). The coefficients A_ρ of the generalized Dirichlet series (4.58) are given by

$$A_\rho = \prod_{i=1}^{r} a_i\ , \qquad a_i := \frac{(-1)^{m_i}\, e^{-\frac{m_i}{2}(m_i-1)l_{\gamma_i}}}{\prod_{k=1}^{m_i}(1 - e^{-kl_{\gamma_i}})}\ , \tag{4.59}$$

for $\rho = \{\gamma_1^{m_1}\}_p \oplus \ldots \oplus \{\gamma_r^{m_r}\}_p$. Recall that a pseudo-orbit ρ is a formal combination of finitely many (not necessarily primitive) closed geodesics, the integers $m_1, \ldots, m_r$ denoting the numbers of respective repetitions of the primitive closed geodesics corresponding to $\{\gamma_1\}_p, \ldots, \{\gamma_r\}_p$. The asymptotic relation $\mathcal{N}(l) \sim \mathcal{N}_p(l)$, $l \to \infty$, which is an analogue of (3.9), expresses the fact that the number of all closed geodesics of lengths not exceeding l is in the asymptotic regime $l \to \infty$ already given by the respective number of primitive geodesics. In the limit $L \to \infty$ therefore the number $\mathcal{N}^{(P)}(L)$ of pseudo-orbits with $L_\rho \leq L$ is dominated by those ρ that are completely composed of primitive closed geodesics traversed only once, i.e. $m_1 = \ldots = m_r = 1$. Their respective coefficients A_ρ look like

$$A_\rho = \prod_{i=1}^{r} \frac{(-1)}{1-e^{-l_{\gamma_i}}} \sim (-1)^{|\rho|} , \tag{4.60}$$

where $|\rho| := r$ denotes the number of primitive geodesics the pseudo-orbit ρ consists of.

Since already the lengths of primitive closed geodesics are at least twice degenerate because of the time-reversal symmetry, the pseudo-lengths L_n will in general also possess multiplicities exceeding one. Assuming that these multiplicities $g^{(P)}(L_n)$ are exclusively caused by degenerate primitive lengths and not by different combinations of primitive lengths yielding the same pseudo-length, degenerate ρ's have identical coefficients A_ρ. Thus

$$Z(s) = \sum_{\{L_n\}} g^{(P)}(L_n)\, A_n\, e^{-sL_n} \tag{4.61}$$

for $Re\, s > 1$. Numerical calculations now hint at an exponential increase of the multiplicities according to [6, 5]

$$< g^{(P)}(L) > \sim d\, e^{\alpha L} , \qquad L \to \infty ; \tag{4.62}$$

d and α are constants to be determined numerically. Thus

$$Z(s) = \sum_{\{L_n\}} d\, A_n \frac{g^{(P)}(L_n)}{d e^{\alpha L_n}}\, e^{-(s-\alpha)L_n} , \tag{4.63}$$

which can be considered as a generalized Dirichlet series in the variable $s-\alpha$. The theory of Dirichlet series [48] as e.g. also briefly reviewed in appendix A now allows to determine the convergence properties of the Dirichlet series for $Z(s)$ from (4.63) once the pseudo-orbits have been arranged in ascending order of their respective pseudo-lengths, $0 = L_0 < L_1 < L_2 < \ldots$. According to this theory (4.58) converges for $Re\, s > \sigma_c$ and diverges for $Re\, s < \sigma_c$; furthermore it converges absolutely for $Re\, s > \sigma_a$, $\sigma_a \geq \sigma_c$. The abscissa of convergence σ_c and of absolute convergence σ_a are determined by the formulae (A.16). The asymptotic growth (3.89) of the number $\mathcal{N}^{(P)}(L)$ of pseudo-orbits with pseudo-lengths not exceeding L then fixes the abscissa of absolute convergence to be $\sigma_a = 1$. In [6] a statistical model was established that yielded $\sigma_c = \frac{1+\alpha}{2}$. The model assumes that after applying the approximation (4.60) the coefficients A_n in (4.63) represent random signs. This conjecture is based on the observation that by (4.60) $A_\rho \sim \pm 1$, the sign depending on whether the number of primitive closed geodesics comprising the pseudo-orbit ρ is even or odd, respectively. Arranging the pseudo-lengths in ascending order and taking into account a supposed irregularity in the distribution of primitive lengths should make the numbers $|\rho_n|$ and $|\rho_{n+1}|$ modulo two independent of one another.

Taking this randomness hypothesis for granted one can obtain the result for σ_c also slightly differently, although in the same spirit as in [6]. For it is known [61] that a series

$$\sum_{k=1}^{\infty} (-1)^{\omega_k} c_k \tag{4.64}$$

of positive coefficients c_k and random signs $(-1)^{\omega_k}$ either converges or diverges, depending on whether the series $\sum_{k=1}^{\infty} c_k^2$ converges or diverges, respectively. Recalling (3.89) and (4.62), one observes that the number of distinct pseudo-lengths up to a value of L grows asymptotically proportional to $e^{(1-\alpha)L}$. Thus the criterion for convergence stated after (4.64) requires $Re\, s >$

$\frac{1+\alpha}{2}$ for the Dirichlet series (4.63) to converge, reproducing hence the outcome of the model in [6]. In conclusion, one learns from this model that the Dirichlet series representing the Selberg zeta function is not expected to converge on the critical line $Re\, s = \frac{1}{2}$. The distance of σ_c to the critical line is determined by the growth of the multiplicities of pseudo-lengths.

For arithmetic systems α is expected to be large since already the multiplicities of primitive lengths grow exponentially. Indeed, by numerical calculations of pseudo-length spectra up to some cut-off value L_{max} it was observed in [6] that $\alpha = 0.4658$ for the regular octagon group, and $\alpha = 0.279$ for Artin's billiard. However, non-arithmetic and completely desymmetrized systems should possess a considerably smaller value for α. In [5] an example for such a system was studied numerically and $\alpha = 0.0572$ was found. It even may be that in these cases the multiplicities will not really show an exponential behaviour, but will rather follow a power law, leading to an effective vanishing of α. Then the Dirichlet series would converge for $Re\, s > \frac{1}{2}$ and diverge for $Re\, s < \frac{1}{2}$, but it would not be known whether it converged on the critical line. In any case it would be possible to evaluate $Z(s)$ close to the critical line and to obtain the non-trivial zeros as minima. One could even try to go onto the critical line and hope that a divergence would not show up when using the finitely many available pseudo-lengths in (4.58). An example of such a procedure is presented in [5].

A remark on the (arithmetic) case of Artin's billiard seems to be in place now. It is expected that σ_c will in this case be well above $\frac{1}{2}$, keeping in mind the rather large value of $\alpha = 0.279$. However, a numerical evaluation of the formula (A.16) for σ_c yields a value below $\frac{1}{2}$ [68, 6], at least in the finite range of available pseudo-lengths. An explanation for this observation is that the randomness hypothesis the statistical model is built upon is apparently violated [6] in the computed range of the pseudo-length spectrum. Thus Artin's billiard can apparently not be understood by the above considerations. The question whether or not this phenomenon pertains to higher values of L remains open.

In the following the question for the location of the abscissa of convergence σ_c of the Dirichlet series (4.58) will be approached from a different side, employing methods from analytic number theory and inverse quantum chaology. As guiding references [97, 51] may be consulted. The idea to be pursued below is similar to the one how to obtain the PNT from the analytic properties of $\frac{\zeta'(s)}{\zeta(s)}$ (see (A.12)), or how to obtain the PGT from $\frac{R'(s)}{R(s)}$ (see (3.84)). To this end define the function

$$\psi_Z(L) := \sum_{\rho, L_\rho \leq L} A_\rho , \tag{4.65}$$

where the above notation should indicate that the pseudo-lengths have to be counted with their respective multiplicities. The abscissa of convergence is then according to (A.16) given by

$$\sigma_c = \limsup_{L_\rho \to \infty} \frac{1}{L_\rho} \log |\psi_Z(L_\rho)| . \tag{4.66}$$

If it were possible to derive an O–estimate for $\psi_Z(L)$, this would yield an upper bound for σ_c, i.e. $\psi_Z(L) = O(L^a e^{bL})$, $a \in \mathbb{R}$, $b > 0$, results in the bound $\sigma_c \leq b$. Accordingly, an Ω–result for $\psi_Z(L)$ would give a lower bound on σ_c. Thus it will be attempted to estimate $\psi_Z(L)$ for $L \to \infty$. The principle tools to be employed have already been used in section 3.6 and in appendix A.

Using the Dirichlet series (4.58) and the integral (A.11) one easily obtains

$$\psi_Z(L) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{ds}{s}\, e^{sL}\, Z(s) , \qquad c > 1, \tag{4.67}$$

in analogy to (3.84) and (A.12). However, it does not proof particularly useful to extend the contour of integration in (4.67) from $c-i\infty$ to $c+i\infty$. In the following the integral will therefore be restricted to the finite interval $[c-iT, c+iT]$, $T>0$, and the remainder that has been left out in comparison to (4.67) will be estimated. This is achieved by the following

Lemma: With the notations introduced above, and $c>1$, $L>0$, $L \neq L_\rho$,

$$\psi_Z(L) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{ds}{s}\, e^{sL}\, Z(s) + O\left(\frac{e^{cL}}{(c-1)T}\right) + O\left(\frac{L}{T}e^{L}\right) . \tag{4.68}$$

Proof: The proof is a standard calculation in analytic number theory, see e.g. [97], pp.60. Due to the importance of the result for the further considerations the main ideas shall, however, be reproduced here.

Inserting the Dirichlet series (4.58) into the integral on the r.h.s. of (4.68) one is left with an integrand of $\frac{1}{s}e^{s(L-L_\rho)}$. Depending on the sign of $L-L_\rho$ one has to choose different contours to render the following integrals finite:

1. $L > L_\rho$:

$$\frac{1}{2\pi i}\left\{\int_{-\infty-iT}^{c-iT} + \int_{c-iT}^{c+iT} + \int_{c+iT}^{-\infty+iT}\right\}\frac{ds}{s}\, e^{s(L-L_\rho)} = 1 , \tag{4.69}$$

because the pole of the integrand at $s=0$ is enclosed by the contour.

2. $L < L_\rho$:

$$\frac{1}{2\pi i}\left\{\int_{\infty-iT}^{c-iT} + \int_{c-iT}^{c+iT} + \int_{c+iT}^{\infty+iT}\right\}\frac{ds}{s}\, e^{s(L-L_\rho)} = 0 . \tag{4.70}$$

A typical integral to be estimated now can be treated by an integration by parts ($L > L_\rho$),

$$\begin{aligned}\int_{-\infty-iT}^{c-iT}\frac{ds}{s}\, e^{s(L-L_\rho)} &= \left.\frac{e^{s(L-L_\rho)}}{s(L-L_\rho)}\right|_{-\infty-iT}^{c-iT} + \frac{1}{L-L_\rho}\int_{-\infty-iT}^{c-iT}\frac{ds}{s^2}\, e^{s(L-L_\rho)} \\ &= \frac{e^{(c-iT)(L-L_\rho)}}{(c-iT)(L-L_\rho)} + \frac{e^{-iT(L-L_\rho)}}{L-L_\rho}\int_{-\infty}^{c} d\sigma\, \frac{e^{\sigma(L-L_\rho)}}{(\sigma-iT)^2} .\end{aligned} \tag{4.71}$$

To obtain an upper bound for the absolute value of the above expression one extracts the maximal value of the exponential under the integral on the r.h.s. and obtains

$$\begin{aligned}\left|\int_{-\infty-iT}^{c-iT}\frac{ds}{s}\, e^{s(L-L_\rho)}\right| &\leq \frac{e^{c(L-L_\rho)}}{|c-iT|(L-L_\rho)} + \frac{e^{c(L-L_\rho)}}{L-L_\rho}\int_{-\infty}^{+\infty} d\sigma\, \frac{1}{\sigma^2+T^2} \\ &= O\left(\frac{e^{c(L-L_\rho)}}{T(L-L_\rho)}\right) .\end{aligned} \tag{4.72}$$

The remaining three integrals from (4.69) and (4.70) are of the same type and obey the same bounds. Thus

1. $L > L_\rho$:

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{ds}{s}\, e^{s(L-L_\rho)} = 1 + O\left(\frac{e^{c(L-L_\rho)}}{T(L-L_\rho)}\right) , \tag{4.73}$$

2. $L < L_\rho$:

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{ds}{s}\, e^{s(L-L_\rho)} = O\left(\frac{e^{c(L-L_\rho)}}{T(L-L_\rho)}\right) . \tag{4.74}$$

Using these bounds one finds with the help of (4.65) that

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{ds}{s}\, e^{sL}\, Z(s) = \psi_Z(L) + O\left(\frac{e^{cL}}{T}\sum_\rho |A_\rho|\frac{e^{-cL_\rho}}{|L-L_\rho|}\right) . \tag{4.75}$$

One is therefore left with the task of bounding the sum on the r.h.s. of (4.75). Somewhat tedious but straightforward calculations that may e.g. be found in [97], pp.60, and will not be reproduced here yield

$$\sum_\rho |A_\rho|\frac{e^{-cL_\rho}}{|L-L_\rho|} = O\left(\frac{1}{c-1}\right) + O(L\, e^{(1-c)L}) , \tag{4.76}$$

finally proving the lemma.

The next task in order to derive a bound for $\psi_Z(L)$ is to estimate the integral that is left on the r.h.s. of (4.68). Because of the factor of e^{sL} under the integral it proofs useful to move the contour from $Re\, s = c > 1$ to the left in the complex s-plane as far as possible. It turns out that this can be achieved up to directly before the critical line. What is missing yet is an estimate of $Z(s)$ on the contour. It is at this point where an inverse quantum chaology argument enters the game. Namely, Hejhal can prove an estimate for $Z(s)$ in the half-plane $Re\, s > \frac{1}{2}$ depending on an upper bound for $S(p) = N_{fl}(E(p)) = \frac{1}{\pi}\arg Z(\frac{1}{2}+ip)$. Define

$$\Delta(p) := p^\mu(\log p)^\nu(\log\log p)^\lambda , \tag{4.77}$$

where the exponents μ, ν, λ are chosen such that $\Delta(p)$ tends to infinity for $p \to \infty$. Using the notation $s = \sigma + ip$, $\sigma \in I\!R$, $p > 0$, *Theorem 10.10* in [51] then states that $|S(p)| = O(\Delta(p))$ implies

$$\log Z(s) = O\left(\Delta(p)^{2\max[0,1-\sigma]}\log\Delta(p)\right) , \tag{4.78}$$

for $\sigma = Re\, s \geq \frac{1}{2} + \frac{1}{\log\Delta(p)}$ and p large enough, $p \geq p_0(\Delta)$. It will henceforth be assumed that $p_0(\Delta)$ is chosen that large that $\Delta(p)$ is monotonically increasing for $p \geq p_0(\Delta)$. From now on σ shall be restricted to the domain $\frac{1}{2}+\frac{1}{\log\Delta(p)} < \frac{1}{2}+\frac{1}{\log\log\Delta(p)} \leq \sigma \leq 1$, on which $2\max[0,1-\sigma] = 1 - \frac{2}{\log\log\Delta(p)}$. Thus $\log Z(s) = O(\Delta(p)\, e^{-\frac{2\log\Delta(p)}{\log\log\Delta(p)}+\log\log\Delta(p)}) = O(\Delta(p)\, e^{-\frac{\log\Delta(p)}{\log\log\Delta(p)}})$. Since therefore $|\log Z(s)| \leq \eta\, \Delta(p)^{1-\frac{1}{\log\log\Delta(p)}}$ for some constant $\eta > 0$ and $p \geq p_0(\Delta)$, which again must be chosen large enough, one obtains that $|\log Z(s)| \leq \varepsilon\Delta(p)$ for all $\varepsilon > 0$ and p large enough. Thus in the domain alluded to above

$$|Z(s)| \leq e^{\varepsilon\Delta(p)} , \qquad \forall \varepsilon > 0 . \tag{4.79}$$

The integration contour in (4.68) can by Cauchy's theorem now be moved to $Re\, s = \frac{1}{2}+\delta$, $\delta := \frac{1}{\log\log\Delta(T)}$, without loosing control on the magnitude of $Z(s)$,

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{ds}{s}\, e^{sL}\, Z(s) = \frac{1}{2\pi i}\left\{\int_{c-iT}^{\frac{1}{2}+\delta-iT} + \int_{\frac{1}{2}+\delta-iT}^{\frac{1}{2}+\delta+iT} + \int_{\frac{1}{2}+\delta+iT}^{c+iT}\right\}\frac{ds}{s}\, e^{sL}\, Z(s) . \tag{4.80}$$

The first and the third integral on the r.h.s. behave alike,

$$\left|\int_{c\pm iT}^{\frac{1}{2}+\delta\pm iT}\frac{ds}{s}\,e^{sL}\,Z(s)\right| \le \int_{\frac{1}{2}+\delta}^{c} d\sigma\,\frac{e^{\sigma L}}{|\sigma\pm iT|}|Z(\sigma\pm iT)| = O\left(\frac{1}{T}\,e^{\epsilon\Delta(T)}\,e^{cL}\right). \tag{4.81}$$

The second integral, however, can be bounded according to

$$\begin{aligned}\left|\int_{\frac{1}{2}+\delta-iT}^{\frac{1}{2}+\delta+iT}\frac{ds}{s}\,e^{sL}\,Z(s)\right| &\le e^{(\frac{1}{2}+\delta)L}\int_{-T}^{+T}dt\,\frac{|Z(\frac{1}{2}+\delta+it)|}{|\frac{1}{2}+\delta+it|}\\ &= O\left(e^{(\frac{1}{2}+\delta)L}\,e^{\epsilon\Delta(T)}\int_0^T dt\,\frac{1}{1+t}\right)\\ &= O\left(e^{(\frac{1}{2}+\delta)L}\,e^{\epsilon\Delta(T)}\log T\right).\end{aligned} \tag{4.82}$$

Combining the estimates (4.81) and (4.82) with the lemma one obtains the following

PROPOSITION: The function $\psi_Z(L)$ can be estimated for $L\to\infty$ using the notations introduced above as

$$\psi_Z(L) = O\left(\frac{1}{T}e^{\epsilon\Delta(T)}e^{cL}\right) + O\left(e^{(\frac{1}{2}+\delta)L}e^{\epsilon\Delta(T)}\log T\right) + O\left(\frac{1}{T}e^{cL}\right) + O\left(\frac{L}{T}e^{L}\right). \tag{4.83}$$

Recall that one seeks for an upper bound of the type $\psi_Z(L) = O(L^a e^{bL})$, $1 > b > 0$, in order to bound the abscissa of convergence by $\sigma_c \le b < 1$. The third and the fourth term on the r.h.s. of (4.83) therefore require to take $T = e^{dL}$ for some appropriate $d > c-1 > 0$. Then, however, $\Delta(T) = e^{\mu dL}(dL)^{\nu}(\log dL)^{\lambda}$, and the first two terms prohibit to obtain the desired form of the estimate unless $\mu = 0$. Once, however, $\Delta(p) = (\log p)^{\nu}(\log\log p)^{\lambda}$ for $\nu < 1$ or $\nu = 1$, $\lambda \le 0$, and thus $|S(p)| = O(\log p)$, one observes with the choice $d = c > 1$

$$\psi_Z(L) = O(e^{\epsilon cL}) + O(L\,e^{(\frac{1}{2}+\delta+\epsilon c)L}) + O(L\,e^{-(c-1)L}) \tag{4.84}$$

for $L\to\infty$ and for all $\varepsilon > 0$. Since $\delta = \frac{1}{\log\log\Delta(T)}$ vanishes for $L\to\infty$ and ε can be made as small as required, one draws from (4.66) determining the abscissa of convergence the bound

$$\sigma_c \le \frac{1}{2} + \varepsilon' \qquad \text{for all } \varepsilon' > 0\,. \tag{4.85}$$

Therefore, the Dirichlet series (4.58) for the Selberg zeta function converges (conditionally) for all s with $Re\,s > \frac{1}{2}$, since one can then always choose ε' as small as desired.

This being a conditional result, the question for the validity of the input $|S(p)| = |N_{fl}(E(p))| = O(\log p)$ immediately arises. For arithmetic groups the lower bound already employed in sections 3.6 and 4.2, $S(p) = \Omega_{\pm}(\frac{\sqrt{p}}{\log p})$, forces to chose $\mu \ge \frac{1}{2}$ in (4.77), therefore ruling out an application of (4.83) to obtain an upper bound for σ_c. This negative observation comes in accordance with the result obtained from the statistical model, $\sigma_c = \frac{1+\alpha}{2} > \frac{1}{2}$, where α, describing the growth of the multiplicities of pseudo-lengths, is rather large for arithmetic groups, and thus σ_c violates the lower bound (4.85). The lower bound for general (cocompact) Fuchsian groups, $S(p) = \Omega_{\pm}(\sqrt{\frac{\log p}{\log\log p}})$, still allows for expecting $\mu = 0$ and $\nu \le 1$. Now suppose

that $S(p) = O((\log p)^{\nu})$, i.e. $N_{fl}(E) = O((\log E)^{\nu})$. In order to obtain the saturation value $\Delta_{\infty}(E)$ of the spectral rigidity in the semiclassical limit (see (2.24)) one has to evaluate

$$\frac{\overline{d}}{2L}\int_{E-\frac{L}{\overline{d}}}^{E+\frac{L}{\overline{d}}} dE'\,[N_{fl}(E')]^2 = O((\log E)^{2\nu}) + O\left((\log E)^{2\nu-1}\frac{L^2}{E^2}\right) \tag{4.86}$$

in the limit $L\to\infty$ and $E\to\infty$. The interval of length $2\frac{L}{\overline{d}}$ to be integrated over has to be kept small compared to E. Choosing $L = E^a$, $0 < a < 1$, then yields the semiclassical asymptotics

$$\Delta_{\infty}(E) \sim O((\log E)^{2\nu})\,. \tag{4.87}$$

Berry's semiclassical theory for the spectral rigidity yielding for generic classically chaotic systems $\Delta_{\infty}(E) \sim \frac{1}{2\pi^2}\log E$ now implies $\nu = \frac{1}{2} < 1$. Once one believes in the applicability of this heuristic theory to non-arithmetic Fuchsian groups one has to draw the conclusion that the Dirichlet series (4.58) for the respective Selberg zeta functions converge conditionally for $Re\,s > \frac{1}{2}$. The above reasoning can be supported by the numerical evaluation of the formula (A.16) in an example of a non-arithmetic group in [5], which yielded a result in accordance with the bound (4.85). From the part of the pseudo-length spectrum calculated in [5] one can, however, not draw a clear-cut conclusion on the precise value of σ_c because the plotted curve still oscillates rather strongly.

A further confirmation of the above results may be provided by considering the Riemann zeta function $\zeta(s)$. Assuming the Riemann hypothesis its non-trivial zeros are given by $s_n = \frac{1}{2}\pm i\gamma_n$, $\gamma_n \ge 0$. Supposing that either γ_n or γ_n^2 correspond to quantum energies of a yet unknown physical system, the spectral statistics show a behaviour as if the classical limit of this system were chaotic without time-reversal invariance, i.e. the level spacings can be well described by the GUE random matrix ensemble, and the spectral rigidity and the number-variance saturate for $L\to\infty$, see e.g. [74, 21]. The function $S(p) = \frac{1}{\pi}\arg\zeta(\frac{1}{2}+ip)$ is known to obey the same lower bound $S(p) = \Omega_{\pm}(\sqrt{\frac{\log p}{\log\log p}})$ [72] as the analogous quantity for a generic Selberg zeta function. Its upper bound, however, is given by $S(p) = O(\frac{\log p}{\log\log p})$, see e.g. [50]. Berry's theory of the rigidity stating that $\Delta_{\infty}(E) \sim \frac{1}{4\pi^2}\log E$ would predict that $|S(p)|$ is asymptotically given by $\sqrt{\log p}$, possibly times some power of $\log\log p$, therefore clearly being within the rigorous bounds. This is also in accordance with the belief that the Ω–estimate is "likely to be best possible" [72] and possibly sharp. The upper bound for $S(p)$ leads to an analogue of (4.78), namely $\log\zeta(s) = O((\log p)^{2\max[0,1-\sigma]}\log\log p)$ for $\sigma = Re\,s \ge \frac{1}{2} + \frac{1}{\log\log p}$. The estimate corresponding to (4.79), $\zeta(s) = O(p^{\varepsilon})$, $\forall\varepsilon > 0$, is then equivalent to the Lindelöf hypothesis, see [97] for details.

In conclusion it appears that as long as the geodesic flow on a hyperbolic surface $\Gamma\backslash\mathcal{H}$ is generic in the sense that the spectral statistics can be described by RMT on small scales and follow Berry's prediction for the rigidity, especially if the prescribed saturation occurs with the predicted energy dependence, the Dirichlet series for the Selberg zeta function converges for $Re\,s > \frac{1}{2}$. The arithmetic systems once again drop out of this general scheme by reasons that seem to be understood: sticking to the statistical model they are provided by the exponential degeneracies of the (pseudo-) length spectra, whereas consulting inverse quantum chaology it is the exceptional spectral statistics, showing much less correlations, that cause the trouble. Hence the arithmetic case once again exemplifies the duality of classical and quantum properties and demonstrates the mechanisms of their interplay.

It is tempting now to express the expectation that generic systems with a chaotic classical limit might show an analogous behaviour regarding the convergence properties of their dynamical zeta functions. As long as their phase spaces are compact and Pesin's theorem [75] on the equality of metric and topological entropy holds, the Dirichlet series for the dynamical zeta functions should converge conditionally for $Re\, s > \frac{\bar{\lambda}}{2} = \frac{\tau}{2}$, i.e. up to immediately before the critical line. The expectation is based on the fact that although the above inverse quantum chaology argument required detailed knowledge about the analytic properties of the Selberg zeta functions, the mechanism seems to be universal as it dwells on the magnitude of $N_{fl}(E)$. One only needs a "rigid" spectrum with a saturating spectral rigidity that can be described by Berry's general scheme. For generic chaotic systems all evidence is for this to be satisfied, and thus, although the knowledge about the associated dynamical zeta functions is much poorer than for the Selberg zeta function, at least there seems to be no obvious obstacle around against the expectation to hold. One is, however, far from proving this since the technical problems are enormous in the general case.

5 Summary

This investigation contained a discussion of the semiclassical quantization for a class of strongly chaotic systems. The relevant aspects of classical and quantum mechanics for the unconstrained motions of single particles on hyperbolic surfaces with arithmetic fundamental groups were studied. The main body of this text consisted of two major parts: chapter 3 discussed classical mechanics, i.e. the geometry of the surfaces the systems are defined on; chapter 4 then was devoted to an investigation of the quantum mechanical energy spectra of the arithmetical systems. It was worked out that the arithmetic nature of the fundamental groups involved had consequences for the geometry of the respective surfaces. In particular the length spectra of closed geodesics reveal high degrees of degeneracies.

In the context of the semiclassical quantization of classically chaotic systems the philosophy of studying "generic" systems includes the requirement of only dealing with completely desymmetrized systems. Symmetries may lead to unwanted effects that superimpose the structures one tries to explore, and in many cases they can rather easily be removed. Once the systems have been desymmetrized one can compare them irrespective of their differences in detail. Those quantities that appear to share common properties can then be used to characterize the class of chaotic dynamical systems. One manifestation of a symmetry in a classical system is the presence of degeneracies in the spectrum of periodic orbits with respect to their actions. After removing symmetries two classes of generic systems remain. The first one comprises of time-reversal invariant systems, whereas the second one consists of systems without time-reversal invariance. The philosophy referred to above continues in assuming that then periodic orbits generically can at most be twofold degenerate in action due to a time-reversal invariance. Further multiplicities would be considered as accidental. The discussion of the arithmetical systems, however, revealed that there exist perfectly chaotic Hamiltonian dynamical systems with multiplicities of lengths of periodic orbits that even grow exponentially with length. These multiplicities are not really accidental since they can be traced back to the structure of the set out of which the geodesic lengths are allowed to be taken. The arithmetic structure inherent in this set forces the lengths of closed geodesics not to cluster too strongly for $l \to \infty$. Since the total number of closed geodesics with lengths up to l has to grow according to the universal prime geodesic theorem, $\mathcal{N}(l) \sim \frac{e^l}{l}$, $l \to \infty$, the low number of distinct lengths up to l, $\hat{\mathcal{N}}(l) \sim \frac{2}{c_\Gamma} e^{l/2}$, $l \to \infty$, has to be compensated by exponentially increasing multiplicities, $< g(l) > \sim c_\Gamma \frac{e^{l/2}}{l}$, $l \to \infty$. In chapter 3 details on the mechanism resulting in the exponential law for the mean multiplicities occurring for arithmetic fundamental groups have been worked out. Since there exist infinitely many arithmetic Fuchsian groups these exceptional systems form a whole class of strongly chaotic dynamical systems that cannot be neglected. In addition, there is a further (discrete) dynamical system known with similar properties. This is the so-called *cat map*, whose classical and quantum properties have been discussed in detail by Keating [62].

It has long been known that the arithmetical systems excel by a further property, namely by the existence of infinitely many pseudosymmetries. Although their definition includes geometric symmetries, non-trivial pseudosymmetries cannot be removed in a kind of desymmetrization procedure. It might appear that the impossibility to "de-pseudosymmetrize" the arithmetical systems suffices to consider them as generic, but the discussion of multiplicities in their length spectra showed that the algebraic and geometric structures induced by pseudosymmetries are important enough to result in considerable effects. On the quantum mechanical side of the problem, which was discussed in chapter 4, these structures affect quantities that are commonly be considered as characteristic for a distinction of classically chaotic and integrable systems. In particular statistical properties of quantum mechanical energy spectra were discussed. Regarding the latter, symmetries manifest themselves as independent superpositions of spectra referring to individual symmetry classes. Thus the total quantum energy spectrum of a system possessing a discrete and finite symmetry group contains finitely many subspectra, which can each be viewed as generic, since they are spectra of desymmetrized systems. It was observed that regarding non-trivial pseudosymmetries the situation is somehow reversed. The eigenvalue spectrum of an arithmetical system is a subspectrum of infinitely many other, also arithmetic, spectra. If it were an independent superposition of infinitely many spectra, one could immediately identify the result as showing Poissonian fluctuations, see e.g. [70]. However, the fact that it is only a subspectrum in such an infinite superposition complicates the use of this point of view to draw conclusions on the spectral statistics for arithmetical systems.

This is one reason for chapter 4 to proceed differently in its investigation of arithmetic energy spectra. It appeared to be more convenient to employ the exponential growth of multiplicities in the geodesic length spectra. The spectral form factor turned out to be a useful means for a periodic-orbit investigation of spectral statistics. The two quantities that were picked out to be studied were the level spacings distribution and the number variance. The former yields information on short-range correlations, whereas the latter takes medium- and long-range correlations into account. The exponential increase of multiplicities of lengths allowed for the development of a simplified model for the form factor. This model was essentially only based on the obtained exponential increase for small τ and the saturation for $\tau \to \infty$ of the form factor. Applied to the level spacings distribution and to the number variance the model was found to describe the numerically observed phenomena qualitatively correctly. Quantum energy spectra of arithmetical systems are reminiscent of those for classically integrable systems. Their fluctuations are much stronger than those for generic classically chaotic systems. They show a level attraction that grows with increasing energy and the level spacings approach a Poissonian behaviour for $E \to \infty$. This finding is in contrast to the integrable case that seems to yield stationary distributions already at finite energies. On larger scales the correlations in arithmetical spectra appear to be slightly stronger than those observed for integrable systems. This is reflected in the energy dependence of the saturation value of the spectral rigidity. The latter was found by Berry [20] to be $\Delta_\infty(E) \sim const.\ \sqrt{E}$, $E \to \infty$, whereas the model for the number variance of arithmetical systems yielded $\Delta_\infty(E) \sim \frac{2\bar{d}}{\pi} \frac{\sqrt{E}}{\log E}$, $E \to \infty$. However, the spectral statistics in arithmetical quantum chaos are much more similar to those of classically integrable systems than to the ones of generic classically chaotic systems.

Sections 3.6 and 4.5 on fluctuations in geodesic length spectra and on convergence properties of the Selberg zeta function, respectively, had to take the different spectral statistics for the hyperbolic Laplacian on arithmetic and non-arithmetic surfaces into account. It turned out that in both sections an application of inverse quantum chaology proved useful. The desired results, however, could only be obtained in the non-arithmetical case. There Berry's observation on the saturation value of the spectral rigidity, $\Delta_\infty(E) \sim \frac{1}{2\pi^2} \log E$, $E \to \infty$, sufficed as an input to apply Hejhal's theorems [51] of inverse quantum chaology. In section 3.6 the remainder term to the leading asymptotics in the PGT thus followed to be of the form $Q_R(l) = e^{\frac{1}{2}l}\,\omega(l)$, with $\omega(l)$ denoting some unknown function containing powers and logarithms of l. The result of section 4.4 on $\Delta_\infty(E)$ for arithmetical systems could only bring down the upper bound of $e^{\frac{3}{4}l}\,\omega(l)$ for $Q_R(l)$ to $e^{\frac{2}{3}l}\,\omega(l)$. Numerical evidence obtained from three arithmetic groups, however, suggested that the exponent for the remainder term $Q_R(l)$ should also be $\frac{1}{2}$ for arithmetical systems. Thus the inapplicability of the inverse quantum chaology reasoning for arithmetic groups rather seems to be of a technical nature than of a fundamental one.

Regarding convergence properties of the Dirichlet series for the Selberg zeta function, however, the difference between the arithmetic and the non-arithmetic case seems to be not void of consequences. Using Hejhal's *Theorem 10.10* [51] of inverse quantum chaology and Berry's result on $\Delta_\infty(E)$ for generic systems revealed a convergence of the Dirichlet series at least until directly before the critical line, which is the physically interesting domain. Again, the strong spectral fluctuations present for arithmetical systems prevented an application of this method to the latter. The statistical model for the convergence properties that was introduced in [6] now hints at the reason for this obstruction. The exponentially increasing multiplicities of geodesic lengths yield an exponential growth of the multiplicities of pseudo-lengths. The exponent α describing the latter increase is a measure for the distance of the domain of convergence to the critical line.

In conclusion, the arithmetical systems that were studied in the present text excel by properties of important classical and quantum mechanical quantities that distinguish them from those of strongly chaotic systems that are commonly considered as generic. Rather convincing heuristic reasons for the exceptional spectral statistics could be derived from the classical properties of arithmetical chaos. It was mainly a combination of heuristic reasoning with the intuition gained from numerical observations that could be used in conjunction with rigorous results. This amalgam of different methods proved to be particularly fruitful. The arithmetical systems thus turned out to provide a convenient test-ground for the ideas and methods developed in the framework of periodic-orbit theory.

Acknowledgements

I want to thank Professor Frank Steiner for uncountable fruitful discussions and for his constant support throughout the time I have been his student.

I enjoyed many useful discussions with Ralf Aurich and Holger Ninnemann. Furthermore, Ralf Aurich, Claudia Matthies, Holger Ninnemann and Gunther Steil generously supplied me with their numerical data. Many thanks are due to them.

I acknowledge financial support by the University of Hamburg through a Doktorandenstipendium.

A The Riemann Zeta Function

In the main body of the present work the Selberg trace formula and the Selberg zeta function have been extensively used. Historically, Selberg introduced his formalism in close analogy to the theory of the Riemann zeta function and the distribution of prime numbers. Since many techniques appearing in the latter theory can be carried over to the case of the Selberg zeta function, some important tools that were developed to study the Riemann zeta function will be briefly introduced in this appendix. Those details that will be omitted can be found in e.g. [97, 59].

The Riemann zeta function $\zeta(s)$ is a meromorphic function for all $s \in \mathbb{C}$ that has a simple pole at $s = 1$ with residue one. For $Re\, s > 1$ it is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1} . \tag{A.1}$$

The *Dirichlet series* for $\zeta(s)$ extends over all integers n, whereas its *Euler product* runs over all primes p. The analogy to Selberg's zeta function $Z(s)$ leads to an identification of the primes p with $e^{l(\gamma)}$, where $l(\gamma)$ denotes the length of the primitive closed geodesic related to $\gamma \in \Gamma$, and an identification of the integers n with the pseudo-orbits ρ. These identifications describe what for the Selberg zeta function is the "classical" side of its theory. The "quantum" side is missing in the theory of $\zeta(s)$ in that there is no self-adjoint operator known, whose eigenvalues are related to the non-trivial zeros of the Riemann zeta function. The trivial zeros of $\zeta(s)$ are, however, explicitly known to be located at $s_k = -2k$, $k \in \mathbb{N}$. If a self-adjoint operator related to the non-trivial zeros were known, the Riemann hypothesis (RH) would be true, since then (depending on how the operator is defined) either γ_n or γ_n^2 is a real eigenvalue of it. This means that the non-trivial zeros $s_n = \frac{1}{2} \pm i\gamma_n$ lie on the critical line $Re\, s = \frac{1}{2}$. But, the RH still being unproven, the γ_n can be complex. It is only known that $0 < Re\, s_n < 1$. There is, however, tremendous evidence in favour of the RH from extensive numerical computations of non-trivial zeros [74]. Therefore it seems to be justified to assume the validity of the RH throughout, if not stated otherwise.

The importance of the Riemann zeta function derives not only from its connection to the RH, being one of the most famous unsolved problems in mathematics, but also from its decisive role played in the proof of the *prime number theorem* (PNT) and for estimating the remainder term appearing in the PNT. Thus the analytic properties of $\zeta(s)$ are essential for describing the distribution of prime numbers, which is a central issue of number theory. It will now be explained how the validity of the RH influences the magnitude of the remainder in the PNT. Let therefore

$$\pi(x) := \#\{p;\ p \le x\} \tag{A.2}$$

be the counting function of prime numbers. The PNT now states that

$$\pi(x) = li(x) + Q(x) , \qquad Q(x) = o(x/\log x) , \quad x \to \infty . \tag{A.3}$$

$\pi(x)$ may also be expressed by two other functions that have been introduced by Chebyshev. Using the *von Mangoldt function*

$$\Lambda(n) := \begin{cases} \log p & , \quad n = p^k \\ 0 & , \quad \text{otherwise} \end{cases} \tag{A.4}$$

these are defined as

$$\begin{aligned} \theta(x) &:= \sum_{p \le x} \log p , \\ \psi(x) &:= \sum_{n \le x} \Lambda(n) = \sum_{k \ge 1} \sum_{p^k \le x} \log p \\ &= \sum_{k \ge 1} \theta(x^{\frac{1}{k}}) . \end{aligned} \tag{A.5}$$

When $x^{\frac{1}{k}} < 2$ is reached the last series breakes off, i.e. when $k > \frac{\log x}{\log 2}$. Writing $\psi(x) = \theta(x) + \mathcal{R}(x)$ one can easily estimate that $\mathcal{R}(x) = O(\sqrt{x}(\log x)^2)$, $x \to \infty$. From the definition of $\theta(x)$ one obtains

$$d\theta(x) = \sum_p \log p \, \delta(x - p)\, dx , \tag{A.6}$$

and thus

$$\int_2^x \frac{d\theta(t)}{\log t} = \sum_p \log p \int_2^x \frac{dt}{\log t} \delta(t - p) = \sum_{p \le x} 1 = \pi(x) . \tag{A.7}$$

An integration by parts yields

$$\pi(x) = \int_2^x \frac{d\theta(t)}{\log t} = \frac{\theta(x)}{\log x} + \int_2^x dt \, \frac{\theta(t)}{t(\log t)^2} + O(1) . \tag{A.8}$$

Using $\theta(x) = \psi(x) + O(\sqrt{x}(\log x)^2)$ leads to

$$\pi(x) = \frac{\psi(x)}{\log x} + \int_2^x dt \, \frac{\psi(t)}{t(\log t)^2} + O(\sqrt{x} \log x) . \tag{A.9}$$

Proving the PNT is thus equivalent to determining the leading asymptotic behaviour of the Chebyshev function $\psi(x)$ for $x \to \infty$, and estimating the remainder term $Q(x)$ can be achieved by knowing the remainder to the asymptotics of $\psi(x)$.

At this stage now the Riemann zeta function enters the game. Using its Euler product for $Re\, s > 1$ one observes that

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} . \tag{A.10}$$

Employing the Cauchy integral theorem one can derive that (for $b > 1$, $a > 0$, $a \ne 1$)

$$\frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} ds \, \frac{a^s}{s} = \begin{cases} 1 & , \quad a > 1 \\ 0 & , \quad a < 1 \end{cases} . \tag{A.11}$$

This result may be used to show that ($x \notin \mathbb{N}$)

$$\psi(x) = \sum_{n=1}^{\infty} \Lambda(n) \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{ds}{s} \left(\frac{x}{n}\right)^s = -\frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} ds \, \frac{x^s}{s} \frac{\zeta'(s)}{\zeta(s)} , \tag{A.12}$$

a relation that together with (A.9) clearly shows how the PNT is related to the analytic properties of $\zeta(s)$. One can now use the Weierstraß representation of $\zeta(s)$ as a product over its zeros to obtain from (A.12) the *explicit formula* of Riemann and von Mangoldt,

$$\psi(x) = x - \sum_{s_n} \frac{x^{s_n}}{s_n} - \frac{1}{2} \log(1 - x^{-2}) - \frac{\zeta'(0)}{\zeta(0)} , \tag{A.13}$$

where the (conditionally convergent) sum runs over all non-trivial zeros $s_n = \beta_n + i\gamma_n$ of $\zeta(s)$. Denoting $\sigma_0 := \sup\{\beta_n;\ s_n = \beta_n + i\gamma_n\}$, the RH is equivalent to $\sigma_0 = \frac{1}{2}$. An estimate for the sum over the non-trivial zeros in (A.13) can be found in [59], $\sum_{s_n} \frac{x^{s_n}}{s_n} = O(x^{\sigma_0}(\log x)^2)$, leading to

$$\psi(x) = x + P(x)\,, \qquad P(x) = O(x^{\sigma_0}(\log x)^2)\,. \tag{A.14}$$

Since $\int_2^x \frac{dt}{(\log t)^2} = li(x) - \frac{x}{\log x} + O(1)$, one observes from (A.9), using (A.14), that

$$\pi(x) = li(x) + O(x^{\sigma_0}\log x)\,, \tag{A.15}$$

which is the PNT with an estimate for the remainder term $Q(x)$. This relation shows the influence of the validity of the RH on the PNT. Notice that in order that $Q(x) = o(x/\log x)$ one has to show that $\sigma_0 < 1$, i.e. $\zeta(s)$ must not have a zero on $Re\,s = 1$, nor a subsequence of zeros with real parts accumulating at 1. To show this was the main achievement of Hadamard and de la Vallée Poussin in proving the PNT in 1896.

The point of view that will be taken now introduces a connection of the RH (equivalently the PNT) to the convergence properties of a certain Dirichlet series. The general theory of Dirichlet series is presented in [48], where one can find the following results. A *generalized Dirichlet series* is a series of the form $F(s) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n s}$, $a_n \in \mathbb{C}$, $\lambda_1 < \lambda_2 < \lambda_3 < \dots$. If $\lambda_n = \log n$, it is called an *ordinary Dirichlet series*. In any case there exists a number $\sigma_c \in \mathbb{R} \cup \{\pm\infty\}$ such that the Dirichlet series converges for $Re\,s > \sigma_c$ and diverges for $Re\,s < \sigma_c$. Since the series in addition converges uniformly on compact sets, $F(s)$ is a holomorphic function in the domain of convergence. There also exists a number σ_a, $\sigma_a \geq \sigma_c$, such that the Dirichlet series converges absolutely for $Re\,s > \sigma_a$. The ***abscissae of convergence*** σ_c and σ_a are determined by

$$\begin{aligned}\sigma_c &= \limsup_{N\to\infty} \frac{1}{\lambda_N} \log\left|\sum_{n=1}^{N} a_n\right|\,,\\ \sigma_a &= \limsup_{N\to\infty} \frac{1}{\lambda_N} \log\sum_{n=1}^{N} |a_n|\,.\end{aligned} \tag{A.16}$$

Applying (A.16) to the (ordinary) Dirichlet series for $\zeta(s)$ one obtains ($a_n = 1$, $n \in \mathbb{N}$) that $\sigma_c = 1 = \sigma_a$. The location of the abscissae of convergence follows from the existence of the pole of $\zeta(s)$ at $s = 1$, which prevents the series to converge for $Re\,s \leq 1$.

Our goal now is to define a Dirichlet series that yields a meromorphic function for $s \in \mathbb{C}$ and whose abscissa of convergence σ_c is given by σ_0, i.e. by the non-trivial zero of $\zeta(s)$ with largest real part. In the vicinity of $s = 1$ the Riemann zeta function behaves like $\zeta(s) = \frac{1}{s-1} + \gamma + O((s-1))$, γ being the Euler constant. Thus $\frac{\zeta'(s)}{\zeta(s)} = -\frac{1}{s-1}$ + regular terms, $s \to 1$, is a meromorphic function with poles at $s = 1$ and at the zeros of $\zeta(s)$. Then

$$f(s) := \zeta(s) + \frac{\zeta'(s)}{\zeta(s)} \tag{A.17}$$

defines a meromorphic function that is holomorphic for $Re\,s > \sigma_0$. In the *critical strip* $0 < Re\,s < 1$ its poles are located at the non-trivial zeros s_n of $\zeta(s)$. Inserting for $Re\,s > 1$ the Dirichlet series' for $\zeta(s)$ and $\frac{\zeta'(s)}{\zeta(s)}$ yields

$$f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}\,, \qquad a_n = 1 - \Lambda(n)\,. \tag{A.18}$$

By (A.16) the abscissae of convergence for this Dirichlet series are

$$\begin{aligned}\sigma_a &= \limsup_{N\to\infty} \frac{1}{\log N} \log\sum_{n=1}^{N} |1-\Lambda(n)|\,,\\ \sigma_c &= \limsup_{N\to\infty} \frac{1}{\log N} \log\left|\sum_{n=1}^{N}(1-\Lambda(n))\right|\\ &= \limsup_{N\to\infty} \frac{1}{\log N} \log|N-\psi(N)|\\ &= \limsup_{N\to\infty} \frac{\log|P(N)|}{\log N}\,.\end{aligned} \tag{A.19}$$

Since $\Lambda(n) \neq 0$ only for $n = p^k$, and $\hat{\pi}(x) = \#\{p^k;\ p^k \leq x\} \sim \pi(x) \sim \frac{x}{\log x}$, $x \to \infty$, one concludes that $\sum_{n\leq N} |1-\Lambda(n)| \sim 2N$ for $N \to \infty$. Therefore the Dirichlet series for $f(s)$ converges absolutely for $Re\,s > \sigma_a = 1$. The interesting observation one makes with (A.19) is that the abscissa of conditional convergence is determined by the first pole that is encountered when moving with the axis $Re\,s = const.$ to the left. Namely, since σ_c has obviously to fulfill $\sigma_c \geq \sigma_0$, but the asymptotics $P(x) = O(x^{\sigma_0}(\log x)^2)$ gives $\sigma_c \leq \sigma_0$, one concludes $\sigma_c = \sigma_0$. The Dirichlet series for $f(s)$ therefore converges in the maximal possible domain. In sloppy terms one could call this Dirichlet series a "detector" for the non-trivial zero of $\zeta(s)$ with largest real part, indicating through its convergence properties.

The condition $\sigma_c \geq \sigma_0$ now also gives a lower bound for $P(x)$, namely

$$P(x) = \Omega(x^{\sigma_0 - \varepsilon}) \qquad \forall \varepsilon > 0\,. \tag{A.20}$$

Thus the true magnitude of the remainder term $Q(x)$ in the PNT is asymptotically bounded from below by $x^{\sigma_0-\varepsilon}$ for all $\varepsilon > 0$ and from above by $x^{\sigma_0}\log x$. Therefore $Q(x) = x^{\sigma_0}\cdot\omega(x)$, where $\omega(x)$ is some combination of logarithmic functions. It is thus the "leading" term x^{σ_0} that determines the fine structure in the PNT. If now the RH were true, the remainder term in the PNT would have the "minimal" asymptotic behaviour $|Q(x)| \propto \sqrt{x}\cdot\omega(x)$. In this case the best lower bound available is $\omega(x) = \Omega_{\pm}(\frac{\log\log\log x}{\log x})$ [59].

B Desymmetrizing the Hyperelliptic Involution

The general procedure of desymmetrizing the quantum problem of a particle on a hyperbolic surface possessing symmetries has been reviewed in section 4.1. This appendix now contains an explicit application of the general formulation derived in [100, 101] to a rather simple case, namely the so-called hyperelliptic involution, emphasizing the point of view employing the Selberg trace formula. A discussion of the example used here can be found in [5]. The following presentation, however, differs a little from the one given in [5] by being closer to [100, 101] in order to serve more explicitly as an example for the general situation.

The symmetry under consideration is the *hyperelliptic involution* (see e.g. [40]) present on all hyperelliptic (compact) Riemann surfaces. The latter ones can be realized as two-sheeted coverings of the sphere. If the surface $\Gamma\backslash\mathcal{H}$ is hyperelliptic and of genus g, this covering is branched at the $2g+2$ Weierstraß points. The operation that interchanges the two sheets of the covering is an involution (i.e. the symmetry group is isomorphic to $\mathbb{Z}_2$ if no other symmetries are present, as will be assumed henceforth), called the hyperelliptic involution. The hyperelliptic surfaces form a $(2g-1)$–dimensional subvariety of the $(3g-3)$–dimensional moduli space of compact Riemann surfaces of genus g; for $g=2$ all compact surfaces are hyperelliptic.

For the following it proves useful to change the model of hyperbolic geometry and to pass to the Poincaré unit-disc $\mathcal{D}=\{w\in\mathbb{C};\ |w|<1\}$ by mapping the upper half-plane $\mathcal{H}$ via $z\mapsto w=Cz$, $C=\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}1 & i\\ i & 1\end{smallmatrix}\right)$, for $z\in\mathcal{H}$. $SL(2,\mathbb{R})$ is then mapped to $SU(1,1)=\{\left(\begin{smallmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{smallmatrix}\right);\ |\alpha|^2-|\beta|^2=1\}$ by conjugation with C. A $g\in SU(1,1)$ operates on $\mathcal{D}$ via fractional linear transformations of the form

$$g\,w=\frac{\alpha w+\beta}{\bar\beta w+\bar\alpha}\ , \qquad w\in\mathcal{D}. \tag{B.1}$$

The conjugation of a Fuchsian group $\Gamma\subset SL(2,\mathbb{R})$ by C yields a discrete subgroup of $SU(1,1)$ that will also be denoted as Γ by abuse of notation. The geodesics of the appropriately transformed hyperbolic metric are the half-circles and straight lines perpendicular to $\partial\mathcal{D}=\{w\in\mathbb{C};\ |w|=1\}$. [17] may serve as a reference to find further details concerning the geometry of this model.

Any Fuchsian group Γ leading to a hyperelliptic surface $\Gamma\backslash\mathcal{D}$ may be obtained in the following manner (see also [12]). One can choose a fundamental domain $\mathcal{F}\subset\mathcal{D}$ whose boundary $\partial\mathcal{F}$ consists of $4g$ geodesic segments. The corner points $w_1,\dots,w_{4g}$ are enumerated in ascending order when going counterclockwise along $\partial\mathcal{F}$ and starting with w_1 on the positive real axis. $w_2,\dots,w_{2g}$ are placed in the upper half of $\mathcal{D}$ ($Im\ w_i>0$). The remaining corner points are obtained as $w_{2g+1}=-w_1,\dots,w_{4g}=-w_{2g}$. $\mathcal{F}$ is therefore symmetric under the operation $w\mapsto -w$. Having fixed $w_2,\dots,w_{2g}$ one has to vary w_1 (hence also $w_{2g+1}=-w_1$) on the real axis until the constraint area$(\mathcal{F})=4\pi(g-1)$ is fulfilled. The fractional linear transformations of the form (B.1) that identify opposite geodesic segments comprising $\partial\mathcal{F}$ then serve as generators for a strictly hyperbolic Fuchsian group Γ possessing $\mathcal{F}$ as a fundamental domain. This identification of pairs of edges of $\mathcal{F}$ is compatible with the symmetry $w\mapsto -w$, so that $\Gamma\backslash\mathcal{D}$ is a compact hyperelliptic surface of genus g. The above construction yields every such surface by choosing the corner points $w_2,\dots,w_{2g}$ appropriately in the upper half of $\mathcal{D}$, therefore clearly showing that the subvariety of hyperelliptic surfaces in the moduli space of compact surfaces of genus g is of (complex) dimension $2g-1$.

The generators $b_1,\dots,b_{4g}$ will be enumerated such that b_i identifies the geodesic segment connecting w_i and w_{i+1} with the one connecting $-w_i$ and $-w_{i+1}$, $i=1,\dots,2g$. By their very construction then $b_{i+2g}=b_i^{-1}$ for $i=1,\dots,2g$. The generators obey the constraint

$$b_1b_2^{-1}b_3b_4^{-1}\dots b_{2g-1}b_{2g}^{-1}\,b_1^{-1}b_2\dots b_{2g-3}^{-1}b_{2g-2}b_{2g-1}^{-1}b_{2g}=\mathbf{1}\ . \tag{B.2}$$

The hyperelliptic involution is realized as the mapping $w\mapsto -w$, which can be represented by the matrix $S=\left(\begin{smallmatrix}-i & 0\\ 0 & i\end{smallmatrix}\right)\in SU(1,1)$, $S\notin\Gamma$. Since the b_i's identify opposite edges of $\mathcal{F}$, one notices that $Sb_iS=b_i^{-1}$, where $S^2=\mathbf{1}$ as an identity in $PSU(1,1)=SU(1,1)/\{\pm\mathbf{1}\}$ has been used. The $2g+2$ fixed points of S on $\Gamma\backslash\mathcal{D}$, i.e. the Weierstraß points, can now be identified. Obviously, $u_1=0$ is fixed by S. The second Weierstraß point u_2 is represented by the corner points of $\mathcal{F}$ that are all being identified to one another by Γ. The remaining fixed points $u_3,\dots,u_{2g+2}$ are given by the mid-points of the $2g$ pairs of opposite edges of $\mathcal{F}$ as determined by the hyperbolic metric.

A new Fuchsian group Γ' is now introduced by adjoining S to Γ, i.e. Γ' consists of all words in the generators $b_1,\dots,b_{4g},S$, subject to the constraint (B.2) and fulfilling $b_i^{-1}=b_{i+2g}$, $S^2=\mathbf{1}$, $Sb_iS=b_i^{-1}$. It is thus possible to rewrite any word in those generators as being of the form $S^\epsilon b_{i_1}\dots b_{i_n}$, $\epsilon\in\{0,1\}$. One therefore obtains that Γ is a normal subgroup of index two in Γ'.

The latter decomposes disjointly according to

$$\Gamma'=\Gamma\,\dot\cup\,S\Gamma\ , \tag{B.3}$$

thus reproducing for the hyperelliptic involution S the algebraic setting of symmetries reviewed in section 3.5. The symmetry group Σ is yielded as $\Sigma=\{\mathbb{1},S\}$, $S^2=\mathbb{1}$. Thus $\Sigma\cong\mathbb{Z}_2$, and also for the unitary dual $\Sigma^*\cong\mathbb{Z}_2$. Explicitly, Σ^* is given by the two representations χ_+ and χ_- of Γ'; χ_+ denotes the trivial representation $\chi_+(\gamma')=1$ for all $\gamma'\in\Gamma'$, whereas χ_- is defined as

$$\chi_-(\gamma'):=\begin{cases}+1\ , & \gamma'\in\Gamma\\ -1\ , & \gamma'\in S\Gamma\end{cases}\ . \tag{B.4}$$

Since $S\in SU(1,1)$ is elliptic ($|tr\,S|<2$), Γ' is not strictly hyperbolic. Fortunately, it is possible to identify all elliptic conjugacy classes of Γ' explicitly by their fixed points on the surface $\Gamma'\backslash\mathcal{D}$. An elliptic $R\in\Gamma'$ has one fixed point z_R in the interior of $\mathcal{D}$, and its conjugacy class $\{R\}_{\Gamma'}$ fixes the set of points $\Gamma' z_R$ that are identified under Γ'. To each elliptic conjugacy class there hence corresponds the point $\Gamma' z_R$ on the surface $\Gamma'\backslash\mathcal{D}$; but this must be one of the Weierstraß points $u_1,\dots,u_{2g+2}$. The Fuchsian group Γ' therefore contains $2g+2$ elliptic conjugacy classes $\{R\}_{\Gamma'}$, all of them of order $m(R)=2$ (meaning the minimal positive integer with $R^{m(R)}=\pm\mathbf{1}$). One can easily determine representatives for the elliptic classes,

- u_1 is fixed by S,
- u_2 is fixed by $Sb_1b_2^{-1}b_3b_4^{-1}\dots b_{2g-1}b_{2g}^{-1}$,
- u_i is fixed by Sb_{i-2}^{-1}, for $i=3,\dots,2g+2$.

All other conjugacy classes in Γ' are hyperbolic ones.

The general receipt of [101] now proceeds in constructing the unitary representation of Γ' that is received as being induced from the trivial representation of its subgroup Γ. To this end define the one dimensional representation (on $\mathbb{C}$)

$$\overline{\chi}(\gamma'):=\begin{cases}1\ , & \gamma'\in\Gamma\\ 0\ , & \gamma'\in S\Gamma\end{cases}\ . \tag{B.5}$$

The *induced representation* $\rho:\ \Gamma'\to\mathrm{End}(\mathbb{C}\oplus\mathbb{C})$ is then obtained as

$$\rho(\gamma')(v\oplus w):=[\overline{\chi}(\gamma')v+\overline{\chi}(\gamma'S)w]\oplus[\overline{\chi}(S\gamma')v+\overline{\chi}(S\gamma'S)w]\ , \tag{B.6}$$

for $\gamma'\in\Gamma'$ and $v,w\in\mathbb{C}$. On Γ this representation operates trivially, $\rho(\gamma)(v\oplus w)=v\oplus w$, $\gamma\in\Gamma$; on $S\Gamma$, however, it acts according to $\rho(S\gamma)(v\oplus w)=w\oplus v$, $\gamma\in\Gamma$. As a matrix representation on $\mathbb{C}^2$ the induced representation is hence given by

$$\rho(\gamma')=\begin{cases}\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)\ , & \gamma'\in\Gamma\\ \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)\ , & \gamma'\in S\Gamma\end{cases}\ . \tag{B.7}$$

Later, in the Selberg trace formula, $tr\,\rho(\gamma')^k$, $k\in\mathbb{N}$, is needed. For $\gamma\in\Gamma$, clearly $tr\,\rho(\gamma)^k=2$, but for $\gamma'\in S\Gamma$ one obtains that $tr\,\rho(\gamma')^k=1+(-1)^k$. Altogether, one can reformulate this as $tr\,\rho(\gamma')^k=\chi_+(\gamma')^k+\chi_-(\gamma')^k$ for all $\gamma'\in\Gamma'$.

Venkov and Zograf now demonstrate [101] that the hyperbolic and the elliptic contributions to the Selberg trace formulae for Γ endowed with the trivial representation and for Γ' endowed with the induced representation ρ coincide, respectively. The hyperbolic terms thus yield

$$\begin{aligned}\sum_{\{\gamma\}_\Gamma}\sum_{k=1}^{\infty}\frac{l(\gamma)\,g(kl(\gamma))}{2\sinh(\frac{k}{2}l(\gamma))} &= \sum_{\{\gamma'\}_{\Gamma'}}\sum_{k=1}^{\infty}\frac{tr\,\rho(\gamma')^k\,l(\gamma')\,g(kl(\gamma'))}{2\sinh(\frac{k}{2}l(\gamma'))} \\ &= \sum_{\{\gamma'\}_{\Gamma'}}\sum_{k=1}^{\infty}\frac{l(\gamma')\,g(kl(\gamma'))}{2\sinh(\frac{k}{2}l(\gamma'))} + \sum_{\{\gamma'\}_{\Gamma'}}\sum_{k=1}^{\infty}\frac{\chi_-(\gamma')^k\,l(\gamma')\,g(kl(\gamma'))}{2\sinh(\frac{k}{2}l(\gamma'))}\,,\end{aligned} \tag{B.8}$$

where the outer sums extend over the respective primitive hyperbolic conjugacy classes. This result yields a factorization formula for the Selberg zeta function,

$$Z_\Gamma(s) = Z^+_{\Gamma'}(s)\cdot Z^-_{\Gamma'}(s)\,, \tag{B.9}$$

where for $Re\, s > 1$

$$\begin{aligned} Z^+_{\Gamma'}(s) &:= \prod_{\{\gamma'\}_{\Gamma'}}\prod_{n=0}^{\infty}\left(1 - e^{-(s+n)l(\gamma')}\right)\,, \\ Z^-_{\Gamma'}(s) &:= \prod_{\{\gamma'\}_{\Gamma'}}\prod_{n=0}^{\infty}\left(1 - \chi_-(\gamma')\,e^{-(s+n)l(\gamma')}\right)\,, \end{aligned} \tag{B.10}$$

which should be compared with (4.2) and (4.3). The factorization (B.10) comprises the desymmetrization with respect to the hyperelliptic involution on the level of the Selberg zeta function because the non-trivial zeros of $Z^\pm_{\Gamma'}(s)$ correspond to the eigenvalues of $-\Delta$ referring to the two symmetry classes of S. The reason for this is that the respective wavefunctions transform under $\chi_\pm$ according to $\psi_\pm(\gamma' z) = \chi_\pm(\gamma')\psi_\pm(z)$, $\gamma' \in \Gamma'$. Hence the $\psi_\pm(z)$ are invariant under Γ ($\chi_\pm(\gamma) = 1$ for $\gamma \in \Gamma$), and transform under S as $\psi_\pm(Sz) = \chi_\pm(S)\psi_\pm(z) = \pm\psi_\pm(z)$.

Finally it should be remarked that the explicit knowledge of the elliptic conjugacy classes of Γ' allows to determine the elliptic contributions to the Selberg trace formulae for Γ' endowed with the representations $\chi_\pm$ as

$$\pm\,\frac{g+1}{4}\int_{-\infty}^{+\infty} dp\;\frac{h(p)}{\cosh(\pi p)}\,, \tag{B.11}$$

see [5]. There one can also find a numerical evaluation of the Selberg zeta functions $Z^\pm_{\Gamma'}(s)$ on the critical line for a specific surface of genus $g = 2$.

C Estimates of Remainder Terms

In analytic number theory several estimates are used to describe remainder terms to the leading asymptotic behaviour of functions of interest, the most prominent example being the remainder $Q(x)$ in the PNT, see appendix A. Since in other fields, like in physics, some of these estimates are not so commonly used, their definitions will be supplied in this appendix.

Let therefore $f(x)$ be a function of the real variable $x \geq 0$ that shall be estimated for $x \to \infty$ and compared to the positive and monotonic function $g(x)$.

The first estimate is

$$f(x) = O(g(x)) \quad :\Leftrightarrow \quad \limsup_{x\to\infty}\frac{|f(x)|}{g(x)} < \infty\,. \tag{C.1}$$

Asymptotically $|f(x)|$ is thus bounded by $g(x)$, which might therefore also be referred to as an *upper bound*.

Another, stronger, upper bound is

$$f(x) = o(g(x)) \quad :\Leftrightarrow \quad \limsup_{x\to\infty}\frac{|f(x)|}{g(x)} = 0\,. \tag{C.2}$$

A sort of asymptotic *lower bounds* is provided by Ω–estimates, which are given as

$$\begin{aligned} f(x) = \Omega(g(x)) \quad &:\Leftrightarrow \quad \limsup_{x\to\infty}\frac{|f(x)|}{g(x)} > 0\,, \\ f(x) = \Omega_+(g(x)) \quad &:\Leftrightarrow \quad \limsup_{x\to\infty}\frac{f(x)}{g(x)} > 0\,, \\ f(x) = \Omega_-(g(x)) \quad &:\Leftrightarrow \quad \liminf_{x\to\infty}\frac{f(x)}{g(x)} < 0\,. \end{aligned} \tag{C.3}$$

The O–estimate can also be formulated slightly differently: $f(x) = O(g(x))$, if there exist some $x_0 \geq 0$ and $M > 0$ such that $|f(x)| \leq M\,g(x)$ for $x \geq x_0$. Analogously, $f(x) = \Omega(g(x))$, if $|f(x)| \geq M\,g(x)$ for $x \geq x_0$, etc.

References

[1] V.I. Arnold, *Mathematical Methods in Classical Mechanics* (Springer, Berlin–Heidelberg–New York, 1980)

[2] E. Artin, *Ein mechanisches System mit quasiergodischen Bahnen*, Abh. Math. Sem. Univ. Hamburg **3**(1924)170-175

[3] R. Aurich, *private communication*

[4] R. Aurich, E.B. Bogomolny and F. Steiner, *Periodic Orbits on the Regular Hyperbolic Octagon*, Physica **D48**(1991)91-101

[5] R. Aurich and J. Bolte, *Quantization Rules for Strongly Chaotic Systems*, Mod. Phys. Lett. **B6**(1992)1691-1719

[6] R. Aurich, J. Bolte, C. Matthies, M. Sieber and F. Steiner, *Crossing the Entropy Barrier of Dynamical Zeta Functions*, Physica **D63**(1993)71-86

[7] R. Aurich, C. Matthies, M. Sieber and F. Steiner, *Novel Rule for Quantizing Chaos*, Phys. Rev. Lett. **68**(1992)1629-1632

[8] R. Aurich and M. Sieber, *An Exponentially Increasing Spectral Form Factor $K(\tau)$ for a Class of Strongly Chaotic Systems*, DESY-preprint, DESY 92-171, 1992

[9] R. Aurich, M. Sieber and F. Steiner, *Quantum Chaos of the Hadamard-Gutzwiller Model*, Phys. Rev. Lett. **61**(1988)483-487

[10] R. Aurich and F. Steiner, *On the Periodic Orbits of a Strongly Chaotic System*, Physica **D32**(1988)451-460

[11] R. Aurich and F. Steiner, *Periodic-Orbit Sum Rules for the Hadamard-Gutzwiller Model*, Physica **D39**(1989)169-193

[12] R. Aurich and F. Steiner, *Energy-Level Statistics of the Hadamard-Gutzwiller Ensemble*, Physica **D43**(1990)155-180

[13] R. Aurich and F. Steiner, *From Classical Periodic Orbits to the Quantization of Chaos*, Proc. Roy. Soc. London **A437**(1992)693-714

[14] R. Aurich and F. Steiner, *Staircase Functions, Spectral Rigidity and a Rule for Quantizing Chaos*, Phys. Rev. **A45**(1992)583-592

[15] R. Aurich and F. Steiner, *Asymptotic Distribution of the Pseudo-Orbits and the Generalized Euler Constant γ_Δ for a Family of Strongly Chaotic Systems*, Phys. Rev. **A46**(1992)771-781

[16] R. Aurich and F. Steiner, *Statistical Properties of Highly Excited Quantum Eigenstates of a Strongly Chaotic System*, Physica **D**(1993), in the press

[17] N.L. Balazs and A. Voros, *Chaos on the Pseudosphere*, Phys. Rep. **143**(1986)109-240

[18] P.T. Bateman and H.G. Diamond, *Asymptotic Distribution of Beurling's Generalized Prime Numbers*, in *Studies in Number Theory, Vol. 6*, pp.153-211 (The Mathematical Association of America, Englewood Cliffs, 1969)

[19] M.V. Berry, *Regular and Irregular Semiclassical Wavefunctions*, J. Phys. **A10**(1977)2083-2091

[20] M.V. Berry, *Semiclassical Theory of Spectral Rigidity*, Proc. Roy. Soc. London **A400**(1985)229-251

[21] M.V. Berry, *Semiclassical Formula for the Number Variance of the Riemann Zeros*, Nonlinearity **1**(1988)399-407

[22] M.V. Berry and J.P. Keating, *A Rule for Quantizing Chaos?*, J. Phys. **A23**(1990)4839-4849

[23] M.V. Berry and J.P. Keating, *A New Asymptotic Representation for $\zeta(\frac{1}{2}+it)$ and Quantum Spectral Determinants*, Proc. Roy. Soc. London **A437**(1992)151-173

[24] M.V. Berry and M. Tabor, *Level Clustering in the Regular Spectrum*, Proc. Roy. Soc. London **A356**(1977)375-394

[25] P.M. Bleher, *Distribution of Energy Levels in Quantum Systems with Integrable Classical Counterpart. Rigorous Results*, in *Mathematical Physics X*, Proc. of the X^{th} Congress on Mathematical Physics, Leipzig 1991, pp. 298-302, ed.: K. Schmüdgen (Springer, Berlin–Heidelberg–New York, 1992)

[26] E.B. Bogomolny, *Semiclassical Quantization of Multidimensional Systems*, Nonlinearity **5**(1992)805-866

[27] E.B. Bogomolny, B. Georgeot, M.-J. Giannoni and C. Schmit, *Chaotic Billiards Generated by Arithmetic Groups*, Phys. Rev. Lett. **69**(1992)1477-1480

[28] E.B. Bogomolny, B. Georgeot, M.-J. Giannoni and C. Schmit, *Trace Formulae for Arithmetical Systems*, Orsay–preprint IPNO/TH 92-78, 1992

[29] E.B. Bogomolny and C. Schmit, *Semiclassical Computations of Energy Levels*, Orsay–preprint, IPNO/TH 92-77, 1992

[30] O. Bohigas, M.-J. Giannoni and C. Schmit, *Characterization of Chaotic Quantum Spectra and Universality of Level Fluctuation Laws*, Phys. Rev. Lett. **52**(1984)1-4

[31] O. Bohigas, M.-J. Giannoni and C. Schmit, *Spectral Fluctuations of Classically Chaotic Systems*, in *Quantum Chaos and Statistical Nuclear Physics*, eds.: T.H. Seligman and H. Nishioka, Springer Lecture Notes in Physics **263**(1986)18-40

[32] J. Bolte, G. Steil and F. Steiner, *Arithmetical Chaos and Violation of Universality in Energy Level Statistics*, Phys. Rev. Lett. **69**(1992)2188-2191

[33] J. Bolte, *Periodic Orbits in Arithmetical Chaos*, DESY-preprint, DESY 92-104, 1992

[34] S.I. Borewicz and I.R. Šafarevič, *Zahlentheorie* (Birkhäuser, Basel–Stuttgart, 1966)

[35] T.A. Brody, J. Flores, P.A. Mores, A. Pandey, and S.S.M. Wong, *Random-Matrix Physics: Spectrum and Strength Fluctuations*, Rev. Mod. Phys. **53**(1981)385-479

[36] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces* (Birkhäuser, Boston–Basel–Berlin, 1992)

[37] I.P. Cornfeld, S.V. Fomin and Ya.G. Sinai, *Ergodic Theory* (Springer, Berlin–Heidelberg–New York, 1982)

[38] F.J. Dyson and M.L. Mehta, *Statistical Theory of the Energy Levels of Complex Systems. IV*, J. Math. Phys. **4**(1963)701-719

[39] A. Erdélyi (ed.), *Higher Transcendental Functions, Vol. II* (McGraw–Hill, New York, 1953)

[40] H.M. Farkas and I. Kra, *Riemann Surfaces* (Springer, Berlin–Heidelberg–New York, 1980)

[41] M.-J. Giannoni, A. Voros and J. Zinn-Justin (eds.), *Chaos and Quantum Physics*, Proc. of the Les Houches Summer School 1989 (North-Holland, Amsterdam, 1991)

[42] M.C. Gutzwiller, *Phase-Integral Approximation in Momentum Space and the Bound States of an Atom*, J.Math.Phys. **8**(1967)1979-2000; *Phase-Integral Approximation in Momentum Space and the Bound States of an Atom II*, J.Math.Phys. **10**(1969)1004-1020; *Energy Spectrum According to Classical Mechanics*, J.Math.Phys. **11**(1970)1791-1806; *Periodic Orbits and Classical Quantization Conditions*, J.Math.Phys. **12**(1971)343-358

[43] M.C. Gutzwiller, *Classical Quantization with Ergodic Behavior*, Phys. Rev. Lett. **45**(1980)150-153

[44] M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, Berlin–Heidelberg–New York, 1990)

[45] F. Haake, *Quantum Signatures of Chaos* (Springer, Berlin–Heidelberg, 1991)

[46] J. Hadamard, *Les surfaces à courbures opposées et leur lignes géodésique*, J. de Math. pure et appl. **4**(1898)27-73

[47] J.H. Hannay and A.M. Ozorio de Almeida, *Periodic Orbits and a Correlation Function for the Semiclassical Density of States*, J. Phys. **A17**(1984)3429-3440

[48] G.H. Hardy and M. Riesz, *The General Theory of Dirichlet's Series* (Cambridge Univ. Press, Cambridge, 1915)

[49] E. Hecke, *Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung I*, Math. Annalen **114**(1937)1-28; *dito II*, Math. Annalen **114**(1937)316-351

[50] D.A. Hejhal, *The Selberg Trace Formula and the Riemann Zeta Function*, Duke Math. J. **43**(1976)441-482

[51] D.A. Hejhal, *The Selberg Trace Formula for $PSL(2,\mathbb{R})$, Vol. I and II*, Springer Lecture Notes in Mathematics **548**(1976) and **1001**(1983)

[52] D.A. Hejhal, *Eigenvalues of the Laplacian for $PSL(2,\mathbb{Z})$: Some New Results and Computational Techniques*, in *International Symposium in Memory of Hua Loo-Keng, Vol. I*, pp. 59-102, eds.: S. Gong, Q. Lu Y. Wang and L. Yang (Science Press, New York, 1991)

[53] D.A. Hejhal and S. Arno, *On Fourier Coefficients of Maaß Waveforms for $PSL(2,\mathbb{Z})$*, Chalmers University of Technology and Göteborg University preprint No. 1991-24, 1991 (to appear in Math. of Comp., Lehmer Memorial Vol.)

[54] D.A. Hejhal and B. Rackner, *On the Topography of Maaß Waveforms for $PSL(2,\mathbb{Z})$: Experiments and Heuristics*, Univ. of Minnesota Supercomputer Institute Research Report UMSI 92/162, 1992

[55] D. Hilbert, *Die Theorie der algebraischen Zahlkörper*, Jahresberichte der Deutschen Mathematikervereinigung **4**(1897)175-546, reprinted in: D. Hilbert, *Gesammelte Abhandlungen*, Band I: Zahlentheorie, pp. 63-363 (Springer, Berlin–Heidelberg–New York, 1970)

[56] E. Hopf, *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91**(1939)261-304

[57] R.D. Horowitz, *Characters of Free Groups Represented in the Two-Dimensional Special Linear Group*, Commun. Pure and Appl. Math. **25**(1972)635-649

[58] H. Huber, *Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen*, Math. Annalen **138**(1959)1-26

[59] A.E. Ingham, *The Distribution of Prime Numbers* (Cambridge Univ. Press, Cambrige, 1932)

[60] H. Iwaniec, *Prime Geodesic Theorem*, J. Reine Angew. Math. **349**(1984)136-159

[61] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory* (J. Wiley, New York, 1959)

[62] J.P. Keating, *Asymptotic Properties of the Periodic Orbits of the Cat Map*, Nonlinearity **4**(1991)277-307; *The Cat Map: Quantum Mechanics and Classical Motion*, Nonlinearity **4**(1991)309-341

[63] J.P. Keating, *Periodic Orbit Resummation and the Quantization of Chaos*, Proc. Roy. Soc. London **A436**(1991)99-108

[64] E. Krätzel, *Zahlentheorie* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1981)

[65] S. Lang, *Algebraic Numbers* (Addison–Wesley, Reading–Palo Alto–London, 1964)

[66] C.G. Latimer and C.C. MacDuffee, *A Correspondence between Classes of Ideals and Classes of Matrices*, Annals of Math. **34**(1933)313-316

[67] G.A. Margulis, *Arithmetic Properties of Discrete Subgroups*, Russ. Math. Surveys **29**(1974)107-156

[68] C. Matthies and F. Steiner, *Selberg's ζ-Function and the Quantization of Chaos*, Phys. Rev. **A44**(1991)R7877-R7880

[69] H.P. McKean, *Selberg's Trace Formula as Applied to a Compact Riemann Surface*, Commun. Pure and Appl. Math. **25**(1972)225-246

[70] M.L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels* (Academic Press, New York, 1967); new revised and enlarged edition (Academic Press, New York, 1990)

[71] T. Miyake, *Modular Forms* (Springer, Berlin–Heidelberg–New York, 1989)

[72] H.L. Montgomery, *Extreme Values of the Riemann Zeta Function*, Comm. Math. Helv. **52**(1977)511-518

[73] H. Ninnemann, (PhD Thesis, Univ. Hamburg, in preparation)

[74] A.M. Odlyzko, *On the Distribution of Spacings Between Zeros of the Zeta Function*, Math. of Comp. **48**(1987)273-308; *The 10^{20}th Zero of the Riemann Zeta Function and 175 Million of its Neighbours*, preprint, 1991

[75] Yu. B. Pesin, *Lyapunov Characteristic Exponents and Ergodic Properties of Smooth Dynamical Systems with an Invariant Measure*, Sov. Math. Dokl. **17**(1976)196-199

[76] T. Pignataro, *Hausdorff Dimension, Spectral Theory and Applications to the Quantization of Geodesic Flows on Surfaces of Constant Negative Curvature* (PhD Thesis, Princeton Univ., 1984)

[77] C.E. Porter (ed.), *Statistical Theory of Spectra: Fluctuations* (Academic Press, New York, 1965)

[78] B. Randol, *The Length Spectrum of a Riemann Surface is Always of Unbounded Multiplicity*, Proc. of the AMS **78**(1980)455-456

[79] P. Sarnak, *Arithmetic Quantum Chaos*, Schur Lectures, Tel Aviv, 1992

[80] P. Sarnak, *Arithmetic Quantum Chaos*, Talk given at the Ecole Normale Supérieure, Paris, January 13^{th}, 1993

[81] D. Schleicher, *Bestimmung des Energiespektrums in einem chaotischen System mit Hilfe der Selbergschen Spurformel* (Diploma Thesis, Univ. Hamburg, 1991)

[82] C. Schmit, *Triangular Billiards on the Hyperbolic Plane: Spectral Properties*, Orsay-preprint IPNO/TH 91-68, 1991

[83] P. Schmutz, *Small Eigenvalues on Riemann Surfaces of Genus 2*, Invent. Math. **106**(1991)121-138

[84] A. Selberg, *Harmonic Analysis and Discontinuous Groups in Weakly Symmetric Riemannian Spaces with Application to Dirichlet Series*, J. Indian Math. Soc. **20**(1956)47-87

[85] G. Shimura, *On the Zeta-Function of the Algebraic Curves Uniformized by Certain Automorphic Functions*, J. Math. Soc. Japan **13**(1961)275-331

[86] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions* (Princeton University Press, Princeton, 1971)

[87] M. Sieber, *The Hyperbola Billiard: A Model for the Semiclassical Quantization of Chaotic Systems*, DESY-preprint, DESY 91-030 (PhD Thesis, Univ. Hamburg, 1991)

[88] M. Sieber and F. Steiner, *Generalized Periodic-Orbit Sum Rules for Strongly Chaotic Systems*, Phys.Lett. **A144**(1990)159-163

[89] M. Sieber and F. Steiner, *Quantization of Chaos*, Phys. Rev. Lett. **67**(1991)1941-1944

[90] Ya.G. Sinai, *Mathematical Problems in the Theory of Quantum Chaos*, Distinguished Raimond and Beverly Sackler Lectures, Tel Aviv University, 1990 (to appear in Springer Lecture Notes in Mathematics)

[91] G. Steil, *Über das diskrete Energiespektrum des Artinschen Billards* (Diploma Thesis, Univ. Hamburg, 1992)

[92] F. Steiner, Talk given at ITP, University of California, Santa Barbara, May 1989

[93] H.P.F. Swinnerton-Dyer, *Arithmetic Groups*, in *Discrete Groups and Automorphic Functions*, Proc. of an international conference organized by the London Mathematical Society and the Univ. of Cambridge, Cambridge 1975, pp. 377-401 (Academic Press, London–New York–San Francisco, 1977)

[94] K. Takeuchi, *On some Discrete Subgroups of* $SL_2(\mathbb{R})$, J. Fac. Sci. Univ. Tokyo **16**(1969)97-100; *A Characterization of Arithmetic Fuchsian Groups*, J. Math. Soc. Japan **27**(1975)600-612

[95] G. Tanner, P. Scherer, E.B. Bogomolny, B. Eckhardt and D. Wintgen, *Quantum Eigenvalues from Classical Periodic Orbits*, Phys. Rev. Lett. **67**(1991)2410-2413

[96] A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications I* (Springer, Berlin–Heidelberg–New York, 1985)

[97] E.C. Titchmarsh, *The Theory of the Riemann Zeta-function*, second edition, revised by D.R. Heath-Brown (Clarendon Press, Oxford, 1986)

[98] B.L. van der Waerden, *Algebra I*, 5. Auflage (Springer, Berlin–Göttingen–Heidelberg, 1960)

[99] A.B. Venkov, *Selberg's Trace Formula for the Hecke Operator Generated by an Involution, and the Eigenvalues of the Laplace-Beltrami Operator on the Fundamental Domain of the Modular Group* $PSL(2,\mathbb{Z})$, Math. USSR Izvestiya **12**(1978)448-462

[100] A.B. Venkov, *Spectral Theory of Automorphic Functions*, Proc. Steklov Math. Inst. **153**(1981)

[101] A.B. Venkov and P.G. Zograf, *On Analogues of the Artin Factorization Formulas in the Spectral Theory of Automorphic Functions Connected with Induced Representations of Fuchsian Groups*, Math. USSR Izvestiya **21**(1983)435-443

[102] M.-F. Vignéras, *Arithmétique des Algèbres de Quaternions*, Springer Lecture Notes in Mathematics **800**(1980)

[103] A. Weil, *Basic Number Theory* (Springer, Berlin–Heidelberg–New York, 1967)