

# THE SPECTRUM OF THE LAPLACIAN ON $\Gamma \backslash \mathfrak{H}$ & THE TRACE FORMULA

OWEN BARRETT

ABSTRACT. Maaß cusp forms on  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$  are the basic building blocks of automorphic forms on  $\mathrm{GL}(2)$ . Yet, their existence is known almost entirely through indirect means; no one has found a single example of an everywhere-unramified cuspidal Maaß form for  $\mathrm{PSL}(2, \mathbf{Z})$ . The spectral theory of  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$  is deeply important and remains exceedingly mysterious. In this expository article, we introduce some of the key questions in this area, and the main tool used to study them, the Selberg trace formula.

## CONTENTS

1. Introduction . . . . .	2
2. The Trace Formula: Compact Case . . . . .	6
3. The Trace Formula: Arithmetic Case . . . . .	12
3.1. Introducing the Selberg Trace Formula on $\Gamma \backslash \mathfrak{H}$ . . . . .	14
3.2. The Spectral Decomposition of $\Gamma \backslash \mathfrak{H}$ . . . . .	17
3.3. The Selberg Trace Formula on $\Gamma \backslash \mathfrak{H}$ . . . . .	25
4. The Spectrum in Detail . . . . .	36
4.1. The Continuous Spectrum: Eisenstein series on $\Gamma \backslash \mathfrak{H}$ . . . . .	36
4.2. The Discrete Cuspidal Spectrum: Maaß forms on $\Gamma \backslash \mathfrak{H}$ . . . . .	40
4.3. The Ramanujan Conjecture . . . . .	50
4.4. Weyl's Law . . . . .	51
4.5. Quantum Ergodicity . . . . .	54
4.6. Duke's Theorem . . . . .	56
References . . . . .	59

*Elle est retrouvée.  
Quoi ? L'Éternité.  
C'est la mer allée  
Avec le soleil.  
Arthur Rimbaud.*

---

Date: 27 April 2015.

The author would like to thank Peng Zhao, who supervised this essay, for many generous conversations.

## 1. INTRODUCTION

The goal of this exposition is to discuss the spectrum of the Laplace-Beltrami operator  $\Delta$  on (compact and arithmetic) manifolds given by quotients of the upper half-plane  $\mathfrak{H} = \{z \in \mathbb{C} : \Im z > 0\}$ , and to introduce the main tool available to study the structure of the spectrum; the Selberg trace formula.

On a compact hyperbolic surface  $S$ ,  $\Delta$  can be interpreted as the quantum Hamiltonian of a particle whose classical dynamics are described by the (chaotic) geodesic flow on the unit tangent bundle of  $S$ . The spectrum of  $\Delta$  on  $S$  resists straightforward characterization, and indeed it is only by considering the spectrum all at once that Selberg was able to find the identity that bears his name. He called this identity a ‘trace formula’ since it was obtained by computing the trace of an integral operator on  $S$  of trace class in two different ways. This method generalizes the Poisson summation formula, which is obtained in precisely the same way. In every case in which a trace formula exists, the formula exploits the duality afforded by Fourier theory to relate, on average, a collection of spectral data to a collection of geometric data.

An understanding of the spectrum and eigenfunctions of the Laplacian on a compact Riemannian manifold  $S$  is of interest to physicists. For example, a basic problem that probes the interaction between classical and quantum mechanics is the question of whether or not the  $L^2$  mass of high-energy eigenfunctions becomes equidistributed over  $S$  in the limit. It is known that when the geodesic flow on  $S$  (corresponding to the classical dynamics) is ergodic, such equidistribution is achieved generically; we call such equidistribution *quantum ergodicity*. We would also like to know whether the genericity hypothesis may be lifted; if not, there would be a subset of ‘bad’ eigenfunctions that resist equidistribution. If we need no assumption on genericity, and *all* eigenfunctions equidistribute in the limit of high energy, we say the system exhibits *quantum unique ergodicity*. This phenomenon, and related conjectures, are discussed in §4.5.

Automorphic forms and their associated L-functions are principal objects of study by many contemporary number theorists. Perhaps the most classical motivation may be found in the connection between modular forms and elliptic integrals, or modular functions’ relationship to generating functions for combinatorial and number-theoretic quantities, of which the connection between the partition function and the Dedekind eta function is but one example. More recently, connections have been made between the coefficients of certain modular forms and the representation theory of some sporadic groups. In contemporary number theory, modularity of elliptic curves has connected holomorphic cusp forms to elliptic curves, and to the theory of Galois representations. The automorphy of Galois representations, or how to obtain a Galois representation from an automorphic form, are questions that are of central interest to number theorists today. Statements relating to special values of motivic L-functions, some of which are theorems, like Dirichlet’s class number formula, most of which are conjectures, among them, the Birch and Swinnerton-Dyer conjecture, and, more generally, the equivariant Tamagawa conjecture, have provided clear evidence that global information of the utmost importance is encoded in L-functions attached

to arithmetic objects. Langlands's conjectures, among them, (a) the functoriality conjecture, which, loosely speaking, posits the existence of a kind of functorial relationship between automorphic forms and automorphic L-functions (operations on L-functions should correspond to operations [lifts] on automorphic forms, and vice versa), (b) conjectures extending endoscopic transfer of automorphic forms to more exotic sorts of transfer, and (c) conjectures that suggest that all L-functions of motivic origin should factor as products of L-functions belonging to a certain class of 'standard' automorphic L-functions, all contribute to making the study of automorphic forms, representations, and L-functions a major focus of contemporary number theory.

The majority of meaningful statements one could hope to make about L-functions attached to automorphic forms, such as subconvexity bounds, distributional statistics of critical zeros and, in particular, low-lying zeros, bounds on residues, moments, bounds and average values at the central point, all require statements about spectral properties of the automorphic forms themselves. On the other hand, spectral statements about automorphic forms and related objects have been found to have striking and ingenious applications to statements as seemingly unrelated as equidistribution of geodesics on the modular surface (c.f. §4.6). Trace formulæ are the primary route to obtain necessary spectral insight, and are crucial ingredients in obtaining statements like the above. Trace formulæ are also crucial for understanding finer points about the decomposition of the spectrum of the Laplacian on arithmetic manifolds such as Weyl's law.

For these reasons, we will focus most of our attention in this article on the arithmetic manifold  $\Gamma \backslash \mathfrak{H}$ , where  $\Gamma = \mathrm{PSL}(2, \mathbf{Z})$ . The modular surface  $\Gamma \backslash \mathfrak{H}$  is not compact, and therefore the spectrum of the Laplace-Beltrami operator admits a continuous part spanned by Eisenstein series, as discussed in §4.1. The discrete part of the spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H})$  is spanned by Maaß cusp forms; as discussed in §4.2, these are the eigenfunctions that are the primary object of study in arithmetic quantum unique ergodicity, which is now a theorem thanks to work of Lindenstrauss [38] and Soundararajan [61]. Those Maaß cusp forms that are also eigenfunctions of the Hecke operators on the modular surface are real-analytic but not holomorphic, and are the central automorphic objects on  $\mathrm{GL}(2)$ , but remain considerably more mysterious than their holomorphic counterparts. Even writing down even Hecke-Maaß forms for  $\Gamma \backslash \mathfrak{H}$  has proven elusive, and associating a Galois representation to a Maaß form is not known in general.

Additionally, any connection to motivic cohomology, natural in the case of holomorphic Hecke eigencuspforms, is absent. In particular, this means that Deligne's proof [11] of the Ramanujan conjecture on Fourier coefficients (i.e. Hecke eigenvalues) of holomorphic cusp forms does not apply to Maaß cusp form coefficients. Indeed, though the Ramanujan conjecture for a Hecke-Maaß cusp form  $f$  would follow from the automorphy of all symmetric power L-functions  $L(s, \mathrm{sym}^r f)$ , this is likely out of reach in the near future, and the best bounds towards  $\mathrm{GL}(2)$  Ramanujan come from harmonic analysis and known automorphy of symmetric power L-functions. This is discussed further in §4.3.

Nevertheless, it is possible to formulate a trace formula for  $L^2(\Gamma \backslash \mathfrak{H})$  using methods not unlike in the case of  $S$  a compact hyperbolic surface. The Selberg trace formula in this case connects spectral data of automorphic forms in  $L^2(\Gamma \backslash \mathfrak{H})$  to the geometry of the modular surface. This duality can be seen as complementary, in the cases where it is relevant to compare, to the Grothendieck-Lefschetz fixed-point theorem, which connects geometric fixed points to motivic cohomology. There is another trace formula, the Kuznetsov trace formula, that has been developed very explicitly on  $GL(2)$ , and, more recently, on  $GL(3)$ , and which is a favorite tool of analytic number theorists when dealing with Maaß forms and spectral automorphic L-functions, since it expresses a trace of spectral data in terms of familiar character sums and special functions.

This article seeks to expose and elucidate some of the spectral questions surrounding automorphic forms on  $GL(2)$ , and provide a conceptual introduction to the Selberg trace formula in this setting from a classical perspective. It should be viewed as complementary to a technical discussion of the trace formula, as a complete technical discussion is not the goal here. Instead, we choose to focus on (a) introducing the many arithmetic applications of the trace formula and spectral theory, and (b) providing a tour of key problems in this area.

In §2, we introduce the trace formula in its original setting of a Riemannian manifold. This is technically more straightforward, and prepares us to tackle the trace formula in the noncompact arithmetic case in §3. We then proceed to describe some of the properties and mysteries of the spectrum of the Laplacian on the modular surface in §4.

Before all of this, however, we pause to recall the Poisson summation formula and its proof, which provided the basic concept that Selberg initially sought to generalize. We ask that  $f \in C^2(\mathbf{R})$  and either  $f, f', f'' \in L^1(\mathbf{R})$ , or  $|f(x)| \ll 1/(1 + |x|)^{1+\delta}$ . We could instead require that  $f, \hat{f} \in L^1(\mathbf{R})$  and have bounded variation, where  $\hat{f}$  is the Fourier transform, which is given by

$$\hat{f}(\xi) = \int_{\mathbf{R}} f(x)e(-x\xi) dx.$$

(Throughout,  $e(z) = e^{2\pi iz}$ .) Then, the Poisson summation formula is

$$\sum_{m \in \mathbf{Z}} f(m) = \sum_{n \in \mathbf{Z}} \hat{f}(n).$$

The Poisson summation formula is of constant use in number theory, since it allows one to pass back and forth between the time and frequency domains, depending on which is easier to understand. It may be proved in a straightforward way by expanding the periodic function

$$g(y) = \sum_{m \in \mathbf{Z}} f(y + m)$$

as its Fourier series and setting  $y = 0$ . However, to draw a better analogy with the Selberg trace formula, we give the proof of Lapid [35].

We first form the convolution operator  $R(f)$  on  $L^2(\mathbf{T})$  ( $\mathbf{T} = \mathbf{Z} \backslash \mathbf{R}$ ). For a  $\phi \in L^2(\mathbf{T})$ ,

$$\begin{aligned} R(f)\phi(x) &= \int_{\mathbf{R}} f(y)\phi(x+y) \, dy = \int_{\mathbf{R}} f(y-x)\phi(x) \, dy \\ &= \int_{\mathbf{T}} \sum_{n \in \mathbf{Z}} f(y+n-x)\phi(y) \, dy = \int_{\mathbf{T}} K_f(x,y)\phi(y) \, dy, \end{aligned}$$

where  $K_f(x,y) = \sum_{n \in \mathbf{Z}} f(y+n-x) \in C^\infty(\mathbf{T}^2)$ .

We now compute the trace of  $R(f)$  in two different ways. First,

$$\text{tr } R(f) = \int_{\mathbf{T}} K_f(x,x) \, dx = \sum_{n \in \mathbf{Z}} f(n).$$

On the other hand,  $R(f)$  may be diagonalized using the orthonormal basis  $e_n(z) = e(nz)$ . It is immediately checked that  $R(f)e_n = \hat{f}(n)e_n$ . This shows that

$$\text{tr } R(f) = \sum_{n \in \mathbf{Z}} \hat{f}(n),$$

and we have arrived at the Poisson formula.

The basic idea of computing the trace of an appropriately-chosen integral operator in two different ways shall prove quite fruitful in the development of more sophisticated versions of the ‘trace formula.’ We conclude the introduction with what is probably the most famous application of the Poisson summation formula, namely, the proof of the functional equation of  $\zeta(s)$  using the modularity of Jacobi’s theta function

$$\theta(z) = \sum_{n \in \mathbf{Z}} e^{-\pi n^2 z}, \quad \Re z > 0.$$

If we choose  $f(x) = e^{-\pi x^2}$  in the Poisson summation formula, then since  $\hat{f}(x) = f(x)$ , we have that for  $y > 0$ ,

$$\sum_{n \in \mathbf{Z}} e^{-\pi n^2 x} = \frac{1}{\sqrt{y}} \sum_{n \in \mathbf{Z}} e^{-\pi n^2 / y}.$$

It is then straightforward to obtain the identity

$$\theta(1/z) = \sqrt{z} \theta(z).$$

Riemann expressed  $\xi(s)$  as the Mellin integral

$$2\pi^{-s/2} \Gamma(s/2) \zeta(s) =: 2\xi(s) = \int_0^\infty (\theta(u) - 1) u^{s/2} \frac{du}{u}, \quad \Re s > 0.$$

Breaking up the integral and changing variables, we arrive at

$$\xi(s) + \frac{1}{s} + \frac{1}{1-s} = \frac{1}{2} \int_1^\infty (\theta(u) - 1) (u^{s/2} + u^{(1-s)/2}) \frac{du}{u}.$$

Since the right-hand side is invariant under  $s \leftrightarrow 1 - s$ , so, too, is the left-hand side. Both are analytic by the exponential decay of  $\theta(u)$  (c.f. [14] for more details). This concludes our dalliance with theta functions, and our review of Poisson summation.

## 2. THE TRACE FORMULA: COMPACT CASE

In his epochal 1956 paper [56], Selberg announced a ‘general relation which can be considered as a generalization of the so-called Poisson summation formula.’ This ‘general relation,’ which he referred to as a the ‘trace formula,’ began as a way to study the spectrum of differential operators of finite order on a Riemannian space  $S$  that are invariant by a locally compact group of isometries of  $S$ . When  $S = \Sigma_g$  is a compact hyperbolic surface of genus  $\geq 2$ ,  $\Sigma_g$  may be formed as a quotient  $\pi_1(\Sigma_g)\backslash\mathfrak{H}$ , where  $\pi_1(\Sigma_g) \subset \mathrm{PSL}(2, \mathbf{R})$ , the fundamental group of  $\Sigma_g$  (and group of deck transformations with respect to the universal cover  $\mathfrak{H}$ ) is a strictly hyperbolic Fuchsian group. For ease of exposition, and to conform with [22, Chapter 1], whose techniques we refer to throughout, we restrict to this case for the remainder of this section.

We recall that  $\mathfrak{H}$  is a Riemannian manifold with the Poincaré metric  $ds^2 = \frac{dx^2 + dy^2}{y^2}$  and volume form  $d\mu(z) = \frac{dx \, dy}{y^2}$ . The Poincaré metric has constant negative Gaussian curvature  $K = -1$ . The Laplacian on  $\mathfrak{H}$  is the familiar operator

$$\Delta = y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

It is the unique fundamental differential operator invariant under the action of  $\mathrm{SL}(2, \mathbf{R})$  (i.e. all others are polynomials in  $\Delta$ ). Since it is invariant by  $\mathrm{SL}(2, \mathbf{R})$ , and hence by the Fuchsian group  $\pi_1(\Sigma_g)$ ,  $\Delta$  descends to the operator  $\mathcal{D}$  on a Riemann surface  $\Sigma_g$  via the projection  $\mathfrak{H} \rightarrow \Sigma_g$ . Explicitly, if  $\xi$  and  $\eta$  are local coordinates on  $\Sigma_g$  such that  $ds^2 = a \, d\xi^2 + 2b \, d\xi \, d\eta + c \, d\eta^2$ , then  $\mathcal{D}$  is given explicitly by

$$(2.1) \quad \mathcal{D} = \frac{1}{ac - b^2} \left[ \frac{\partial}{\partial \xi} \left( \frac{c\partial/\partial\xi - b\partial/\partial\eta}{\sqrt{ac - b^2}} \right) + \frac{\partial}{\partial \eta} \left( \frac{a\partial/\partial\eta - b\partial/\partial\xi}{\sqrt{ac - b^2}} \right) \right].$$

Since  $\Sigma_g$  is compact,

(a) the spectrum of  $\mathcal{D}$  is discrete:

$$0 = \mu_0 < \mu_1 \leq \mu_2 \leq \dots, \quad \lim_{n \rightarrow \infty} \mu_n = \infty;$$

(b) corresponding to the eigenvalues  $\mu_n$ , we can construct an orthonormal basis  $\{\phi_n\}$  for  $L^2(\Sigma_g)$  satisfying  $\mathcal{D}\phi_n + \mu_n\phi_n = 0$ ;

(c) the normalized eigenfunctions  $\phi_n$  are real-valued;

(d) we have a version of Bessel’s inequality

$$\sum_{k=1}^{\infty} \frac{1}{\mu_k^2} < \infty.$$

It follows from essentially the classical Hilbert-Schmidt theorem that the Fourier series  $\sum_{n=0}^{\infty} c_n \phi_n$  of any  $C^2(\Sigma_g)$  function converges uniformly and absolutely; c.f. [15, 1, p.383] and [27, 1, p.234–235]. We renote as  $\lambda_n := -\mu_n$  to conform to Selberg's [56] notation. Then

$$\phi_n \in C^\infty(\pi_1(\Sigma_g) \backslash \mathfrak{H}) \quad \text{and} \quad \mathcal{D}\phi_n = \lambda_n \phi_n.$$

As a consequence of the Stone-Weierstrass theorem, as in [26, p.151] we have the spectral decomposition

$$(2.2) \quad L^2(\pi_1(\Sigma_g) \backslash \mathfrak{H}) = \bigoplus_{n=0}^{\infty} [\phi_n].$$

It was Selberg's crucial insight that the spectral theory of  $\mathcal{D}$ , the Laplacian on  $\Sigma_g$ , can be formulated in terms of integral operators of the form

$$(2.3) \quad R(f)(z) := \int_{\Sigma_g} k(z, w) f(w) \, d\mu(w),$$

where  $d\mu(w)$  denotes the invariant element of volume derived from the metric, and the kernel is necessarily  $\pi_1(\Sigma_g)$ -invariant; i.e.

$$\begin{aligned} k(z, w) &= k(w, z), \quad \text{and} \\ k(mz, mw) &= k(z, w) \quad \text{for } w, z \in \Sigma_g, m \in \pi_1(\Sigma_g). \end{aligned}$$

Such a  $k$  is called a 'point-pair invariant.' The purpose of the rest of this section is to make explicit the way in which the operator  $R(f)$  provides a window into the spectrum of the Laplacian on a compact hyperbolic surface. We have selected this notation for  $R(f)$  to be in close analogy to the example of Poisson summation given in the introduction. Proofs may be found in [22, Chapter 1].

We begin by recalling that  $\Sigma_g$  admits a Fuchsian model as a quotient of  $\mathfrak{H}$  by  $\pi_1(\Sigma_g)$ , a Fuchsian group (discrete subgroup of  $\text{PSL}(2, \mathbf{R})$ ). The quotient space  $\pi_1(\Sigma_g) \backslash \mathfrak{H}$  has a polygon  $F$  as a representative element that is known as the standard fundamental polygon with  $4g$  sides, where  $g$  is the genus of  $\Sigma_g$ . Additionally,

$$\text{area}(\Sigma_g) = \mu(F) = 4\pi(g - 1),$$

with  $\mu$  as above.

Next, given a  $\Phi \in C_{00}(\mathbf{R})$ , there is an associated point-pair invariant  $k(z, w)$  given by

$$(2.4) \quad k(z, w) = \Phi \left( \frac{|z - w|^2}{\Im z \Im w} \right).$$

**Proposition 2.1** ([22, Ch. 1, Prop. 3.1]). *Let  $f$  be any eigenfunction of  $\Delta$  on  $\mathfrak{H}$  with eigenvalue  $\lambda$ . Then,*

$$\int_{\mathfrak{H}} k(z, w) f(w) \, d\mu(w) = \Lambda(\lambda) f(z),$$

where  $\Lambda(\lambda)$  depends solely on  $\lambda$  and the test function  $\Phi$ . ■

If we form the automorphic kernel function

$$K(z, w) := \sum_{\sigma \in \pi_1(\Sigma_g)} k(\sigma z, w), \quad z, w \in \mathfrak{H},$$

then, provided, as always,  $\Phi$  is a real test function with complex support,  $K(z, w)$  is actually a sum of uniformly bounded length,  $K(z, w) = K(w, z)$ , and  $K(z, w)$  is  $\pi_1(\Sigma_g) \times \pi_1(\Sigma_g)$ -invariant.

Returning now to the integral operator  $R(f)$  as in (2.3), and with  $\Phi \in C_{00}(\mathbf{R})$  as usual, then it can be shown that  $R(f)$  is a bounded linear operator on  $L^2(\pi_1(\Sigma_g) \backslash \mathfrak{H})$  and may be expressed as

$$R(f)(z) = \int_{\mathbb{F}} K(z, w) f(w) \, d\mu(w) \quad \text{for any } f \in L^2(\pi_1(\Sigma_g) \backslash \mathfrak{H}).$$

It follows (see [70, p.277]) that  $R(f)$  is an operator of Hilbert-Schmidt type on  $L^2(\pi_1(\Sigma_g) \backslash \mathfrak{H})$ . This allows us to state a few facts about  $R(f)$  that justify our choice of  $R(f)$  and make the notion of a ‘trace formula’ both viable and deeply significant.

**Proposition 2.2** ([22, Ch. 1, Prop. 3.8]). *Let  $\Phi \in C_{00}^2(\mathbf{R})$ . Then*

- (a)  $K(z, w) = \sum_{n=0}^{\infty} \Lambda(\lambda_n) \phi_n(z) \phi_n(w)$ , with uniform absolute convergence on  $\mathfrak{H} \times \mathfrak{H}$ .
- (b)  $\sum_{n=0}^{\infty} |\Lambda(\lambda_n)| \neq \infty$ .
- (c)  $\int_{\mathbb{F}} K(z, z) \, d\mu(z) = \sum_{n=0}^{\infty} \Lambda(\lambda_n)$ .

*Proof.* In view of the spectral decomposition (2.2), a consequence of the Fourier expansion of  $C^2$  functions on  $\Sigma_g \times \Sigma_g$  and the identity

$$\int_{\mathbb{F}} K(z, w) \phi_n(z) \, d\mu(z) = \int_{\Sigma_g} k(z, w) \phi_n(z) \, d\mu(z) = \Lambda(\lambda_n) \phi_n(w). \quad \blacksquare$$

The significance of Proposition 2.2 is immediately illustrated when considered alongside the spectral decomposition (2.2), and Proposition 2.1, which allow us to conclude that

$$\text{tr } R = \sum_{n=0}^{\infty} \Lambda(\lambda_n) = \int_{\mathbb{F}} K(z, z) \, d\mu(z),$$

where  $K(z, z) = \sum_{\sigma \in \pi_1(\Sigma_g)} k(\sigma z, z)$ , as above. The Selberg trace formula arises from a term-by-term expansion of the last integral. This expansion is somewhat technical, and full details may be found in [22, §1.4–1.6]. We limit our concern to providing some flavor for the expansion and characterizing its general shape without getting bogged down in technicalities.

We notate conjugacy classes in  $\pi_1(\Sigma_g)$  as  $[\sigma]$ . We observe that for a fixed  $\sigma \in \pi_1(\Sigma_g)$ ,  $g^{-1}\sigma g = h^{-1}\sigma h \Leftrightarrow hg^{-1}\sigma gh^{-1} = \sigma \Leftrightarrow gh^{-1} \in Z(\sigma) \Leftrightarrow g \in Z(\sigma)h$ , where  $Z(\sigma)$  is the



centralizer of  $\sigma \in \pi_1(\Sigma_g)$ . We may then write

$$\begin{aligned}
\mathrm{tr} R &= \sum_{n=0}^{\infty} \Lambda(\lambda_n) = \int_{\mathbb{F}} K(z, z) \, d\mu(z) \\
&= \int_{\mathbb{F}} \left( \sum_{\sigma \in \pi_1(\Sigma_g)} k(\sigma z, z) \right) d\mu(z) \\
&= \sum_{\sigma \in \pi_1(\Sigma_g)} \int_{\mathbb{F}} k(\sigma z, z) \, d\mu(z) \\
&= \sum_{\substack{[\sigma] \\ \text{distinct}}} \sum_{\mathbb{R} \in [\sigma]} \int_{\mathbb{F}} k(\mathbb{R}z, z) \, d\mu(z) \\
&= \sum_{\substack{[\sigma] \\ \text{distinct}}} \sum_{\rho \in Z(\sigma) \backslash \pi_1(\Sigma_g)} \int_{\mathbb{F}} k(\rho^{-1} \sigma \rho z, z) \, d\mu(z) \\
&= \sum_{[\sigma]} \sum_{\rho \in Z(\sigma) \backslash \pi_1(\Sigma_g)} \int_{\mathbb{F}} k(\sigma \rho z, \rho z) \, d\mu(z) \\
&= \sum_{[\sigma]} \sum_{\rho \in Z(\sigma) \backslash \pi_1(\Sigma_g)} \int_{\rho(\mathbb{F})} k(\sigma \xi, \xi) \, d\mu(\xi).
\end{aligned}$$

It can be shown that

$$\bigcup_{\rho \in Z(\sigma) \backslash \pi_1(\Sigma_g)} \rho(\mathbb{F})$$

is a fundamental region for the Fuchsian group  $Z(\sigma)$ , and that the integral of  $k(\sigma z, z)$  over any reasonable fundamental region for  $Z(\sigma)$  is independent of the choice of fundamental region. We are then justified in letting  $\mathfrak{F}[Z_{\Gamma}(\sigma)]$  denote any fundamental region for  $Z(\sigma)$ , and writing

$$\mathrm{tr} R = \sum_{\substack{[\sigma] \\ \text{distinct}}} \int_{\mathfrak{F}[Z_{\Gamma}(\sigma)]} k(\sigma z, z) \, d\mu(z).$$

This sum is absolutely convergent, since, as remarked above,  $K(z, w)$  is actually a sum of uniformly bounded length. It becomes clear that, to compute the other side of the trace formula, we must select a fundamental region  $\mathfrak{F}[Z_{\Gamma}(\sigma)]$  in such a way that we might be able to execute the resulting computations effectively.

For an arbitrary element  $\sigma \in \pi_1(\Sigma_g) \subset \mathrm{PSL}(2, \mathbf{R})$ ,  $\sigma$  is either hyperbolic or  $\sigma = I$ . This illustrates how the case under consideration is almost a ‘toy case;’ in any event, it is highly non-generic, in the sense that in general, there will be a large variety of other kinds of contributions to these orbital integrals than just the ones we will see here. Also, since we are working on  $\mathfrak{H}$ , there is only one fundamental invariant operator, a further simplification.

This is perhaps why Dennis Hejhal said ‘Of course, we won’t really understand the trace formula until it is written down for  $\mathrm{SL}(4, \mathbf{Z})$ ’

If  $\sigma$  is hyperbolic, then  $Z(\sigma)$  is a cyclic subgroup of  $\pi_1(\Sigma_g)$  with a generator  $\sigma_0$  that is uniquely determined by  $\sigma$ . Eventually putting  $z = \eta w$  with  $\eta \in \mathrm{PSL}(2, \mathbf{R})$ , we may write

$$\begin{aligned} \mathcal{E} &:= \int_{\mathfrak{F}[Z_\Gamma(\sigma)]} k(\sigma z, z) \, d\mu(z) = \int_{\mathfrak{F}[Z_\Gamma(\sigma)]} k(\sigma z, z) \, d\mu(z) \\ &= \int_{\eta^{-1}\mathfrak{F}[Z_\Gamma(\sigma)]} k(\sigma\eta w, \eta w) \, d\mu(w) \\ &= \int_{\mathfrak{F}[Z_\Gamma(\eta^{-1}\sigma\eta)]} k(\eta^{-1}\sigma\eta w, w) \, d\mu(w). \end{aligned}$$

We may choose  $\eta \in \mathrm{PSL}(2, \mathbf{R})$  such that

$$\eta^{-1}\sigma_0\eta(w) = N(\sigma_0)w, \quad 1 < N(\sigma_0) < \infty.$$

Then

$$\eta^{-1}\sigma\eta(w) = N(\sigma)w,$$

where  $N(\sigma)$  is the multiplier or characteristic constant of the transformation  $\sigma$ . It can be shown that a fundamental region for  $\eta^{-1}\sigma_0\eta$  is given explicitly by

$$\mathfrak{F}[Z_\Gamma(\eta^{-1}\sigma_0\eta)] = \{1 \leq \Im w < N(\sigma_0)\}.$$

This choice of fundamental region allows us to compute this orbital integral fully explicitly, and we obtain

**Proposition 2.3** ([22, Ch. 1, Prop. 6.3]). *For hyperbolic  $\sigma \in \pi_1(\Sigma_g)$ , and denoting the generator of the cyclic group  $Z(\Gamma)$  by  $\sigma_0$ , we have*

$$\int_{\mathfrak{F}[Z_\Gamma(\sigma)]} k(\sigma z, z) \, d\mu(z) = \frac{\log N(\sigma_0)}{N(\sigma)^{1/2} - N(\sigma)^{-1/2}} g(\log N(\sigma)),$$

where

$$g(u) := \int_x^\infty \frac{\Phi(t)}{\sqrt{t-x}} \, dt \quad \text{for } x = e^u + e^{-u} - 2 \text{ and } u \in \mathbf{R}. \quad \blacksquare$$

The other case is when  $\sigma = \mathbf{I}$ . In this case,  $\mathfrak{F}[Z_\Gamma(\mathbf{I})] = \mathfrak{F}[Z_\Gamma(\pi_1(\Sigma_g))] = \mathbf{F}$ , and

$$\int_{\mathbf{F}} k(z, z) \, d\mu(z) = \Phi(0)\mu(\mathbf{F}).$$

It is quite technical to get ahold of  $\Phi(0)$ , but, with perseverance, one finds that

**Proposition 2.4** ([22, Ch. 1, Prop. 6.4]). *If  $\Phi \in C_{00}^2(\mathbf{R})$ ,*

$$\Phi(0) = \frac{1}{4\pi} \int_{-\infty}^{\infty} r \left( \int_{-\infty}^{\infty} g(u) e^{iru} du \right) \tanh(\pi r) dr. \quad \blacksquare$$

This is enough to arrive at a preliminary version of the Selberg trace formula for a compact hyperbolic surface  $\Sigma_g$ . However, it requires some more work to arrive in the final form. Namely, to arrive at Selberg's version of the trace formula [56, p.74] for  $\Sigma_g$ , we need to compute the abscissa of convergence of a particular spectral sum; i.e. we would need to show that

$$\inf \left\{ \nu \geq 1 : \sum_{n=1}^{\infty} |\lambda_n|^{-\nu} < \infty \right\} = 1.$$

This, plus an approximation argument, yields the final form of the Selberg trace formula, which is considerably more flexible than the one obtained so far. In particular, we can replace the as-of-yet-implicitly-defined function  $\Lambda$  with a class of even test functions satisfying the following standard conditions:

- (1)  $h(r)$  is analytic on  $|\Im r| \leq \frac{1}{2} + \delta$ ,
- (2)  $h(-r) = h(r)$ , and
- (3)  $|h(r)| \ll (1 + |\Re r|)^{-2-\delta}$  for some  $\delta > 0$ .

We recall some notation and state the final version of the trace formula straightaway. The  $L^2(\pi_1(\Sigma_g) \backslash \mathfrak{H})$  spectrum of  $\Delta$  is  $\{\lambda_n\}$ , where

$$0 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots, \quad \lim_{n \rightarrow \infty} \lambda_n = -\infty.$$

**Theorem 2.5** (The Selberg trace formula for a compact hyperbolic surface, [22, Ch. 1, Theorem 7.5]). *Let  $\Sigma_g$  be a compact hyperbolic surface, let  $\Phi \in C_{00}^2(\mathbf{R})$ , and suppose that  $h(r)$  satisfies the above assumptions. Then, writing  $\lambda_n = \frac{1}{2} + ir_n$ , we have*

$$\begin{aligned} \sum_{n=0}^{\infty} h(r_n) &= \frac{\mu(F)}{4\pi} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr \\ &+ \sum_{[\sigma]} \frac{\log N(\sigma_0)}{N(\sigma)^{1/2} - N(\sigma)^{-1/2}} g(\log N(\sigma)), \end{aligned}$$

where

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) e^{-iru} dr, \quad u \in \mathbf{R}. \quad \blacksquare$$

The final form of the trace formula given in Theorem 2.5 is tantalizing when compared to the explicit formula attached to the zeta function or any L-function. If  $\gamma \in \mathbf{C}$  is an ordinate

of a nontrivial zero of  $\zeta(s)$ ; i.e.  $\zeta(\rho) = 0$  and  $\rho = \frac{1}{2} + i\gamma$ , then the explicit formula takes the form of the following sum over ordinates of critical zeros for some test function  $h$ .

$$\begin{aligned} \sum_{\gamma} h(\gamma) &= h\left(\frac{1}{2}\right) - g(0) \log \pi + \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \frac{\Gamma'}{\Gamma} \left(\frac{1}{4} + \frac{1}{2}ir\right) dr \\ &\quad - 2 \sum_{n=1}^{\infty} \frac{\Lambda(n)}{\sqrt{n}} g \log n, \end{aligned}$$

where  $\Lambda$  is von Mangoldt's function and  $g$  is related to  $h$  as above. The Riemann hypothesis is equivalent to the positivity of the right hand side of the explicit formula for convolution-type functions  $g(u)$  of the form

$$g(u) = \int_{-\infty}^{\infty} g_0(u+t) \overline{g_0(t)} dt.$$

The hope, then, would be to gain some insight into  $\zeta(s)$  and the Riemann hypothesis by studying trace formulæ attached to arithmetic groups  $\Gamma$  such as the arithmetic lattice  $\mathrm{PSL}(2, \mathbf{Z})$ . This is sometimes referred to as 'Hilbert and Pólya's dream;' in any case, this resemblance motivates the study in the next section.

### 3. THE TRACE FORMULA: ARITHMETIC CASE

We now arrive at the arithmetic case that is the primary concern of our discussion; namely, the noncompact manifold  $\Gamma \backslash \mathfrak{H}$ , where  $\Gamma = \mathrm{PSL}(2, \mathbf{Z})$  is the full modular group. Arthur [3, p. 2–3] presents eloquent motivation for the general study of Laplace-Beltrami operators attached to reductive algebraic groups. He explains that the eigenforms of these differential operators, automorphic forms, are expected to 'characterize some of the deepest objects of arithmetic.' The case of  $\Gamma \backslash \mathfrak{H}$  is the most classical, and home to so much arithmetic. As before, we will consider this space in as elementary and explicit language as possible, in the spirit of Hejhal, whose development we shall follow. In [23], Hejhal formulates three versions of the Selberg trace formula for noncompact quotients of the half-plane by Fuchsian groups whose fundamental regions have finite non-Euclidean area. Recall that a Fuchsian group acts on  $\mathfrak{H}$  by fractional linear transformations. Among these examples, we have the case of the full modular group, and congruence subgroups such as the principal congruence subgroup of level  $N$

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, d \equiv \pm 1 \pmod{N} \text{ and } b, c \equiv 0 \pmod{N} \right\},$$

the congruence subgroup

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : c \equiv 0 \pmod{N} \right\},$$

and the intermediate congruence subgroup

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a \equiv d \equiv 1 \pmod{N} \text{ and } c \equiv 0 \pmod{N} \right\}.$$

To these congruence subgroups there are associated modular curves  $X(N) = \Gamma(N) \backslash \mathfrak{H}$ ,  $X_0(N) = \Gamma_0(N) \backslash \mathfrak{H}$ , and  $X_1(N) = \Gamma_1(N) \backslash \mathfrak{H}$ . There is a formula for the number of cusps of various modular curves; for example,

$$|\{\text{cusps of } \Gamma_0(N)\}| = \sum_{d|N} \phi(\gcd(d, N/d)),$$

where  $\phi$  is Euler's totient function.

Let  $G$  denote any Fuchsian group whose fundamental region  $F$  has finite (non-Euclidean) area. We take a moment to review some vocabulary. A matrix  $A \in \text{SL}(2, \mathbf{C})$  is said to be

- (i) *loxodromic* when  $\text{tr } A \notin [-2, 2]$ ,
- (ii) *hyperbolic* when  $\text{tr}^2 A \in (4, \infty)$ ,
- (iii) *elliptic* when  $\text{tr } A \in (-2, 2)$ , and
- (iv) *parabolic* when  $\text{tr } A = \pm 2$  and  $\text{tr } A \notin \{1, -1\}$ .

When developing a theory of group actions on hyperbolic space, we think of matrices in  $\text{PSL}(2, \mathbf{R})$  as fractional linear transformations of the half-plane  $\mathfrak{H}$ . We recall the notion of normal form for a hyperbolic transformation  $L \in \text{PSL}(2, \mathbf{R})$ , which has two fixed points, say,  $\xi$  and  $\eta$ . Setting

$$g(z) := \frac{z - \xi}{z - \eta},$$

then the transformation  $gLg^{-1}$  fixes 0 and  $\infty$  and is thus a dilation  $\kappa z$ . This complex number  $\kappa$  is called the multiplier of  $L$ , and satisfies

$$\lambda^{1/2} + \lambda^{-1/2} = |\text{tr } L|.$$

To be consistent with [22, 23], we use the notation  $N(L) = \lambda$  for the multiplier of a hyperbolic transformation  $L$ .

Let  $\chi$  be an  $r \times r$  unitary representation of  $G$ . We then define the  $L^2$  space

$$L^2(G \backslash \mathfrak{H}, \chi) = \left\{ f : f \text{ is measurable on } \mathfrak{H}, \right. \\ \left. \begin{aligned} f(\sigma x) &= \chi(\sigma) f(x) \text{ for } \sigma \in G, \\ \int_{\mathbb{F}} f(z) {}^t \overline{f(z)} \, d\mu(z) &< \infty \end{aligned} \right\}.$$

(Here, we regard the functions  $f$  as column vectors.) It follows from the assumption  $\text{area}(G \backslash \mathfrak{H}) < \infty$  that  $G$  is a finitely-generated Fuchsian group of the first kind. We say that a point  $a \in \mathbb{H}$  is an *elliptic fixpoint* for  $G$  if there is an elliptic element  $\sigma \in G$  such that  $\sigma$  stabilizes  $a$ . The attached isotropy subgroup  $G_a$  then reduces to a finite cyclic group; i.e.

$$G_a = \{\sigma \in G : \sigma(a) = a\}$$

is generated by an elliptic element of  $G$  of finite order.

We call a point  $b \in \mathbf{R} \cup \{\infty\}$  a *cuspid* if there is a parabolic transformation  $\eta \in G$  such that  $\eta(b) = b$ . In this case, the isotropy subgroup  $G_b$  is infinite cyclic with parabolic generator. Two elliptic fixpoints or cusps are deemed equivalent when one can be taken to the other by an element of  $G$ .

The above assumption on the finiteness of the area of a fundamental region for  $G$  implies that the number of inequivalent elliptic fixpoints and the number of inequivalent cusps are both finite. We denote these integers  $s$  and  $\kappa$ , respectively.

The analogue of the hyperbolic Laplacian for automorphic forms of weight  $k$  is

$$\Delta_k = y^2 \left( \frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y} \right) - iky \frac{\partial}{\partial x};$$

see Maaß [44]. Among other things, recall that an automorphic form of weight  $k$  on, say,  $\Gamma \backslash \mathfrak{H}$ , is an eigenfunction of  $\Delta_k$ . We will be somewhat relaxed in this exposition about our notation for the Laplace-Beltrami operator on  $\Gamma \backslash \mathfrak{H}$ . It descends as in the compact case (2.1), but we will just persist in notating it as  $\Delta$ , for sake of notational simplicity.

On the way towards developing the final version of Selberg's classical trace formula, Hejhal [23] formulates progressively more general versions. Explicitly, with  $r$  the dimension of the unitary representation  $\chi$  of  $G$ ,  $s$  the number of inequivalent elliptic fixpoints,  $\kappa$  the number of inequivalent cusps, and  $k$  the weight, which specifies the differential operator  $\Delta_k$ , Hejhal's three formulations of the trace formula on  $L^2(G \backslash \mathfrak{H}, \chi)$  for  $G$  a Fuchsian group of the first kind with  $\text{area}(G \backslash \mathfrak{H}) < \infty$  are given by the letters A, B, and C corresponding to the following choice of parameters.

	$\kappa$	$s$	$r$	$k$
A	1	finite	1	0
B	finite	finite	finite	0
C	finite	finite	finite	arbitrary real

Table 1: Versions of the Selberg trace formula.

We note that the modular surface  $\text{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$  has a cusp at  $\infty$  and elliptic fixpoints  $\{1\}$  and  $\{0\}$ ; hence  $\kappa = 1$  and  $s = 2$ . Version A of the trace formula contains the central ideas in the simplest form, and since it addresses the case of the modular surface, we will present only this case. The technicalities required to arrive at versions B and C are substantial, and may be found in [23, Ch. 8, 9], respectively. The case of a Riemann surface, discussed in §2, corresponds to  $\kappa = 0$ ,  $s = 0$ ,  $\chi$  trivial, and  $k = 0$ .

**3.1. Introducing the Selberg Trace Formula on  $\Gamma \backslash \mathfrak{H}$ .** In this section,  $\Gamma$  denotes any Fuchsian group whose fundamental region has finite area and one cusp, but we might as well just think about it as  $\text{PSL}(2, \mathbf{Z})$ . Without loss of generality, we can place the single cusp at  $\infty$ ; then, from the above, the stabilizer  $\Gamma_\infty$  of the cusp is infinite cyclic. From the action of  $\Gamma$  by fractional linear transformations, we see that  $\Gamma_\infty$  is generated by the matrix  $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ .

The value that the character  $\chi$  assumes on the generator  $S$  changes the matter of formulating the Selberg trace formula for  $\Gamma \backslash \mathfrak{H}$  drastically. For, while the spectral theory of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  is not far away from the case of a compact surface when  $\chi(S) \neq 1$ , it is quite different when  $\chi(S) = 1$ , and the usual attempt to characterize the spectrum in terms of integral operators fails badly. It is illustrative to first discuss how we would like to formulate the spectral theory of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  in general; this approach actually works when  $\chi(S) \neq 1$ . The basic idea is to try to apply the resolvent formalism to the spectral theory of  $\Delta$  on  $\Gamma \backslash \mathfrak{H}$ . This can be understood as a classical exercise in applying Fredholm integral operator techniques to a problem in PDE.

As in the compact case, we would like to reformulate the spectral theory of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  in terms of integral operators. If this is possible, then we should be able to use Green's functions as resolvents to characterize the spectrum in a standard way. More concretely, here are the steps we would like to take to understand the spectral theory of  $L^2(\Gamma \backslash \mathfrak{H})$ .

- (a) Construct a Green's function for the PDE  $\Delta u + s(1-s)u = 0$  on  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  that will serve as the integral kernel of a resolvent for the Laplace-Beltrami operator on  $\Gamma \backslash \mathfrak{H}$ . In particular, the eigenfunctions of the operators arising from the Green's functions coincide with those of  $\Delta$ .
- (b) Show that the corresponding integral operators have a discrete countable spectrum.
- (c) Verify that the conditions of the Hilbert-Schmidt theorem are satisfied for the integral operators with Green's functions as kernels.
- (d) Conclude that the spectrum of  $\Delta$  on  $\Gamma \backslash \mathfrak{H}$  is discrete and spanned by a countable orthonormal basis of eigenfunctions.
- (e) Form a new automorphic kernel function  $K$  in analogy with § 2, chosen so that the integral of the trace, on the one hand, coincides with the trace of the spectrum of  $\Delta$  enveloped by a test function, while, on the other hand, can be expressed in terms of orbital integrals of a point-pair invariant that are tractable and can be computed fairly explicitly.
- (f) Explicitly compute the resulting orbital integrals.

Fortunately, this goes through in the case that  $\chi(S) \neq 1$ . It absolutely does not in the case that  $\chi(S) = 1$ , which is the case of most classical interest to number theorists. It is easy to see that it doesn't work out; harder to understand why. To see that it doesn't work out, assume all steps prior to (e), and, following [23, Ch. 6, §9], consider the standard point-pair invariant  $k(z, w)$  as in 2.4, where  $\Phi \in C_{00}(\mathbf{R}^+)$ ,  $\Phi(0) = 1$ , and  $\Phi(t) \geq 0$ . In obvious analogy with before, form the kernel function

$$K(z, w, \chi) := \sum_{\sigma \in \Gamma} \chi(\sigma) k(z, \sigma w).$$

The compact support of  $\Phi(t)$  ensures convergence of  $K(z, w, \chi)$ . Repeating to [23, Ch. 6, Prop. 6.6] shows that  $K(z, w, \chi) \equiv 0$  near  $z = i\infty \forall w \in \mathfrak{H}$ ; hence that  $K(z, w, \chi) \in L^2(\Gamma \backslash \mathfrak{H}, \chi)$ .

If the spectrum of  $L^2(z, w, \chi)$  were discrete and reasonable, we would then have an orthonormal decomposition

$$L^2(\Gamma \backslash \mathfrak{H}, \chi) = \bigoplus_{n=0}^{\infty} [\phi_n]$$

for eigenfunctions  $\phi_n$ . This spectral decomposition immediately implies the spectral expansion

$$K(z, w, \chi) = \sum_{n=0}^{\infty} b(r_n) \phi_n(z) \overline{\phi_n(w)} \quad \forall w \in \mathfrak{H}.$$

In analogy with § 2, we will most definitely need  $\sum |b(r_n)| < \infty$  to develop a sensible notion of trace. The hope would be that by taking  $\Phi$  sufficiently smooth, we would be able to attain this (necessary) condition, in which case we could make a statement like

$$\int_{\mathbb{F}} K(z, z, \chi) \, d\mu(z) = \sum_{n=0}^{\infty} b(r_n).$$

But, the integral over the fundamental region on the left hand side diverges. To see this, we find some  $\alpha$  such that  $\Phi(t) \geq \frac{1}{2}$  for  $t \in [0, \alpha]$ . Then,

$$K(z, z, \chi) = \sum_{\sigma \in [S]} \chi(\sigma) k(z, \sigma z) + \sum_{\sigma \notin [S]} \chi(\sigma) k(z, \sigma z) \quad \text{for } z \in \mathbb{F}.$$

Taking  $y = \Im z$  sufficiently large, and assuming that for  $y$  large,  $\mathbb{F}$  begins to look like the rectangular region  $[0, 1] \times [y, \infty]$ , it can be seen that due to the construction of  $k(w, z)$ , the sum over  $\sigma \notin [S]$  does not contribute due to the compact support of  $\Phi$ . We are left with the sum over the stabilizer of the cusp, which, for large values of  $y$ ,

$$K(z, z, \chi) = \sum_{n=-\infty}^{\infty} k(z, S^n z) = \sum_{n=-\infty}^{\infty} \Phi\left(\frac{n^2}{y^2}\right) \geq \lfloor y\sqrt{\alpha} \rfloor.$$

In particular,

$$\int_Y^{\infty} \int_0^1 K(z, z, \chi) \frac{dx \, dy}{y^2} = +\infty \quad \text{for some } Y.$$

The convergence issues with our naïve effort indicate that our assumptions of a discrete, well-behaved spectrum when  $\chi(S) = 1$  has gotten us in trouble. We will re-assume this matter in the next section, but the above provides evidence for a *continuous spectrum*. Characterizing this continuous spectrum, decomposing it into generators, and arriving at a spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H})$ , will be the primary objective of our efforts to address the case  $\chi(S) = 1$  in what follows. Let

$$(3.1) \quad \delta(\chi) = \begin{cases} 1 & \chi \equiv 1 \\ 0 & \chi \neq 1. \end{cases}$$



We will conclude by stating the spectral decomposition, due to Selberg.

$$(3.2) \quad L^2(\Gamma \backslash \mathfrak{H}, \chi) = \delta(\chi)\mathbf{C} \oplus L_{\text{cusp}}^2(\Gamma \backslash \mathfrak{H}) \oplus L_{\circ}^2(\Gamma \backslash \mathfrak{H}) \oplus L_{\text{cont}}^2(\Gamma \backslash \mathfrak{H}).$$

In order,  $\delta(\chi)\mathbf{C}$  are the constant functions, present only when  $\chi \equiv 1$ .  $L_{\text{cusp}}^2(\Gamma \backslash \mathfrak{H})$  is the space spanned by cusp forms, as in the compact case or the case when  $\chi(S) \neq 1$ .  $L_{\circ}^2(\Gamma \backslash \mathfrak{H})$  is spanned by residues of  $\text{GL}(2)$  Eisenstein series. It has another characterization that we will be useful later. Finally,  $L_{\text{cont}}^2(\Gamma \backslash \mathfrak{H})$  is the continuous spectrum. We will give a more involved description in the next section.

As a final remark, we return to the case that is foremost in our minds, that is, the case of  $\Gamma = \text{SL}(2, \mathbf{Z})$  with trivial character. Then the spectral decomposition is simply

$$L^2(\text{SL}(2, \mathbf{Z}) \backslash \mathfrak{H}) = \mathbf{C} \oplus L_{\text{cusp}}^2(\text{SL}(2, \mathbf{Z}) \backslash \mathfrak{H}) \oplus L_{\text{cont}}^2(\text{SL}(2, \mathbf{Z}) \backslash \mathfrak{H}).$$

We note that the contribution to the discrete spectrum from residues of Eisenstein series is not present. This is because the only pole of the Eisenstein series in  $\Re s \geq \frac{1}{2}$  is at  $s = 1$ , and the residue there is a constant function. Therefore, it doesn't contribute anything new.

**3.2. The Spectral Decomposition of  $\Gamma \backslash \mathfrak{H}$ .** In this section, we present the key techniques used to decompose the (discrete) spectrum in the case  $\chi \neq 1$ , and we elaborate on how to arrive at the decomposition (3.2). The references throughout are [23, Ch. 6–7] and [15, Ch. 10].

First, we address the easier case of  $\chi(S) \neq 1$ . The goal is to use Hilbert-Schmidt to characterize the spectrum of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$ . To do this, we must transfer the spectral theory of  $\Delta$  to that of some easier-to-understand integral operators. That is, we would like to introduce a Green's function for the PDE  $\Delta u + s(1-s)u = 0$  on  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$ .<sup>1</sup> There are a raft of conditions we would like such a function  $k_s(z; z_0)$  to satisfy. Referring to [23, Ch. 6, §6], we insist that

- (a)  $k_s(z; z_0)$  be a point-pair invariant;
- (b)  $k_s(z; z_0)$  should be  $\mathbf{C}^\infty$  for  $z \in \mathfrak{H} - \{z_0\}$ ;
- (c)  $k_s(z; z_0) = \frac{1}{2\pi} \log |z - z_0| + \mathcal{O}(1)$  near  $z = z_0$ ;
- (d)  $\Delta k_s(z; z_0) + s(1-s)k_s(z; z_0) = 0$  for  $z \in \mathfrak{H} - \{z_0\}$ ;
- (e)  $k_s(z; z_0)$  should be as small as possible when  $z \rightarrow \partial \mathfrak{H}$ .

The reason for the last condition is that  $\Delta$  is singular along  $\partial \mathfrak{H}$ , which will introduce difficulties. Using the theory of PDE and Euler's hypergeometric differential equation in particular, we can prove the following statement about the Fourier coefficients of an eigenfunction of  $\Delta$  on  $\mathfrak{H}$ .

**Proposition 3.1** ([23, Ch. 6, Lemma 4.2 & Prop 4.3]). *Suppose that  $z_0 \in \mathfrak{H}$ ,  $0 \leq r_1 < r_2 \leq 1$ ,  $s \in \mathbf{C} - \{0, -1, -2, -3, \dots\}$ . Assume  $f$  is  $\mathbf{C}^2$  and satisfies  $\Delta f + s(1-s)f = 0$  on the set*

$$N(z_0; r_1, r_2) = \left\{ z \in \mathfrak{H} : r_1 < |w| < r_2 \right\}, \quad \text{where } w = re^{i\theta} := \frac{z - z_0}{z - \bar{z}_0}.$$

<sup>1</sup>The introduction of the extra  $s(1-s)u$  term is a standard trick for attacking eigenvalues for which the Green's functions may not otherwise converge; c.f. [22, note 4, p.354]).

Define

$$f_n(\xi) = \xi^n (1 - \xi^2)^s \frac{F(s + n, s; 1 + n; \xi^2)}{\Gamma(1 + n)} \quad \text{for } |\xi| < 1, \text{ and}$$

$$g_n(\xi) = \xi^n (1 - \xi^2)^s \frac{F(s + n, s; 2s; 1 - \xi^2)}{\Gamma(2s)} \quad \text{for } \left\{ 0 < |\xi| < 1, |\operatorname{Arg} \xi| < \frac{\pi}{4} \right\},$$

where  $F(a, b; c; x)$  is the hypergeometric function, a solution of Euler's differential equation

$$x(1 - x)u'' + [c - (a + b + 1)x]u' - abu = 0.$$

Then

- (i)  $f(z) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}$  for  $r_1 < r < r_2$ ;
- (ii)  $c_n(r)$  satisfies the differential equation

$$(3.3) \quad u'' + \frac{1}{r}u' + \left( \frac{4s(1-s)}{(1-r^2)^2} - \frac{n^2}{r^2} \right) u = 0 \quad \text{on } r_1 < r < r_2,$$

while  $f_n(r)$  and  $g_n(r)$  satisfy (3.3) on  $0 < r < 1$ ;

- (iii)  $c_n(r) = A_n f_{|n|}(r) + B_n g_{|n|}(r)$  for appropriate constants  $A_n$  and  $B_n$ .

■

Using Proposition 3.1, we may conclude that  $k_s(z, z_0)$  must be a linear combination of  $f_0(r)$  and  $g_0(r)$ . To satisfy the growth condition (e) in the list of conditions above, and by examining the behavior of  $g_0(r)$  at  $r = 0$  using differential equations techniques and examining the behavior of the linearly independent solutions  $f_0$  and  $g_0$  to (3.3), we find that it behooves us to set

$$k_s(z, z_0) = -\frac{\Gamma(s)^2}{4\pi} g_0(r),$$

since for  $\Re s > \frac{1}{2}$ , the ODE (3.3) has a regular singular point at  $r = 1$ , and setting the coefficient of  $f_0$  in the linear combination constituting  $k_s(z; z_0)$  to zero gives us the smallest possible growth near  $\partial\mathfrak{H}$  (see [23, p.30] for more details). That is, we set

**Definition 3.2** ([23, Ch. 6, Def. 6.1]). For  $\Re s > \frac{1}{2}$ , let

$$k_s(z; w) = -\frac{\Gamma(s)^2}{4\pi\Gamma(2s)} \left( 1 - \left| \frac{z-w}{z-\bar{w}} \right|^2 \right)^s F \left( s, s; 2s; 1 - \left| \frac{z-w}{z-\bar{w}} \right|^2 \right).$$

With this definition,  $k_s(z; w)$  is the *free-space Green's function* for  $\Delta u + s(1-s)u = 0$ . Naturally, we would now like to 'project'  $k_s(z; w)$  down to a function on the coset space  $\Gamma \backslash \mathfrak{H}$ . To construct the Green's function on  $L^2(\Gamma \backslash \mathfrak{H})$ , we would like to do the most naïve thing, which is to define the Poincaré series

$$G_s(z; w; \chi) := \sum_{\sigma \in \Gamma} \chi(\sigma)^{-1} k_s(\sigma z; w).$$

Fortunately, this is not at all far off from what we end up doing. We would like to insist that  $G_s(z; w; \chi)$  be absolutely convergent, since a well-behaved definition of a Poincaré series on  $\Gamma \backslash \mathfrak{H}$  obviously shouldn't depend on the order of summation. We have the following convergence result.

**Proposition 3.3** ([23, Ch. 6, Prop. 6.2 & 6.3]). *Up to dropping a finite number of terms in the series,  $\sum_{\sigma \in \Gamma} \chi(\sigma)^{-1} k_s(\sigma z; w)$  converges uniformly and absolutely on compact subsets of  $\mathfrak{H} \times \mathfrak{H} \times \{s \in \mathbb{C} : \Re s > 1\}$ . On the other hand, the infinite series*

$$\sum_{\sigma \in \Gamma} |k_s(\sigma z; w)|$$

*is divergent whenever  $z \not\equiv w \pmod{\Gamma}$  and  $\Re s = 1$ . ■*

Motivated by Proposition 3.3, we proceed to define the automorphic Green's function on  $\Gamma \backslash \mathfrak{H}$  by

**Definition 3.4** (The automorphic Green's function [23, Ch. 6, Def. 6.4]).

$$G_s(z; w; \chi) = \sum_{\sigma \in \Gamma} \chi(\sigma^{-1}) k_s(\sigma z; w) \quad \text{for } (z, w) \in \mathfrak{H} \times \mathfrak{H} \text{ and } \Re s > 1.$$

We usually assume  $z \not\equiv w \pmod{\Gamma}$  for safety. Next, we summarize some nice properties of  $G_s(z; w; \chi)$ .

**Proposition 3.5** ([23, Ch. 6, Prop. 6.5]). *Assume that  $\Re s > 1$ . Then*

- (a)  $G_{\bar{s}}(z; w; \chi) = \overline{G_s(w; z; \chi)}$ , while  $G_s(z; w; \bar{\chi}) = G_s(w; z; \chi)$ ;
- (b)  $G_s(\sigma z; \eta w; \chi) = \chi(\sigma) G_s(z; w; \chi) \chi(\eta^{-1})$  for  $\sigma, \eta \in \Gamma$ ;
- (c)  $G_s(z; w; \chi)$  is  $C^\infty$  in both variables when  $z \not\equiv w \pmod{\Gamma}$ ,
- (d)  $G_s(z; z_0; \chi)$  satisfies  $\Delta u + s(1-s)u = 0$  on the set  $z \not\equiv z_0 \pmod{\Gamma}$ ,
- (e) when  $\Gamma_{z_0} = \mathbf{I}$ , we have

$$G_s(z; z_0; \chi) = \frac{1}{2\pi} \log |z - z_0| + O(1) \text{ near } z = z_0;$$

- (f) when  $\Gamma_{z_0}$  is infinite cyclic generated by an elliptic  $R \in \Gamma$  of order  $\nu$ , we have

$$G_s(z; z_0; \chi) = \begin{cases} 0 & \chi(R) \neq 1 \\ \frac{\nu}{2\pi} \log |z - z_0| + O(1) \text{ near } z = z_0 & \chi(R) = 1. \end{cases} \quad \blacksquare$$

The significant points of Proposition 3.5 is that  $G_s(z; z_0; \chi)$  has at most a logarithmic singularity near the diagonal, and is an eigenfunction of the Laplacian on  $\Gamma \backslash \mathfrak{H}$ . The point is that, not unlike in the case of a compact Riemann surface, the spectral theory of  $\Delta$  can be reformulated in terms of integral equations involving a kernel with at most a logarithmic singularity. The next proposition shows that  $G_s(z; w; \chi)$  must be as well-behaved near the cusp as we could want.

**Proposition 3.6** ([23, Ch. 6, Prop 6.6]). *Let  $E \times K$  be a compact subset of  $\mathfrak{H} \times \{\Re s > 1\}$ . Assume that  $0 \leq \Re z \leq 1$  and  $w \in E, s \in K$ . Then*

$$\sum_{\sigma \in \Gamma} |k_s(\sigma z; w)| \xrightarrow{z \rightarrow i\infty} 0 \quad \text{uniformly.} \quad \blacksquare$$

The next task is to obtain a Fourier theory for the automorphic Green's function. This is extremely technical, and relies heavily on the theory of Bessel functions, which can also be used to effect a kind of 'separation of variables' in the same way as was achieved in Proposition 3.1 using hypergeometric functions. Once this separation of variables is achieved, we can express a Fourier expansion for  $G_s(z; w; \chi)$  in terms of I- and K-Bessel functions. The full calculation is performed in [23, Ch. 9, §4, 6]. In keeping with the goal of emphasizing concept, we do not reproduce pages of integrals over cosets, and state the main result.

**Proposition 3.7** (Explicit expression for automorphic Green's functions, [23, p. 41–42]). *Let  $\alpha \in [0, 1)$  be defined implicitly by  $\chi(S) = e(\alpha)$ . Set  $\mu := \min_{n+\alpha \neq 0} |n + \alpha|$ . Then*

$$\begin{aligned} G_s(z; w; \chi) &= \delta_{0\alpha} F_0(w; s; \bar{\chi}) \frac{y^{1-s}}{1-2s} \\ &\quad - \sum_{n+\alpha \neq 0} \overline{F_n(w; \bar{s}; \chi)} y^{\frac{1}{2}} K_{s-\frac{1}{2}}(2\pi |n + \alpha| y) e((n + \alpha)z) \\ &= \delta_{0\alpha} F_0(w; s; \bar{\chi}) \frac{y^{1-s}}{1-2s} + O(e^{-2\pi\mu y}), \quad \text{where} \\ F_n(w; s; \chi) &= \sum_{W_0 \in \Gamma_\infty \backslash \Gamma} \chi(W_0^{-1}) \Im(W_0 w)^{\frac{1}{2}} I_{s-\frac{1}{2}}(2\pi |n + \alpha| \Im(W_0 w)) \\ &\quad \times e((n + \alpha)\Re(W_0 w)) \quad \text{when } n + \alpha \neq 0, \text{ and} \\ F_0(w; s; \chi) &= \sum_{W_0 \in \Gamma_\infty \backslash \Gamma} \chi(W_0^{-1}) \Im(W_0 w)^s \quad \text{when } n + \alpha = 0. \quad \blacksquare \end{aligned}$$

Our attention turns next to the functions  $F_n(z; s; \chi)$ . Let  $\mathcal{R} := \{s \in \mathbf{C} : \Re s > 1\}$ . Then  $F_n(z; s; \chi)$  is defined on  $\mathfrak{H} \times \mathcal{R}$  using Proposition 3.7. We summarize some facts about the behavior of these functions on compact subsets of  $\mathcal{R}$ .

**Proposition 3.8** ([23, Ch. 6, Prop. 8.1]). *Let  $K \subset \mathcal{R}$  be compact. Then*

- (a)  $F_n(z; s; \chi)$  is not affected by ambiguities in  $W_0$ ;
- (b)  $F_n(z; s; \chi)$  converges uniformly & absolutely on  $\mathfrak{H} \times K$  compacta;
- (c)  $F_n(\sigma z; s; \chi) = \chi(\sigma) F_n(z; s; \chi)$  for  $\sigma \in \Gamma$ ;
- (d)  $F_n(z; s; \chi)$  is a  $C^\infty$  solution of  $\Delta u + s(1-s)u = 0$  on  $\mathfrak{H}$ ;
- (e) the 'truncated' series

$$\tilde{F}_n(z; s; \chi) := F_n(z; s; \chi) - \left( \frac{y^{\frac{1}{2}} I_{s-\frac{1}{2}}(2\pi |n + \alpha| y) e((n + \alpha)z)}{y^s} \right)$$

tends uniformly to 0 whenever  $\Im z \rightarrow +\infty$  and  $s \in \mathbb{K}$ . ■

3.2.1.  $\chi(S) \neq 1$ . We are now in an optimal position to move forward with the spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H})$  in the case that  $\chi(S) \neq 1$ . We begin by noting that, crucial for an application of Hilbert-Schmidt, the automorphic Green's function is square-integrable. *Throughout this section,  $\chi(S) \neq 1$  is assumed.*

**Proposition 3.9** ([23, Ch. 6, §10]). *Let  $\mathbb{K} \subset \mathbb{R}$ . Then*

$$(3.4) \quad \int_{\mathbb{F}} |G_s(z; w; \chi)|^2 d\mu(z)$$

is uniformly bounded for  $s \in \mathbb{K}$  and  $w \in \mathfrak{H}$ . (The bound depends on  $\chi$ ,  $\mathbb{K}$ , and  $\Gamma$  only.) Additionally, (3.4) is continuous in  $w$ , and tends to 0 whenever  $\Im w \rightarrow \infty$ . ■

Next, we have the crucial proposition that eigenfunctions of the Fredholm integral operator

$$(3.5) \quad v \mapsto \int_{\mathbb{F}} G_s(z; w; \chi)v(w) d\mu(w)$$

on  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  are precisely the ones we want.

**Proposition 3.10** ([23, Ch. 6, Prop. 10.2]). *Assume that  $1 < a < \infty$ . Then*

$$v(z) = \lambda \int_{\mathbb{F}} G_s(z; w; \chi)v(w) d\mu(w) \quad \text{with } v \in L^2(\Gamma \backslash \mathfrak{H}, \chi)$$

if and only if

$$\Delta v + a(1 - a)v = \lambda v \quad \text{with } v \in C^\infty(\mathfrak{H}) \cap L^2(\Gamma \backslash \mathfrak{H}, \chi). \quad \blacksquare$$

There is only one more condition to obtain that the system of eigenfunctions for the integral operator (3.5) is complete in  $C^2(\mathfrak{H}) \cap L^2(\Gamma \backslash \mathfrak{H}, \chi)$ , and thereby arrive at the discrete spectral decomposition.

**Proposition 3.11** ([23, Ch. 6, Prop. 10.3]). *Suppose that  $f \in C^2(\mathfrak{H}) \cap L^2(\Gamma \backslash \mathfrak{H}, \chi)$  and  $\chi(S) \neq 1$ . Assume (for simplicity) that  $f(x + iy) \equiv 0$  near  $y = \infty$ . Then*

$$f(z) = \int_{\mathbb{F}} G_s(z; w; \chi)h(w) d\mu(w),$$

where  $h = \Delta f + a(1 - a)f$ , where  $1 < a < \infty$ . ■

As we will see, Proposition 3.11 allows us to apply Hilbert-Schmidt theorem, since the integral operator with kernel  $G_s(z; w; \chi)$  actually is a resolvent for the Laplacian in the above way. Proposition 3.10 allows to transfer the characterization we can obtain about our integral operator to a statement about  $\Delta$ . We now proceed to a discussion of the Hilbert-Schmidt theorem in a somewhat more basic and restricted setting, which nonetheless contains all the key ideas. We state here according to [15, Ch. 10, §4].

First, some conditions and notation. Suppose  $K(s, t)$  is a symmetric kernel. Suppose we are in the  $L^2$  space of a measure space  $(D, \Omega, \sigma)$ , where  $D$  is a (not-necessarily) compact domain. We use the notation  $K \circ f$  to denote the integral convolution

$$\int_X K(s, t)f(t) \, d\mu(t).$$

We assume that the kernel  $K$  has a weak singularity of the form

$$K(s, t) = \frac{\kappa(s, t)}{r^\ell} \quad 0 \leq \ell < \dim D,$$

where  $\kappa(s, t)$  is a bounded function, and  $r$  is the distance between  $s$  and  $t$ . We use the notation  $K \circ f$  for the operator

$$K \circ f = \int K(s, t)\phi(t) \, d\mu(t).$$

In this notation, (3.5) may be written as  $G_s(z; w; \chi) \circ v$ . Then we have the classical Hilbert-Schmidt theorem.

**Theorem 3.12** (Hilbert-Schmidt). *Any function  $f$  which can be expressed in the form*

$$f = K \circ g,$$

where  $g$  is some square integral function and  $K$  is a symmetric kernel satisfying the above conditions, may be expressed as an absolutely and uniformly convergent series

$$f(s) = \sum_{n=1}^{\infty} (\phi_n \circ f) \phi_n(s) = \sum_{n=1}^{\infty} \frac{(\phi_n \circ g)}{\lambda_n} \phi_n(s),$$

where the  $\phi_n$  are eigenfunctions of  $K$ . In other words, the eigensystem of  $K$  forms a Hilbert space basis for  $L^2(D)$ . ■

In some sense, we would like to apply the Hilbert-Schmidt theorem directly to our situation. Unfortunately, there are additional technicalities relating to the fact that the spectral theory is a bit different for modular functions on  $\Gamma \backslash \mathfrak{H}$  twisted by the unitary representation  $\chi$ . Fortunately, Hejhal [22, Ch. 3] formulates the necessary modifications to [15, Ch. 10] in full detail, and we allow ourselves to conclude that

**Proposition 3.13** (Spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  when  $\chi(S) \neq 1$ , [23, p. 99]). *The eigensystem of  $\Delta$  is orthonormal and complete in  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$ ; i.e.*

$$L^2(\Gamma \backslash \mathfrak{H}, \chi) = \bigoplus_{n=1}^{\infty} [\phi_n] \quad (\text{orthonormal and complete}),$$

$$\Delta \phi_n + \lambda_n \phi_n = 0, \quad \text{and}$$

$$0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots, \quad \lim_{n \rightarrow \infty} \lambda_n = \infty.$$
■

Note that Hilbert-Schmidt allows us to conclude as much for our integral operator with kernel  $G_s(z; w; \chi)$ , and Propositions 3.10 allows us to transfer those statements to the operator we really care about. We emphasize that when  $\chi(S) \neq 1$ , the spectrum of  $\Delta$  is *discrete* and *reasonable*.

3.2.2.  $\chi(S) = 1$ . It is time to give a more detailed, though still mostly non-rigorous and very incomplete, characterization of the spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  in the case that  $\chi(S) = 1$ . This case gives us the rich decomposition (3.2), quite unlike either of the previous cases. The key reference for all statements made in this section is [23, Ch. 6 §9, 13 & Ch. 7].

We begin by trying to isolate the continuous part of the spectrum. We would expect a decomposition like

$$L^2(\Gamma \backslash \mathfrak{H}, \chi) = A \oplus \mathcal{E},$$

where  $A$  is spanned by the  $L^2$  eigenfunctions and  $\mathcal{E}$  corresponds to the continuous spectrum. It can be shown [23, Ch. 10, Claim 9.1, 9.2] that if  $\phi \in C^2(\Gamma \backslash \mathfrak{H}) \cap L^2(\Gamma \backslash \mathfrak{H}, \chi)$  is nonzero and satisfies  $\Delta\phi + \lambda\phi = 0$ , then  $\lambda \in \mathbf{R}$ , and, in fact,  $\lambda \geq 0$ . The case  $\lambda = 0$  occurs iff  $\phi(z)$  is identically constant. Let  $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$  be the eigenvalues of  $\Delta|_A$ , the restriction of  $\Delta$  to  $A$ . It can also be shown that the eigenfunctions corresponding to the eigenvalues  $\lambda_n$  are mutually orthogonal, and (again using the ‘separation of variables’ via Bessel functions) that an eigenfunction  $\phi_n$  has the explicit Fourier expansion

$$(3.6) \quad \phi_n(x + iy) = b_n y^{1-s} + \sum_{m \neq 0} a_m y^{\frac{1}{2}} K_{s-\frac{1}{2}}(2\pi|m|y) e(mx),$$

and  $b_n = 0$  whenever  $\lambda_n \geq \frac{1}{4}$ .

We let  $\mathcal{L}_2(\lambda)$  be the subspace generated by those  $\phi_n$  satisfying  $\Delta\phi_n + \lambda\phi_n = 0$ ,  $A_0(\lambda)$  be the subspace of  $\mathcal{L}_2(\lambda)$  satisfying the condition  $b = 0$  in (3.6), and  $A_1(\lambda)$  be the orthogonal complement of  $A_0(\lambda)$  in  $\mathcal{L}_2(\lambda)$ . We then form the spaces

$$A_0 := \bigoplus_{0 < \lambda < \infty} A_0(\lambda)$$

$$A_1 := \bigoplus_{0 < \lambda < \frac{1}{4}} A_1(\lambda).$$

We arrive at the decompositions

$$A = \delta(\chi)\mathbf{C} \oplus A_0 \oplus A_1$$

$$L^2(\Gamma \backslash \mathfrak{H}, \chi) = \delta(\chi)\mathbf{C} \oplus A_0 \oplus A_1 \oplus \mathcal{E}.$$

We note that the appearance of the  $\delta$  as defined in (3.1) is because  $\lambda_0$  is omitted when  $\chi \neq 1$ .

Proceeding now to the continuous spectrum, since  $\dim(\delta(\chi)\mathbf{C} \oplus A_1) \neq \infty$ , orthogonality with respect to  $A_1$  is the most important condition. With  $\psi(y) \in C^\infty(\mathbf{R}^+)$  with compact

support, we form the Poincaré series

$$\theta_\psi(z) = \sum_{W_0 \in \Gamma_\infty \backslash \Gamma} \chi(W_0^{-1}) \psi(\mathfrak{S}, W_0 z).$$

$\theta_\psi(z)$  is convergent on  $\mathfrak{H}$  compacta and vanishes at the cusp; hence  $\theta_\psi(z) \in L^2(\Gamma \backslash \mathfrak{H}, \chi)$ . It is not too hard to show to show that  $\theta_\psi \perp A_0$ . Namely [23, p.73–74],

$$\begin{aligned} \int_{\mathbb{F}} \theta_\psi(z) \overline{\phi_n(z)} \, d\mu(z) &= \sum_{W_0} \chi(W_0^{-1}) \int_{\mathbb{F}} \phi(\mathfrak{S} W_0 z) \overline{\phi_n(z)} \, d\mu(z) \\ &= \sum_{W_0} \int_{\mathbb{F}} \phi(\mathfrak{S} W_0 z) \overline{\phi_n(W_0 z)} \, d\mu(z) \\ &= \sum_{W_0} \int_{W_0(\mathbb{F})} \psi(\mathfrak{S} \xi) \overline{\phi_n(\xi)} \, d\mu(\xi) \\ &= \int_0^\infty \int_0^1 \psi(y) \overline{\phi_n(x + iy)} \frac{dx \, dy}{y^2}, \end{aligned}$$

since  $\sum W_0(\mathbb{F})$  is a fundamental region for  $\Gamma_\infty$ . But,

$$\int_0^1 \overline{\phi_n(x + iy)} \, dx = \overline{b_n} y^{1-s_n},$$

where  $\lambda_n = s_n(1 - s_n)$  with  $\Re s_n \geq \frac{1}{2}$ . Thus,

$$\langle \theta_\psi, \phi_n \rangle = \overline{b_n} \int_0^\infty \psi(y) y^{-s_n-1} \, dy,$$

ergo  $\theta_\psi \perp A_0$ . Then, by showing the dimension of  $\delta(\xi)\mathbb{C} \oplus A_1$  is finite, we see that the set

$$\mathcal{R} := \{ \psi : \phi \in C_{00}^\infty(\mathbf{R}), \text{supp } \psi \subset (0, \infty), \theta_\psi \perp A \}$$

is quite large. Furthermore, it can be shown that the operator  $L$  sending  $f(y) \mapsto y^2 f''(y)$  has continuous spectrum over  $W := \{ \theta_\psi : \psi \in \mathcal{R} \}$ , and, in fact,  $\mathcal{E} = \overline{W}$ .

Letting

$$L_\psi(s) := \int_0^\infty \psi(y) y^{s-1} \, dy,$$

an application of the Mellin transform to the function  $\theta_\psi$  yields

$$(3.7) \quad \theta_\psi(z) = \frac{1}{2\pi i} \int_{(\sigma)} L_\psi(-s) E(z; s; \chi) \, ds,$$



where  $\sigma > 1$  and  $E(z; s; \chi)$  is the Eisenstein series

$$(3.8) \quad E(z; s; \chi) := \sum_{W_0 \in \Gamma_\infty \backslash \Gamma} \chi(W_0^{-1}) (\Im W_0 z)^s.$$

We postpone the discussion of the meromorphic continuation of Eisenstein series to §4.1, and take it for granted for the time being. Pulling the line of integration to  $\Delta = \frac{1}{2}$ , we encounter no poles, and we obtain a transform of  $\theta_\psi(z)$  that involves  $E(z; \frac{1}{2} + it; \chi)$ . These Eisenstein series can be considered as ‘eigenpackets’ or ‘wave packets’ making up the continuous spectrum. They are themselves not in  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$ , but after convolution with the envelope  $L_\psi(s)$ , they are. Additionally, the corresponding  $\lambda$  are, with  $\sigma = \frac{1}{2}$ , given by  $\frac{1}{4} + t^2$ , and hence belong to  $[\frac{1}{4}, \infty)$ .

The incredible thing, is that, if one lets  $\{s_0, s_1, \dots, s_M\}$  be the  $s$ -values ( $\lambda = s(1-s)$  with  $\Re s \geq \frac{1}{2}$ ) according to  $\delta(\chi)\mathbf{C} \oplus A_1$  (a set that is known to be finite), then

**Proposition 3.14** ([23, Ch. 6, Claim 9.5]).  *$(s_0, s_1, \dots, s_M) \subset \{\text{the poles of } E(z; s; \chi)\}$ , where  $s_0$  exists iff  $\delta(\chi) = 1$ . In fact, the poles of  $E(z; s; \chi)$  located inside  $\{\Re s > \frac{1}{2}\}$  are simple and actually coincide with  $(s_0, s_1, \dots, s_M)$ . ■*

With a little more work, we arrive at the conclusion that the orthogonal complement of  $A_1$  in  $A$  is exactly given by residues  $\text{res}_{s=s_j} E(z; s; \chi)$ , with  $(1 - \delta(\chi)) \leq j \leq M$ . These are residues of the Eisenstein series  $E(z; s; \chi)$ . We also arrive at the characterization of  $\theta_\psi(z)$  given in (3.7). This allows us to conclude that  $\mathcal{E}$  is the closed subspace of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  generated by taking continuous superpositions of  $E(z; \frac{1}{2} + it; \chi)$  for  $0 \leq t < \infty$ .

This discussion has been non-rigorous, but should serve as a conceptual guide to the technical contortions required to put the decomposition (3.2) on solid ground. Every claim above can be justified, but the justifications are far too long to be contained in this exposition.

By way of a last remark in this subsection, we note that in the case of  $\Gamma = \text{SL}(2, \mathbf{Z})$ , the set  $A_1$  is empty; that is, there is no contribution from residues of Eisenstein series. The sensible reason for this is that in this case,  $E_s$  has one pole only in  $\Re s \geq \frac{1}{2}$ , and it is simple with residue a constant. See [17, 18] for a nice discussion of ‘eigenpackets’ arising from Eisenstein series, and a discussion of residues of Eisenstein series when  $\Gamma = \text{SL}(2, \mathbf{Z})$ . Also, note that when formulating the trace formula for  $\text{SL}(3, \mathbf{Z})$ , a case *not* treated here (among other things,  $\text{SL}(3, \mathbf{Z}) \backslash \text{SL}(3, \mathbf{R}) / \text{SO}(3, \mathbf{R})$  is not a quotient of  $\mathfrak{H} = \text{SL}(2, \mathbf{R}) / \text{SO}(2, \mathbf{R})$ ), residues of Eisenstein series very much *do* contribute; c.f. [20, 63–66, 71].

**3.3. The Selberg Trace Formula on  $\Gamma \backslash \mathfrak{H}$ .** In this subsection, we develop the Selberg trace formula for the space  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$ , where  $\Gamma$  is as in §3.1. We are going to form an integral operator of trace class and interpret it in two different ways to arrive at the trace formula. We follow [23, Ch. 6, §10]. We first consider the case  $\chi(S) \neq 1$ .

3.3.1.  $\chi(S) \neq 1$ . In contrast with before, we wish to consider  $\Phi \in C^4(\mathbf{R}^+)$  such that  $|\phi^{(k)}(t)| \leq A(t+4)^{-1-k-\delta}$  for  $0 \leq k \leq 4$  and  $\delta > 0$ . Without loss of generality we may take  $\Phi$  to be real. Note that  $\Phi$  is not assumed to have compact support. As before, we introduce the point-pair invariant

$$k(z, w) = \Phi \left( \frac{|z - w|^2}{\Im z \Im w} \right),$$

and the kernel function

$$K(z, w, \chi) = \sum_{\sigma \in \Gamma} \chi(\sigma) k(z, \sigma w).$$

It is not hard to show that  $k(z, w)$  converges on  $\mathfrak{H} \times \mathfrak{H}$  compacta, and  $K(z, w, \chi) \rightarrow 0$  as  $z \rightarrow i\infty$ . For a fixed  $w$ ,  $K(z, w, \chi)$  expands as a Fourier series

$$K(z, w, \chi) = \sum_{n=1}^{\infty} \Lambda(\lambda_n) \phi_n(z) \overline{\phi_n(w)},$$

which converges absolutely and uniformly on all of  $\mathfrak{H} \times \mathfrak{H}$ . Thanks to our choice of kernel, we have

$$\int_{\mathbb{F}} K(z, z, \chi) \, d\mu(z) = \sum_{n=1}^{\infty} \Lambda(\lambda_n).$$

It now comes down to attacking  $\int_{\mathbb{F}} K(z, z, \chi) \, d\mu(z)$ . After establishing the convergence of

$$\int_{\mathbb{F}} \left( \sum_{|\mathrm{tr}(\sigma)| \neq 2} |k(z, \sigma z)| \right) \, d\mu(z),$$

the spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  permits the assertion ([23, p.102])

$$\begin{aligned} \int_{\mathbb{F}} \left( \sum_{|\mathrm{tr}(\sigma)| \neq 2} |k(z, \sigma z)| \right) \, d\mu(z) &= \sum_{|\mathrm{tr}(\sigma)| \neq 2} \chi(\sigma) \int_{\mathbb{F}} k(z, \sigma z) \, d\mu(z) \\ &= \sum_{\substack{\sigma \\ \text{hyperbolic}}} \chi(\sigma) \int_{\mathfrak{F}[Z_{\Gamma}(\sigma)]} k(z, \sigma z) \, d\mu(z) \\ &+ \sum_{\substack{\eta \\ \text{elliptic}}} \chi(\sigma) \int_{\mathfrak{F}[Z_{\Gamma}(\eta)]} k(z, \eta z) \, d\mu(z) \\ (3.9) \qquad &= \sum_{\substack{\sigma \\ \text{hyperbolic}}} \frac{\chi(\sigma) \log N(\sigma_0)}{N(\sigma)^{1/2} - N(\sigma)^{-1/2}} g(\log N(\sigma)) \\ &+ \sum_{\substack{\eta \\ \text{elliptic}}} \frac{\chi(\eta)}{2m(\eta) \sin \theta(\eta)} \int_{-\infty}^{\infty} \frac{e^{-2\theta(\eta)r}}{1 + e^{-2\pi r}} h(r) \, dr. \end{aligned}$$

Some words of explanation:  $Z_\Gamma(\sigma)$  is the centralizer of  $\sigma$  in  $\Gamma$ .  $\mathfrak{F}[Z_\Gamma(\sigma)]$  is again the fundamental region of the subgroup  $Z_\Gamma(\sigma) \subset \Gamma$ .  $N(\sigma)$  is the multiplier or characteristic constant of the hyperbolic element  $\sigma$ .  $\sigma_0$  generates  $Z(\sigma)$ . Last,  $m(\eta) = |Z_\Gamma(\eta)|$ , and  $\theta(\sigma)$  is defined implicitly by the formula  $\text{tr } \sigma = 2 \cos \theta$ ,  $0 < \theta < \pi$ . This work depends essentially on the setting-up of the trace formula for vector-valued functions [22, p.348–352] and for automorphic forms on a compact surface [22, p.402–407].

In the case that  $|\text{tr } \sigma| = 2$ , we write  $w_0^{-1} S^k w_0$ , where  $w_0 \in \Gamma_\infty \backslash \Gamma$ . This coset decomposition is unique whenever  $\sigma \neq I$ . It again follows from the convergence

$$\int_{\mathbb{F}} \left( \sum_{\substack{\sigma \notin [S] \\ |\text{tr } \sigma| = 2}} |k(z, \sigma z)| \right) d\mu(z) < \infty$$

that

$$\begin{aligned} & \int_{\mathbb{F}} \left( \sum_{|\text{tr } \sigma| = 2} \chi(\sigma) k(z, \sigma z) \right) d\mu(z) \\ &= \int_{\mathbb{F}} \left( \sum_{n=-\infty}^{\infty} \chi(S^n) k(z, S^n z) \right) d\mu(z) \\ &+ \sum_{w_0 \notin [S]} \int_{\mathbb{F}} \left( \sum_{m \neq 0} \chi(w_0^{-1} S^m w_0) k(z, w_0^{-1} S^m w_0 z) \right) d\mu(z) \\ &= \int_{\mathbb{F}} \left( \sum_{n=-\infty}^{\infty} \chi(S^n) k(z, S^n z) \right) d\mu(z) \\ &+ \sum_{w_0 \notin [S]} \int_{\mathbb{F}} \left( \sum_{m \neq 0} \chi(S^m) k(w_0 z, S^m w_0 z) \right) d\mu(z) \\ &= \int_{\mathbb{F}} \left( \sum_{n=-\infty}^{\infty} \chi(S^n) k(z, S^n z) \right) d\mu(z) \\ &+ \sum_{w_0 \notin [S]} \int_{w_0(\mathbb{F})} \left( \sum_{m \neq 0} \chi(S^m) k(w, S^m w) \right) d\mu(w) \\ &= \Phi(0) \mu(\mathbb{F}) + \int_{\mathbb{F}} \left( \sum_{n \neq 0} \chi(S^n) k(z, S^n z) \right) d\mu(z) \\ &+ \sum_{w_0 \notin [S]} \int_{w_0(\mathbb{F})} \left( \sum_{m \neq 0} \chi(S^m) k(w, S^m w) \right) d\mu(w) \\ &= \Phi(0) \mu(\mathbb{F}) + \int_{\sum w_0(\mathbb{F})} \left( \sum_{m \neq 0} \chi(S^m) k(z, S^m z) \right) d\mu(z) \end{aligned}$$

The uniform convergence of  $\sum_{\sigma \in \Gamma} |k(z, \sigma w)|$  on  $\mathfrak{H} \times \mathfrak{H}$  compacta implies the continuity of  $\sum_{m \neq 0} \chi(S^m)k(z, S^m z)$ ; the series is also  $S$ -invariant. We may conclude that

$$\begin{aligned} & \int_{\mathbb{F}} \left( \sum_{|\text{tr } \sigma|=2} \chi(\sigma)k(z, \sigma z) \right) d\mu(z) \\ &= \Phi(0)\mu(\mathbb{F}) + \int_{\mathfrak{F}([\Gamma_\infty])} \left( \sum_{n \neq 0} \chi(S^n)k(z, S^n z) \right) d\mu(z) \\ &= \Phi(0)\mu(\mathbb{F}) + \int_0^1 \int_0^1 \left( \sum_{n \neq 0} \chi(S^n)k(z, S^n z) \right) \frac{dx dy}{y^2}. \end{aligned}$$

The second integral can be evaluated as follows ([23, p.104–105]).

$$\begin{aligned} & \int_0^1 \int_0^1 \left( \sum_{n \neq 0} \chi(S^n)k(z, S^n z) \right) \frac{dx dy}{y^2} \\ &= \int_0^1 \int_0^1 \left( \sum_{n \neq 0} \Phi\left(\frac{n^2}{y^2}\right) e(n\alpha) \right) \frac{dx dy}{y^2} \\ &= 2 \int_0^1 \left( \sum_{n=1}^{\infty} \Phi\left(\frac{n^2}{y^2}\right) \cos(2\pi n\alpha) \right) \frac{dy}{y^2} \\ &= 2 \int_0^1 \left( \sum_{n=1}^{\infty} \Phi(n^2 t^2) \cos(2\pi n\alpha) \right) dt \\ &= 2 \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^1 \left( \sum_{n=1}^{\infty} \Phi(n^2 t^2) \cos(2\pi n\alpha) \right) dt. \end{aligned}$$

Now,

$$\begin{aligned} & 2 \int_{\varepsilon}^1 \left( \sum_{n=1}^{\infty} \Phi(n^2 t^2) \cos(2\pi n\alpha) \right) dt \\ &= 2 \sum_{n=1}^{\infty} \cos(2\pi n\alpha) \int_{\varepsilon}^1 \Phi(n^2 t^2) dt \\ &= 2 \sum_{n=1}^{\infty} \frac{\cos(2\pi n\alpha)}{n} \int_{n\varepsilon}^{\infty} \Phi(u^2) du \\ &= 2 \sum_{n=1}^{\infty} \frac{\cos(2\pi n\alpha)}{n} \left( \int_0^{\infty} \Phi(u^2) du - \int_0^{n\varepsilon} \Phi(u^2) du \right) \end{aligned}$$

$$\begin{aligned}
&= 2 \int_0^\infty \Phi(u^2) \, du \sum_{n=1}^\infty \frac{\cos(2\pi n\alpha)}{n} \\
&- 2 \sum_{n=1}^\infty \frac{\cos(2\pi n\alpha)}{n} \int_0^{n\varepsilon} \Phi(u^2) \, du \\
&= 2 \int_0^\infty \Phi(u^2) \, du \log \left| \frac{1}{1 - e(\alpha)} \right| \\
&- 2 \sum_{n=1}^\infty \frac{\cos(2\pi n\alpha)}{n} \int_0^{n\varepsilon} \Phi(u^2) \, du.
\end{aligned}$$

It follows from summation by parts and trivial estimations that

$$2 \sum_{n=1}^\infty \frac{\cos(2\pi n\alpha)}{n} \int_0^{n\varepsilon} \Phi(u^2) \, du \ll \varepsilon \log \frac{1}{\varepsilon} + \varepsilon,$$

and hence that

$$\begin{aligned}
&2 \int_\varepsilon^\infty \left( \sum_{n=1}^\infty \Phi(n^2 t^2) \cos(2\pi n\alpha) \right) dt \\
&= \int_0^\infty \frac{\Phi(v)}{\sqrt{v}} \, dv \cdot \log \left| \frac{1}{1 - \chi(S)} \right| + O \left( \varepsilon \log \frac{1}{\varepsilon} \right) \\
&= g(0) \log \left| \frac{1}{1 - \chi(S)} \right| + O \left( \varepsilon \log \frac{1}{\varepsilon} \right)
\end{aligned}$$

and

$$\int_0^\infty \int_0^1 \left( \sum_{n \neq 0} \chi(S^n) k(z, S^n z) \right) \frac{dx \, dy}{y^2} = g(0) \log \left| \frac{1}{1 - \chi(S)} \right|.$$

Combining the above, we obtain

$$\begin{aligned}
\int_{\mathbf{F}} \mathbf{K}(z, z, \chi) \, d\mu(z) &= \frac{\mu(\mathbf{F})}{4\pi} \int_{-\infty}^{\infty} r b(r) \tanh(\pi r) \, dr \\
&\quad + g(0) \log \left| \frac{1}{1 - \chi(\mathbf{S})} \right| \\
&\quad + \sum_{\substack{[\sigma] \\ \text{hyperbolic}}} \frac{\chi(\sigma) \log \mathbf{N}(\sigma_0)}{\mathbf{N}(\sigma)^{1/2} - \mathbf{N}(\sigma)^{-1/2}} g(\log \mathbf{N}(\sigma)) \\
&\quad + \sum_{\substack{[\eta] \\ \text{elliptic}}} \frac{\chi(\eta)}{2m(\eta) \sin \theta(\eta)} \int_{-\infty}^{\infty} \frac{e^{-2\theta(\eta)r}}{1 + e^{-2\pi r}} b(r) \, dr,
\end{aligned}$$

where all notation is as above. We arrive at the final statement of the Selberg trace formula for  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  when  $\chi(\mathbf{S}) \neq 1$ .

**Theorem 3.15** (The Selberg trace formula for  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  when  $\chi(\mathbf{S}) \neq 1$ , [23, Ch. 6, Theorem 10.5]). *Suppose that  $b(r)$  satisfies the following hypotheses.*

- (a)  $b(r)$  is analytic on  $|\Im r| \leq \frac{1}{2} + \delta$  f.s.  $\delta > 0$ .
- (b)  $b(-r) = b(r)$ .
- (c)  $|b(r)| \ll (1 + |\Re r|)^{-2-\delta}$ .

Suppose also that

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} b(r) e^{-iru} \, dr \quad u \in \mathbf{R},$$

and let  $\chi$  be a  $1 \times 1$  unitary representation of  $\Gamma$  satisfying  $\chi(\mathbf{S}) \neq 1$ . Introduce  $\lambda_n = \frac{1}{4} + r_n^2$ . Then,

$$\begin{aligned}
\sum_{n=1}^{\infty} b(r_n) &= \frac{\mu(\mathbf{F})}{4\pi} \int_{-\infty}^{\infty} r b(r) \tanh(\pi r) \, dr + g(0) \log \left| \frac{1}{1 - \chi(\mathbf{S})} \right| \\
&\quad + \sum_{\substack{[\sigma] \\ \text{hyperbolic}}} \frac{\chi(\sigma) \log \mathbf{N}(\sigma_0)}{\mathbf{N}(\sigma)^{1/2} - \mathbf{N}(\sigma)^{-1/2}} g(\log \mathbf{N}(\sigma)) \\
&\quad + \sum_{\substack{[\eta] \\ \text{elliptic}}} \frac{\chi(\eta)}{2m(\eta) \sin \theta(\eta)} \int_{-\infty}^{\infty} \frac{e^{-2\theta(\eta)r}}{1 + e^{-2\pi r}} b(r) \, dr.
\end{aligned}$$

■

3.3.2.  $\chi(S) = 1$ . In this section, we follow [23, Ch. 6, §13], which, in turn, mostly follows [56, pp. 64–85]. We would first like to unwind what we might hope for the development of Selberg’s trace formula for  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  when  $\chi(S) = 1$ . We recall the spectral decomposition described in §3.2.2, and stated in (3.2); namely,

$$L^2(\Gamma \backslash \mathfrak{H}, \chi) = \delta(\chi)\mathbf{C} \oplus L^2_{\text{cusp}}(\Gamma \backslash \mathfrak{H}) \oplus L^2_{\text{o}}(\Gamma \backslash \mathfrak{H}) \oplus L^2_{\text{cont}}(\Gamma \backslash \mathfrak{H}).$$

Imprecisely stated, the naïve hope for obtaining a convergent, useful trace formula would be to try to ‘subtract’ off the contribution of the trace of an integral operator arising from the continuous spectrum, and thereby arrive at something well-behaved. Since (c.f. [23, (9.35), p.91], for a nice enough  $f \in L^2(\Gamma \backslash \mathfrak{H}, \chi)$ , we should have, following from the spectral decomposition,

$$(3.10) \quad f(z) = \sum c_n \phi_n(z) + \int_0^\infty g(t) E\left(z; \frac{1}{2} + it, \chi\right) dt,$$

$$\text{where } c_n = \langle f, \phi_n \rangle \text{ and } g(t) = \frac{1}{2\pi} \int_{\mathbb{F}} \overline{f(z)} E\left(z; \frac{1}{2} + it, \chi\right) d\mu(z).$$

That is, if  $k(z, w)$  is a reasonable point-pair invariant, and we form

$$\mathbf{K}(z, w, \chi) = \sum_{\sigma \in \Gamma} \chi(\sigma) k(z, \sigma w)$$

as usual, then (3.10) implies that we should expect that

$$\mathbf{K}(z, w, \chi) = \sum b(r_n) \overline{\phi_n(w)} \phi_n(z) + \int_0^\infty F(t) E\left(z; \frac{1}{2} + it, \chi\right) dt,$$

where

$$F(t) = \frac{1}{2\pi} \int_{\mathbb{F}} \overline{\mathbf{K}(z, w, \chi)} E\left(z; \frac{1}{2} + it, \chi\right) d\mu(z).$$

It actually follows from Proposition 2.1 that

$$F(t) = \frac{1}{2\pi} \int_{\mathfrak{H}} \overline{k(\xi, w)} E\left(\xi; \frac{1}{2} + it, \chi\right) d\mu(\xi) = \frac{1}{2\pi} \overline{h(t)} E\left(w; \frac{1}{2} + it, \chi\right),$$

and therefore that

$$(3.11) \quad \begin{aligned} \mathbf{K}(z, w, \chi) &= \sum b(r_n) \overline{\phi_n(w)} \phi_n(z) + \frac{1}{2\pi} \int_0^\infty \overline{h(t)} E\left(w; \frac{1}{2} + it, \chi\right) E\left(z; \frac{1}{2} + it, \chi\right) dt \\ &= \sum b(r_n) \overline{\phi_n(w)} \phi_n(z) + \mathbf{H}(z, w, \chi), \quad \text{say.} \end{aligned}$$

Our guess, then, is that the Selberg trace formula will result from an identity of the shape

$$\int_{\mathbb{F}} (\mathbf{K}(z, z, \chi) - \mathbf{H}(z, z, \chi)) d\mu(z) = \sum b(r_n).$$

It can be shown that, if one writes down the decomposition (3.9) in the case  $\Phi(t) = (t+4)^{-\gamma}$  and  $\chi \equiv 1$ , one may show that the hyperbolic terms are well-behaved, and

$$\sum_{\substack{[\sigma] \\ \text{hyperbolic}}} \frac{\log N(\sigma_0)}{N(\sigma)^\gamma} < \infty \quad \forall \gamma > 1.$$

Any difficulties in effecting the formula in this case, then, will arise from the parabolic terms. Unfortunately, handling these terms is quite thorny because of the growth properties of  $K(z, w, \chi)$ , but everything works out in the end. We summarize the results.

**Proposition 3.16** ([23, Ch. 6, Prop. 13.3]). *For  $y, v$  sufficiently large,*

$$K(z, w, \chi) = \sqrt{yv}g(\log s/v) + O(1).$$

*The implied constant depends only on  $\Gamma, \chi, F$ , and  $h$ .* ■

**Proposition 3.17** ([23, Ch. 6, Prop. 13.4]). *Assume  $h : \mathbf{R} \rightarrow \mathbf{R}$  is an even test function, analytic and of sufficiently rapid decay ( $\ll e^{-7|r|}$ ) in the vertical strip  $|\Im r| \leq \frac{1}{2} + \delta$ ,  $\delta > 0$ . Let*

$$K_0(z, w, \chi) := K(z, w, \chi) - H(z, w, \chi).$$

*Let  $f \in C^\infty(H) \cap L^2(\Gamma \backslash \mathfrak{H}, \chi)$  satisfy  $\Delta f + (\frac{1}{4} + \xi^2)f = 0$  over  $\mathfrak{H}$ . Then:*

- (i)  $K_0(z, w, \chi) \in C(\mathfrak{H} \times \mathfrak{H})$ ,
  - (ii)  $K_0(w, z, \chi) = \overline{K_0(z, w, \chi)}$ ,
  - (iii)  $K_0(\sigma z, \eta w, \chi) = \chi(\sigma)K_0(z, w, \chi)\overline{\chi(\eta)}$  for  $\sigma, \eta \in \Gamma$ ,
  - (iv)  $K_0(z, w, \chi) = O\left((yv)^{1-\beta}\right)$  for a small, effective constant  $\beta$  for  $y, v \gg 1$ ,
  - (v)  $K_0(z, w, \chi) \in L^2(F \times F)$ ,
  - (vi)  $\int_F K_0(z, w, \chi)f(w) d\mu(w) = h(\xi)f(z)$  for  $z \in \mathfrak{H}$  and some  $\xi$  with  $|\Im \xi| \leq \frac{1}{2} + \delta$ .
- 

The following is a crucial technical result that characterizes the  $L^2$  space and some important integral operators on it fairly completely.

**Proposition 3.18** ([23, Ch. 6, Prop. 13.6]). *We suppose that*

- (i)  $h(r)$  satisfies the assumptions of Proposition 3.17 and that moreover  $h(r) \ll e^{-15|r|}$ ,
- (ii)  $K_0(z, w, \chi)$  is defined as in Proposition 3.17,
- (iii)  $\mathcal{L}[f] = \int_F K_0(z, w, \chi)f(w) d\mu(w)$  for  $f \in L^2(\Gamma \backslash \mathfrak{H}, \chi)$ ,
- (iv)  $\mathbf{A}$  is the closed subspace generated by the  $L^2$  eigenfunctions  $\phi_n$ ,
- (v)  $\mathcal{E}$  is the orthogonal complement of  $\mathbf{A}$ , and
- (vi)  $\lambda_n = s_n(1 - s_n) = \frac{1}{4} + r_n^2$ .

*Then,*

- (a)  $\mathcal{E}$  is closed and  $L^2(\Gamma \backslash \mathfrak{H}, \chi) = \mathbf{A} \oplus \mathcal{E}$ ,
- (b)  $\sum (1 + |r_n|)^{-4} < \infty$ ,



- (c)  $\sum e^{-14|r_n|} |\phi_n(z)\overline{\phi_n(w)}| \ll (\gamma v)^{1-\beta}$  for  $\gamma, v \geq 1$  and  $\beta$  as in Proposition 3.17,
- (d)  $\sum h(r_n)\phi_n(z)\overline{\phi_n(w)}$  converges uniformly on  $\mathfrak{H} \times \mathfrak{H}$  compacta,
- (e)  $\mathcal{L}[\mathcal{E}] = 0$ ,
- (f)  $K_0(z, w, \chi) = \sum h(r_n)\phi_n(z)\overline{\phi_n(w)}$ , and
- (g) equation (3.11) is completely rigorous.

■

With Proposition 3.18 in hand, it is now more or less straightforward to obtain the Selberg trace formula for  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$ . We make a strong growth condition on the test function  $h$ ; namely

$$h(r) \ll e^{-15|r|} \quad \text{for} \quad |\Im r| \leq \frac{1}{2} + \delta.$$

We may proceed roughly in analogy with §3.3.1; see [23, p.196].

$$\begin{aligned}
\sum h(r_n) &= \int_{\mathbb{F}} K_0(z, z, \chi) \, d\mu(z) \\
&= \int_{\mathbb{F}} (K(z, z, \chi) - H(z, z, \chi)) \, d\mu(z) \\
&= \int_{\mathbb{F}} \left( \sum_{|\text{tr } \sigma|=2} \chi(\sigma)k(z, \sigma z) - H(z, z, \chi) \right) \, d\mu(z) \\
&\quad + \int_{\mathbb{F}} \left( \sum_{|\text{tr } \sigma| \neq 2} \chi(\sigma)k(z, \sigma z) \right) \, d\mu(z) \\
&= \int_{\mathbb{F}} \left( \sum_{|\text{tr } \sigma|=2} \chi(\sigma)k(z, \sigma z) - H(z, z, \chi) \right) \, d\mu(z) \\
&\quad + \sum_{\substack{[\sigma] \\ \text{hyperbolic}}} \frac{\chi(\sigma) \log N(\sigma_0)}{N(\sigma)^{1/2} - N(\sigma)^{-1/2}} g(\log N(\sigma)) \\
&\quad + \sum_{\substack{[\eta] \\ \text{elliptic}}} \frac{\chi(\eta)}{2m(\eta) \sin \theta(\eta)} \int_{-\infty}^{\infty} \frac{e^{-2\theta(\eta)r}}{1 + e^{-2\pi r}} h(r) \, dr \\
&= \int_{\mathbb{F}} k(z, z) \, d\mu(z) + \int_{\mathbb{F}} \left( \sum_{n \neq 0} k(z, S^n z) - H(z, z, \chi) \right) \, d\mu(z) \\
&\quad + \int_{\mathbb{F}} \left( \sum_{\substack{\sigma \notin [S] \\ |\text{tr } \sigma|=2}} \chi(\sigma)k(z, \sigma z) \right) \, d\mu(z) + \star
\end{aligned}$$

$$\begin{aligned}
&= \Phi(0)\mu(\mathbb{F}) + \int_{\mathbb{F}} \left( \sum_{n \neq 0} k(z, S^n z) - \mathbf{H}(z, z, \chi) \right) d\mu(z) \\
&+ \sum_{w_0 \notin [\mathbb{S}]} \int_{\mathbb{F}} \left( \sum_{m \neq 0} \chi(w_0^{-1} S^m w_0) k(z, w_0^{-1} S^m w_0 z) \right) d\mu(z) + \star \\
&= \Phi(0)\mu(\mathbb{F}) + \int_{\mathbb{F}} \left( \sum_{n \neq 0} k(z, S^n z) - \mathbf{H}(z, z, \chi) \right) d\mu(z) \\
&+ \sum_{w_0 \notin [\mathbb{S}]} \int_{\mathbb{F}} \left( \sum_{m \neq 0} \chi(S^m) k(w_0 z, S^m w_0 z) \right) d\mu(z) + \star \\
&= \Phi(0)\mu(\mathbb{F}) + \int_{\mathbb{F}} \left( \sum_{n \neq 0} k(z, S^n z) - \mathbf{H}(z, z, \chi) \right) d\mu(z) \\
&+ \sum_{w_0 \notin [\mathbb{S}]} \int_{w_0(\mathbb{F})} \left( \sum_{m \neq 0} k(w, S^m w) \right) d\mu(z) + \star
\end{aligned}$$

(noting that  $\sum_{\sigma \in \Gamma} |k(z, \sigma w)|$  is uniformly convergent on  $\mathfrak{H} \times \mathfrak{H}$  compacta,  $\sum_{m \neq 0} k(z, S^m z)$  is continuous and  $\mathbb{S}$ -invariant,  $\sum_{m \neq 0} |k(z, S^m z)| = O(y^{2+2\delta})$  for  $0 < y < 1$ )

$$\begin{aligned}
&= \Phi(0)\mu(\mathbb{F}) + \int_{\mathbb{F}} \left( \sum_{n \neq 0} k(z, S^n z) - \mathbf{H}(z, z, \chi) \right) d\mu(z) \\
&+ \int_{\mathcal{P}} \left( \sum_{m \neq 0} k(w, S^m w) \right) d\mu(w) + \star,
\end{aligned}$$

where  $\mathcal{P}$  is any region equivalent to  $\sum_{w_0 \notin [\mathbb{S}]} w_0(\mathbb{F})$  under  $\Gamma_\infty$ . With some more work ([23, p.198–199]) to make everything explicit, one obtains ([23, Eq. 13.20, p. 207]):

$$\begin{aligned}
(3.12) \quad \sum b(r_n) &= \frac{\mu(\mathbb{F})}{4\pi} \int_{-\infty}^{\infty} r b(r) \tanh(\pi r) dr - \frac{1}{2\pi} \int_{\mathbb{R}} b(r) \frac{\Gamma'(1+ir)}{\Gamma(1+ir)} dr \\
&+ \frac{1}{4} b(0) \left( 1 - \phi\left(\frac{1}{2}\right) \right) - g(0) \log 2 + \frac{1}{4\pi} \int_{\mathbb{R}} b(t) \frac{\phi'\left(\frac{1}{2}+it\right)}{\phi\left(\frac{1}{2}+it\right)} dt \\
&+ \sum_{\substack{[\sigma] \\ \text{hyperbolic}}} \frac{\chi(\sigma) \log \mathbf{N}(\sigma)}{\mathbf{N}(\sigma)^{1/2} - \mathbf{N}(\sigma)^{-1/2}} g(\log \mathbf{N}(\sigma)) \\
&+ \sum_{\substack{[\eta] \\ \text{elliptic}}} \frac{\chi(\eta)}{2m(\eta) \sin \theta(\eta)} \int_{-\infty}^{\infty} \frac{e^{-2\theta(\eta)r}}{1 + e^{-2\pi r}} b(r) dr,
\end{aligned}$$

where

$$(3.13) \quad \phi(s) := \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \sum_{w_0 \notin [S]} \chi(w_0^{-1}) \frac{1}{|c|^{2s}},$$

where  $w_0$  is understood as  $\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ . All the terms in Equation (3.12) are absolutely convergent. When  $\Gamma = \text{PSL}(2, \mathbf{Z})$  and  $\chi \equiv 1$ ,

$$\phi(s) = \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \sum_{k=1}^{\infty} \frac{\phi(k)}{k^{2s}} = \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \frac{\zeta(2s-1)}{\zeta(2s)}.$$

$\phi(s)$  is important because it is closely related to the singularities of  $E(z; s; \chi)$ . It can be shown that

$$E(z; s; \chi) = y^s + \phi(s)y^{1-s} + O\left(\sqrt{\omega(t)}e^{3|t|-2\pi y}\right) \quad \text{for } y \text{ large,}$$

where  $\omega$  is defined as follows. See also Theorem 4.3 Let  $\{s_0, s_1, \dots, s_M\}$  be the eigenvalues corresponding to  $\delta(\chi)\mathbf{C} \oplus \mathbf{A}_1$ ; that is, the poles of Eisenstein series located inside  $\{\Re s > \frac{1}{2}\}$ . (The  $s_0$  entry is always omitted when  $\delta(\chi) = 0$ .) Let

$$q := \inf \left\{ x > 0 : \sum_{|c|=x} \chi(w_{00}^{-1}) \neq 0, \quad w_{00} \in \Gamma_{\infty} \backslash \Gamma / \Gamma_{\infty} \right\}.$$

Form  $V(s)$  as

$$V(s) := q^{2s-1} \phi(s) \prod_{k=0}^M \frac{s - s_k}{1 - s - s_k}.$$

Then define

$$\omega(r) := 1 - \frac{V'(\frac{1}{2} + ir)}{V(\frac{1}{2} + ir)} \quad \text{for } r \in \mathbf{R}.$$

Evidently,  $\omega$  is a way of counting zeros of  $\phi(s)$ , and measuring the contribution to the spectrum from the Eisenstein series.

We emphasize

$$L^2(\Gamma \backslash \mathfrak{H}, \chi) = \mathbf{A} \oplus \mathcal{E} = \sum [\phi_n] \oplus \mathcal{E},$$

as in Proposition 3.18. We arrive at our final theorem.

**Theorem 3.19** (The Selberg trace formula for  $L^2(\Gamma \backslash \mathfrak{H}, \chi)$  when  $\chi(S) = 1$ , c.f. [23, Theorem 13.8, p. 209–210]). *Suppose that  $h(r)$  satisfies the following hypotheses.*

- (a)  $h(r)$  is analytic on  $|\Im r| \leq \frac{1}{2} + \delta$  f.s.  $\delta > 0$ ,
- (b)  $h(-r) = h(r)$ , and
- (c)  $|h(r)| \ll (1 + |\Re r|)^{-2-\delta}$ .

Suppose further that

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} b(r) e^{-iru} \, dr, \quad u \in \mathbf{R}.$$

Take  $\chi(S) = 1$  and introduce  $\lambda_n = s_n(1 - s_n) = \frac{1}{4} + r_n^2$ , as in Proposition 3.18. Then, with  $r_n$  as in Proposition 3.18,

$$\begin{aligned} \sum b(r_n) &= \frac{\mu(\mathbf{F})}{4\pi} \int_{-\infty}^{\infty} b(r) \tanh(\pi r) \, dr - \frac{1}{2\pi} \int_{\mathbf{R}} b(r) \frac{\Gamma'(1+ir)}{\Gamma(1+ir)} \, dr \\ &\quad + \frac{1}{4} b(0) \left( 1 - \phi\left(\frac{1}{2}\right) \right) - g(0) \log 2 + \frac{1}{4\pi} b(t) \frac{\phi'\left(\frac{1}{2}+it\right)}{\phi\left(\frac{1}{2}+it\right)} \, dt \\ &\quad + \sum_{\substack{[\sigma] \\ \text{hyperbolic}}} \frac{\chi(\sigma) \log N(\sigma_0)}{N(\sigma)^{1/2} - N(\sigma)^{-1/2}} g(\log N(\sigma)) \\ &\quad + \sum_{\substack{[\eta] \\ \text{elliptic}}} \frac{\chi(\eta)}{2m(\eta) \sin \theta(\eta)} \int_{-\infty}^{\infty} \frac{e^{-2\theta(\eta)r}}{1 + e^{-2\pi r}} b(r) \, dr. \end{aligned}$$

All sums and integrals converge absolutely. Further, we have the following version of Weyl's law.

$$(i) \quad \#\{0 \leq r_n \leq R\} + \frac{1}{4\pi} \int_{-R}^R \omega(r) \, dr \sim \frac{\mu(\mathbf{F})}{4\pi} R^2, \text{ and}$$

$$(ii) \quad \omega(r) + \frac{\phi'}{\phi}\left(\frac{1}{2} + ir\right) = O(1).$$

■

Theorems 3.15 and 3.19 define 'version A' of the Selberg trace formula, which includes the case  $L^2(\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}) := L^2(\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}, 1)$  (c.f. Table 1).

#### 4. THE SPECTRUM IN DETAIL

In §3, we considered a general discrete Fuchsian group  $\Gamma \subset \mathrm{PSL}(2, \mathbf{R})$  whose fundamental region  $F$  had finite (non-Euclidean) area. In this section and all that follows, we will restrict to the primary case we care about; the case of the arithmetic lattice  $\Gamma = \mathrm{PSL}(2, \mathbf{Z})$ . Having recorded the genesis and main rank-two development (for our purposes) of the Selberg trace formula in several compact and arithmetic cases, we now proceed to enumerate some of the remarkable properties of the automorphic objects that constitute the spectrum  $L^2(\Gamma \backslash \mathfrak{H})$ .

**4.1. The Continuous Spectrum: Eisenstein series on  $\Gamma \backslash \mathfrak{H}$ .** We assume familiarity with the generalized upper half-plane and coset decomposition of  $\mathfrak{H} = \mathrm{SL}(2, \mathbf{R})/\mathrm{SO}(2, \mathbf{R})$ . See [20, Ch. 1]. We remain in the rank 2 case. We follow the exposition of [20, Ch. 3, §3.1] in this section. Define the function  $I_s(z) := y^s$ .  $I_s$  is an eigenfunction of the hyperbolic Laplacian

$\Delta = -y^2 \left( \frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y} \right)$  with eigenvalue  $s(1-s)$ . If  $\Re s \geq \frac{1}{2}$ , the function  $I_s(z)$  is neither automorphic for  $\mathrm{SL}(2, \mathbf{Z})$ , nor in  $L^2(\Gamma \backslash \mathfrak{H})$  (integration with respect to Haar measure). Perhaps the most naïve thing to do is to try to average  $I_s(z)$  over the group to get an automorphic function. Since  $I_s(z)$  is invariant under  $\Gamma_\infty$ , we factor out by this subgroup. Cosets  $\Gamma_\infty \backslash \mathrm{SL}(2, \mathbf{Z})$  are determined like

$$\Gamma_\infty \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \left\{ \begin{pmatrix} * & * \\ c & d \end{pmatrix} \right\} = \left\{ \begin{pmatrix} u & -v \\ c & d \end{pmatrix} : du + cv = 1 \right\}.$$

Each relatively prime pair  $(c, d)$  determines a coset. This naïve idea produces an Eisenstein series. Compare to (3.8).

**Definition 4.1.** *Let  $z \in \mathfrak{H}$ ,  $\Re s > 1$ . Form the Eisenstein series*

$$E(z, s) := \sum_{\gamma \in \Gamma_\infty \backslash \mathrm{SL}(2, \mathbf{Z})} \frac{I_s(\gamma z)}{2} = \frac{1}{2} \sum_{\substack{c, d \in \mathbf{Z} \\ (c, d) = 1}} \frac{y^s}{|cz + d|^{2s}}.$$

The following proposition establishes that the series we have just written down is well-defined.

**Proposition 4.2** ([20, §3.1, Prop. 3.1.3]). *The Eisenstein series  $E(z, s)$  converges absolutely and uniformly on compact sets for  $z \in \mathfrak{H}$  and  $\Re s > 1$ . It is real-analytic in  $z$  and complex-analytic  $s$ . In addition,*

(1) *Fix  $\varepsilon > 0$ . For  $\sigma = \Re s \geq 1 + \varepsilon > 1$ , there exists a constant  $c(\varepsilon)$  such that*

$$|E(z, s) - y^s| \leq c(\varepsilon) y^{-\varepsilon} \quad \text{for } y \geq 1.$$

(2)  $E\left(\frac{az+b}{cz+d}, s\right) = E(z, s)$  for all  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbf{Z})$ .

(3)  $\Delta E(z, s) = s(1-s)E(z, s)$ .

*Proof.* For  $y \geq 1$ , we have

$$\begin{aligned} |E(z, s) - y^s| &\leq \sum_{\substack{(c, d) = 1 \\ c > 0}} \frac{1}{c^{2\sigma}} \cdot \frac{y^\sigma}{\left|z + \frac{d}{c}\right|^{2\sigma}} \\ &= y^\sigma \sum_{c \geq 1} \sum_{\substack{r=1 \\ (r, c) = 1}}^c \frac{1}{c^{2\sigma}} \sum_{m \in \mathbf{Z}} \frac{1}{\left|z + \frac{r}{c} + m\right|^{2\sigma}}. \end{aligned}$$

Since the set  $\left\{ \left|z + \frac{r}{c} + m\right| : m \in \mathbf{Z}, 1 \leq r \leq c, (r, c) = 1 \right\}$  forms a set spaced by  $\frac{1}{c}$ , we may majorize each term so that

$$\ell \leq \left|z + \frac{r}{c} + m\right| < \ell + 1$$

f.s.  $\ell \in \mathbf{Z}$ . There are at most  $\phi(c)$  such terms for each  $\ell$ . It follows that

$$\begin{aligned} |\mathbf{E}(z, s) - y^s| &\leq y^\sigma \sum_{c=1}^{\infty} \frac{\phi(c)}{c^{2\sigma}} \sum_{\ell \in \mathbf{Z}} \frac{1}{(\ell^2 + y^2)^\sigma} \\ &\leq 2y^\sigma \frac{\zeta(2\sigma - 1)}{\zeta(2\sigma)} \sum_{\ell=0}^{\infty} \frac{1}{(\ell^2 + y^2)^\sigma} \\ &\leq 2y^\sigma \frac{\zeta(2\sigma - 1)}{\zeta(2\sigma)} \left( y^{-2\sigma} + \int_0^{\infty} \frac{du}{(u^2 + y^2)^\sigma} \right) \\ &\ll y^{1-\sigma}. \end{aligned}$$

The second statement follows from the fact that  $\forall \gamma \in \mathrm{SL}(2, \mathbf{Z}), \gamma(\Gamma_\infty \backslash \mathrm{SL}(2, \mathbf{Z})) = (\Gamma_\infty \backslash \mathrm{SL}(2, \mathbf{Z}))$ . The third follows from the eigenvalue of  $\mathbf{I}_s(z)$ .  $\blacksquare$

We now describe the Fourier expansion of  $\mathbf{E}(z, s)$ .

**Theorem 4.3** ([20, p.58, Theorem 3.1.8]). *Let  $\Re s > 1$  and  $z = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} \in \mathfrak{H}$ . The Eisenstein series  $\mathbf{E}(z, s)$  has the Fourier expansion*

$$\mathbf{E}(z, s) = y^s + \phi(s)y^{1-s} + \frac{2\pi^s \sqrt{y}}{\Gamma(s)\zeta(2s)} \sum_{n \neq 0} \sigma_{1-2s}(n) |n|^{s-\frac{1}{2}} K_{s-\frac{1}{2}}(2\pi|n|y) e(nx),$$

where

$$(4.1) \quad \phi(s) = \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \frac{\zeta(2s - 1)}{\zeta(2s)},$$

$$\sigma_s(n) = \sum_{\substack{d|n \\ d>0}} d^s,$$

$$K_s(y) = \frac{1}{2} \int_0^{\infty} e^{-\frac{1}{2}y(u+1/u)} u^s \frac{du}{u}, \text{ and}$$

$$S(n; c) = \sum_{\substack{r=1 \\ (r,c)=1}}^c e\left(\frac{nr}{c}\right) = \sum_{\substack{\ell|n \\ \ell|c}} \ell \mu\left(\frac{c}{\ell}\right) \text{ is the Ramanujan sum.}$$

*Proof.* First note that

$$\zeta(2s)\mathbf{E}(z, s) = \zeta(2s)y^s + \sum_{c>0} \sum_{d \in \mathbf{Z}} \frac{y^s}{|cz + d|^{2s}}.$$

Letting  $\delta_{n,0} = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$  and  $d = mc + r$ , it follows that

$$\begin{aligned} & \zeta(2s) \int_0^1 E(z,s) e(-nx) \, dx \\ &= \zeta(2s) y^s \delta_{n,0} + \sum_{c=1}^{\infty} c^{-2s} \sum_{r=1}^c \sum_{m \in \mathbf{Z}} \int_0^1 \frac{y^s e(-nx)}{|z + m + \frac{r}{c}|^{2s}} \, dx \\ &= \zeta(2s) y^s \delta_{n,0} + \sum_{c=1}^{\infty} c^{-2s} \sum_{r=1}^c \sum_{m \in \mathbf{Z}} \int_{m+r/c}^{1+m+r/c} \frac{y^s e(-n(x-r/c))}{|z|^{2s}} \, dx \\ &= \zeta(2s) y^s \delta_{n,0} + \sum_{c=1}^{\infty} c^{2s} \sum_{r=1}^c e\left(\frac{nr}{c}\right) \int_{-\infty}^{\infty} \frac{y^s e(-nx)}{(x^2 + y^2)^s} \, dx. \end{aligned}$$

Since

$$\sum_{r=1}^c e\left(\frac{nr}{c}\right) = \begin{cases} c & c \mid n \\ 0 & c \nmid n, \end{cases}$$

$$\zeta(2s) \int_0^1 E(z,s) e(-nx) \, dx = \zeta(2s) y^s \delta_{n,0} + \sigma_{1-2s}(n) y^{1-s} \int_{-\infty}^{\infty} \frac{e(-nxy)}{(x^2 + 1)^s} \, dx,$$

where we understand  $\sigma_{1-2s}(0) = \zeta(1-2s)$ . The theorem is now a consequence of the Fourier transform identity

$$(4.2) \quad \int_{-\infty}^{\infty} \frac{e(-xy)}{(x^2 + 1)^s} \, dx = \begin{cases} \sqrt{\pi} \frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)} & y = 0 \\ \frac{2\pi^s |y|^{s-1/2}}{\Gamma(s)} K_{s-\frac{1}{2}}(2\pi|y|) & y \neq 0. \end{cases}$$

This identity, in turn, is established by fact that  $e^{-\pi x^2}$  is self-dual under Fourier transform, applied in the following way.

$$\begin{aligned} \Gamma(s) \int_{-\infty}^{\infty} \frac{e(-xy)}{(x^2 + 1)^s} \, dx &= \int_0^{\infty} \int_{-\infty}^{\infty} e^{-u-2\pi ixy} \left(\frac{u}{1+x^2}\right)^s \, dx \frac{du}{u} \\ &= \int_0^{\infty} e^{-u} u^s \int_{-\infty}^{\infty} e^{-ux^2} e(-xy) \, dx \frac{du}{u} \\ &= \sqrt{\pi} \int_0^{\infty} e^{-u-\pi^2 y^2/u} u^{s-1/2} \frac{du}{u}. \quad \blacksquare \end{aligned}$$

Next, we record the essential meromorphic continuation and functional equation of  $E(z,s)$ ; this is used implicitly in many of the results of the previous section.

**Theorem 4.4** ([20, p.59, Theorem 3.1.10]). *Let  $z \in \mathfrak{H}$  and  $s \in \mathbf{C}$  with  $\Re s > 1$ . The Eisenstein series  $E(z, s)$  and the function  $\phi(s)$  as in 4.1 can be continued to meromorphic functions on  $\mathbf{C}$  satisfying the functional equations*

- (1)  $\phi(s)\phi(1-s) = 1$ ;
- (2)  $E(z, s) = \phi(s)E(z, 1-s)$ .

*The modified function  $E^*(z, s) = \pi^{-s}\Gamma(s)\zeta(2s)E(z, s)$  is regular except for simple poles at  $s = 0, 1$ , and satisfies the functional equation  $E^*(z, s) = E^*(z, 1-s)$ . The residue of the pole at  $s = 1$  (for all  $z \in \mathfrak{H}$ ) is given by*

$$\operatorname{res}_{s=1} E(z, s) = \frac{3}{\pi}. \quad \blacksquare$$

It was Selberg [56, 59] who first established the spectral theory and meromorphic continuation of Eisenstein series on  $GL(2)$ , but Maaß who first made the Definition 4.1. Roelcke studied Eisenstein series attached to discrete groups other than  $SL(2, \mathbf{Z})$ . Selberg [58] established the meromorphic continuation of Eisenstein series for higher-rank groups, and Langlands [36, 37] brought this project to resolution in a pair of significant papers in which he established, in the most general context, the meromorphic continuation of Eisenstein series and complete spectral decomposition of arithmetic quotients  $\Gamma \backslash G$ , where  $G$  is a reductive group and  $\Gamma$  is an arithmetic subgroup. For more, see Arthur's exposition [2] of Eisenstein series, especially Langlands' work, and the trace formula.

**4.2. The Discrete Cuspidal Spectrum: Maaß forms on  $\Gamma \backslash \mathfrak{H}$ .** We turn now to the discrete spectrum, the Maaß forms on  $\Gamma \backslash \mathfrak{H} = SL(2, \mathbf{Z}) \backslash \mathfrak{H}$ . In this section, we follow [20, Ch. 3]. A Maaß form  $f: \Gamma \backslash \mathfrak{H} \rightarrow \mathbf{C}$  is square-integrable; that is,

$$\int_{\Gamma \backslash \mathfrak{H}} |f(z)|^2 \frac{dx \, dy}{y^2} < \infty,$$

$(dx \, dy)/y^2$  is the invariant Haar measure on  $\Gamma \backslash \mathfrak{H}$ . Maaß forms are automorphic objects; that is, they are  $\Gamma$ -invariant hyperbolic eigenfunctions of the Laplacian. More formally,

**Definition 4.5** ([20, §3.3, Defn. 3.3.1]). *Let  $\nu \in \mathbf{C}$ . A Maaß form of type  $\nu$  for  $SL(2, \mathbf{Z})$  is a non-zero function  $f \in L^2(\Gamma \backslash \mathfrak{H})$  that satisfies*

- (1)  $f(\gamma z) = f(z) \forall \gamma \in \Gamma, z = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} \in \mathfrak{H}$ ,
- (2)  $\Delta f = \nu(1-\nu)f$ , and
- (3)  $\int_0^1 f(z) \, dx = 0$  (cuspidality condition).

It is easy to show that if  $f$  is a Maaß form of type  $\nu$  for  $SL(2, \mathbf{Z})$ ,  $\nu(1-\nu)$  is real and nonnegative, and that if  $\nu = 0$  or  $1$ ,  $f$  is a constant function. The first task is to develop a useful Fourier theory for Maaß forms for  $SL(2, \mathbf{Z})$ . The two key ingredients in this development are Whittaker functions on  $\mathfrak{H}$  and multiplicity one for such functions.



4.2.1. *Fourier-Whittaker expansion.* The fact that  $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbf{Z})$  means that  $f(z)$  is a periodic function of  $x$ ; it must therefore admit a Fourier expansion of type

$$(4.3) \quad f(z) = \sum_{m \in \mathbf{Z}} A_m(y) e(mx).$$

Putting  $W_m(z) = A_m(y) e(mx)$ , by the absolute converges of the Fourier expansion,  $W_m(z)$  must satisfy

$$\begin{aligned} \Delta W_m(z) &= \nu(1 - \nu) W_m(z), \\ W_m \left( \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \cdot z \right) &= W_m(z) e(mu). \end{aligned}$$

We say that  $W_m(z)$  is a Whittaker function of type  $\nu$  associated to the additive character  $e(mx)$ . We state this implicit definition outright.

**Definition 4.6.** *A Whittaker function of type  $\nu$  associated to an additive character  $\psi : \mathbf{R} \rightarrow \mathbf{U}$ , where  $\mathbf{U} = \mathbf{C}^\times / (0, \infty)$  is the complex unit circle, is a smooth nonzero function  $W : \mathfrak{H} \rightarrow \mathbf{C}$  satisfying the conditions*

$$\begin{aligned} \Delta W(z) &= \nu(1 - \nu) W(z), \\ W \left( \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \cdot z \right) &= W(z) e(mu). \end{aligned}$$

A Whittaker function  $W(z)$  of type  $\nu$  and character  $\psi$  can always be written in the form

$$W(z) = A_\psi(y) \cdot \psi(x),$$

since the function  $W(z)/\phi(x)$  is invariant under translations  $x \mapsto x + u \forall u \in \mathbf{R}$ ; hence, must be the constant function after fixing  $y$ .

There are two major simplifications of crucial importance to the  $\mathrm{GL}(2)$  theory: multiplicity one, and explicit realization of the one Whittaker function we can construct. Namely, since we know that the function  $I_\nu(z) = y^\nu$  (as in §4.1) satisfies  $\Delta I_\nu(z) = \nu(1 - \nu) I_\nu(z)$ , it is a good place to start if we want to actually write down a Whittaker function. The other condition is the basic observation that if  $h : \mathbf{R} \rightarrow \mathbf{C}$  is  $C^\infty$  and integrable, and  $\psi : \mathbf{R} \rightarrow \mathbf{C}$  is an additive character, then the function  $H(x) := \int_{-\infty}^{\infty} h(u_1 + x) \psi(-u_1) \, du_1$  satisfies the equation  $H(u + x) + \psi(u) H(x)$ . This follows from a simple change of variables.

This statement, plus the fact that, since  $\Delta$  is an invariant differential operator,  $\Delta I_\nu(\gamma z) = \nu(1 - \nu) I_\nu(\gamma z) \forall \gamma \in \mathrm{GL}(2, \mathbf{R})$ , means that the function

$$\begin{aligned} (4.4) \quad W(z, \nu, \psi) &:= \int_{-\infty}^{\infty} I_\nu \left( \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \cdot z \right) \psi(-u) \, du \\ &= \int_{-\infty}^{\infty} \left( \frac{y}{(u+x)^2 + y^2} \right)^\nu \psi(-u) \, du \\ &= \psi(x) \int_{-\infty}^{\infty} \left( \frac{y}{u^2 + y^2} \right)^\nu \psi(-u) \, du \end{aligned}$$

is a Whittaker function of type  $\nu$  and character  $\psi$ .

One fact that sets the case of  $\mathrm{GL}(2)$  apart from the higher-rank cases is that  $W(z, \nu, \psi)$  can be expressed explicitly in terms of common special functions.

**Proposition 4.7** ([20, Prop. 3.4.6, p. 65]). *Let  $\psi_m(u) := e(mu)$ , and let  $W(z, \nu, \psi_m)$  be the Whittaker function (4.4). Then*

$$W(z, \nu, \psi_m) = \sqrt{2} \frac{(\pi|m|)^{\nu-\frac{1}{2}}}{\Gamma(\nu)} \sqrt{2\pi y} K_{\nu-\frac{1}{2}}(2\pi|m|y) \cdot e(mx),$$

where

$$K_\nu(y) = \frac{1}{2} \int_0^\infty e^{-y(u+1/u)/2} u^\nu \frac{du}{u}$$

is the classical K-Bessel function.

*Proof.* The naïve Fourier expansion (4.3) and the definition (4.4) imply that

$$W(z, \nu, \psi_m) = W(y, \nu, \psi_m) \cdot e(mx),$$

where

$$\begin{aligned} W(y, \nu, \psi_m) &= \int_{-\infty}^{\infty} \left( \frac{y}{u^2 + y^2} \right)^\nu \cdot e(-um) \, du \\ &= y^{1-\nu} \int_{-\infty}^{\infty} \frac{e(-uym)}{(u^2 + 1)^\nu} \, du. \end{aligned}$$

See (4.2) to complete the proof. ■

The crucial ingredient in the Fourier theory of Maaß forms for  $\mathrm{SL}(2, \mathbf{Z})$  is the multiplicity one principle, which does generalize to the higher-rank cases. It is of such central importance, since it means that the one Whittaker function that we can construct is in fact the only one we really need to construct.

**Theorem 4.8** (Multiplicity one, [20, Theorem 3.4.8, p. 66]). *Let  $\Psi(z)$  be a Whittaker function for  $\mathrm{SL}(2, \mathbf{Z})$  of type  $\nu \neq 0, 1$  associated to an additive character  $\psi$ , which has rapid decay at the cusp. Then*

$$\Psi(z) = aW(z, \nu, \psi)$$

for some  $a \in \mathbf{C}$ , with  $W(z, \nu, \psi)$  given by (4.4). If  $\psi$  is trivial, then  $a = 0$ .

*Proof sketch.* Using the classical theory of differential equations, and the fact that  $\Psi(z)$  is an eigenfunction of the Laplacian, it can be shown that  $\Psi(y)$  satisfies a differential equation

with exactly two linearly-independent solutions over  $\mathbf{C}$ ; they are

$$\sqrt{2\pi|m|y} \cdot K_{\nu-\frac{1}{2}}(2\pi|m|y), \quad \text{and}$$

$$\sqrt{2\pi|m|y} \cdot I_{\nu-\frac{1}{2}}(2\pi|m|y),$$

where  $I_\nu$  and  $K_\nu$  are the classical I- and K-Bessel functions

$$I_\nu(y) = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}y\right)^{\nu+2k}}{k!\Gamma(k+\nu+1)}, \quad \text{and}$$

$$K_\nu(y) = \frac{\pi}{2} \cdot \frac{I_{-\nu}(y) - I_\nu(y)}{\sin \pi\nu} = \frac{1}{2} \int_0^{\infty} e^{-y(t+1/t)/2} t^\nu \frac{dt}{t}.$$

The asymptotics

$$I_\nu(y) \sim \frac{e^y}{\sqrt{2\pi y}}$$

$$K_\nu(y) \sim \frac{e^{-y}\sqrt{\pi}}{\sqrt{2y}},$$

plus the growth condition on  $\Psi(z)$ , obtain for us the theorem. ■

We put together our explicit expression for  $W(z, \nu, \psi_m)$  (Proposition 4.7), the naïve Fourier expansion (4.3), and multiplicity one (Theorem 4.8) in the following proposition, which is a completely explicit Fourier-Whittaker expansion of a Maaß form for  $SL(2, \mathbf{Z})$ .

**Proposition 4.9** (Fourier-Whittaker expansions on  $GL(2, \mathbf{R})$ , [20, Prop. 3.5.1, p. 67]). *Let  $f$  be a non-constant Maaß form of type  $\nu$  for  $SL(2, \mathbf{Z})$ . Then, for  $z \in \mathfrak{H}$ , we have the Whittaker expansion*

$$f(z) = \sum_{n \neq 0} a_n \sqrt{2\pi|y|} \cdot K_{\nu-\frac{1}{2}}(2\pi|n|y) \cdot e(nx),$$

with complex coefficients  $a_n$ .

*Proof.* Beginning with the naïve Fourier expansion (4.3), i.e.

$$f(z) = \sum_{n \in \mathbf{Z}} A_n(y) e(nx),$$

and noting that it follows from  $\Delta f = \nu(1-\nu)f$  that

$$\Delta(A_n(y)e(nx)) = \nu(1-\nu)A_n(y)e(nx),$$

we conclude that  $A_n(y)e(nx)$  must be a Whittaker function of type  $\nu$  associated to  $e(nx)$ . The assumption that  $f$  is non-constant means that  $\nu \neq 0, 1$ ; since  $f \in L^2(\Gamma \backslash \mathfrak{H})$ ,  $A_n(y)e(nx)$  must have polynomial growth at the cusp. The proposition now follows from the multiplicity one Theorem 4.8. ■

The mysteries of the discrete spectrum are myriad and devilishly hard. For example, questions pertaining to

- (a) the growth and distributional properties of Fourier coefficients of Maaß forms,
- (b) the distribution of ( $L^2$  mass) of eigenfunctions of the Laplacian on the modular surface,
- (c) possible distributional properties admitted by Maaß form zeros on the modular surface,
- (d) automorphy of L-functions attached to symmetric powers  $\text{sym}^r f$  of a fixed Maaß form,

and many others, are incredibly challenging. In particular, this is because the spectral and harmonic nature of these eigenfunctions resists interpretation in terms of algebra and cohomology. This is in blatant contrast to holomorphic cusp forms on  $GL(2)$ , whose coefficients admit an algebraic interpretation that, among other things, allowed Deligne [11] to prove best-possible bounds on the growth of holomorphic cusp form Fourier coefficients. Maaß cusp form coefficients are conjectured to obey the same growth bound as holomorphic cusp form coefficients, but entirely different tools will be necessary to prove the validity of this conjecture, which is known as the Ramanujan conjecture since it was Ramanujan who made the original conjecture for his tau function, which appears in the Fourier expansion of the discriminant cusp form, the first holomorphic cusp form for the full modular group. The Ramanujan conjecture is the topic of § 4.3, where questions (a) and (d) are broached. Question (b) is discussed in §4.5, and question (c) is discussed in §4.2.3.

*4.2.2. Hecke operators.* We would be remiss if we did not touch upon the Hecke theory associated to Maaß forms, since without it, we would be stuck with a far less coherent space of mysterious automorphic forms with many bizarre properties. In particular, though we will *not* delve into the L-theory of Hecke-Maaß forms, the Euler product crucial to the formation of the L-function  $L(s, f)$  attached to a Hecke-Maaß cusp form would not exist if it were not for the multiplicativity that the Hecke operators make transparent. We follow [20, §3.10–3.12].

We first formulate Hecke operators in the the general setting of a group  $G$  acting continuously on a topological space  $X$ . Let  $Z \subset G$  be a discrete subgroup. Assume the quotient  $Z \backslash X$  has a left  $Z$ -invariant measure  $dx$  and define the usual space  $L^2(Z \backslash X)$ . Let  $C_G(Z)$  denote the commensurator of  $Z$ , defined as

$$C_G(Z) := \{g \in G : (g^{-1}Zg) \cap Z \text{ has finite index in both } Z \text{ and } g^{-1}Zg\}.$$

With  $d = [Z : (g^{-1}Zg) \cap Z]$ , we have the coset decompositions

$$Z = \bigcup_{i=1}^d ((g^{-1}Zg) \cap Z) \delta_i$$

$$ZgZ = \bigcup_{i=1}^d Zg\delta_i,$$

which defines  $\{\delta_i\}_{i=1}^d$  implicitly.

The Hecke operators are then defined as follows.

**Definition 4.10** ([20, Defn. 3.10.5, p. 75]). *Let a group  $G$  act continuously on a topological space  $X$  and let  $Z$  be a discrete subgroup of  $G$ . For each  $g \in C_G(Z)$ , we define a Hecke operator*

$$T_g : L^2(Z \backslash X) \rightarrow L^2(Z \backslash X)$$

by the formula

$$T_g(f(x)) = \sum_{i=1}^d f(g\delta_i x),$$

$$\forall f \in L^2(Z \backslash X), x \in X.$$

It is not hard to show that the operators  $T_g$  are well-defined; i.e. they map square-integrable functions to square-integrable functions. It is easy to turn the set of Hecke operators into a  $Z$ -module using the formula  $(mT_g)(f) = mT_g(f)$ , for  $f \in L^2(Z \backslash X)$ . We can also define a multiplication operation on this module in the following way. For  $g, h \in C_G(Z)$ , consider the coset decompositions

$$(4.5) \quad ZgZ = \bigcup_i Z\alpha_i \quad ZhZ = \bigcup_j Z\beta_j.$$

Then

$$(ZgZ) \cdot (ZhZ) = \bigcup_j ZgZ\beta_j = \bigcup_{ij} Z\alpha_i\beta_j = \bigcup_{Zw \subset ZgZhZ} Zw = \bigcup_{Zw \subset ZgZhZ} ZwZ.$$

The multiplication of Hecke operators for  $g, h \in C_G(Z)$  is then given by the formula

$$(4.6) \quad T_g T_h = \sum_{Zw \subset ZgZhZ} m(g, h, w) T_w,$$

where

$$m(g, h, w) = \# \left\{ i, j : Z\alpha_i\beta_j = Zw \right\},$$

where  $\alpha_i, \beta_j$  are given by (4.5). This multiplication is associative, and so we may put a ring structure on the set of Hecke operators per the following definition.

**Definition 4.11** (The Hecke ring, [20, Defn. 3.10.8, p. 76]). *Let  $G$  act continuously on a topological space  $X$  and let  $Z$  be a discrete subgroup of  $G$ . Fix any semigroup  $\Lambda$  such that  $Z \subset \Lambda \subset C_G(Z)$ . The Hecke ring  $\mathcal{H}_{Z,\Lambda}$  is defined as the set of all formal sums*

$$\sum_k c_k \mathbf{T}_{g_k}$$

with  $c_k \in \mathbf{Z}$ ,  $g_k \in \Lambda$ . Multiplication in the ring is induced by (4.6).

Following [20, §3.12], pass now to the specific case of  $Z = \Gamma = \mathrm{SL}(2, \mathbf{Z})$  and  $X = \mathfrak{H}$ . For integers  $n_0, n_1 \geq 1$ , the matrix  $\begin{pmatrix} n_0 n_1 & 0 \\ 0 & n_0 \end{pmatrix} \in C_G(Z)$ . Let  $\Lambda$  denote the semigroup generated by the matrices  $\begin{pmatrix} n_0 n_1 & 0 \\ 0 & n_0 \end{pmatrix}$  for  $n_0, n_1 \geq 1$ , together with the modular group  $\Gamma$ .

**Theorem 4.12.** *The Hecke ring  $\mathcal{H}_{\Gamma,\Lambda}$  is commutative.* ■

For each  $n \geq 1$ , we have a Hecke operator  $\mathbf{T}_n$ . It acts on the space of square-integrable automorphic forms  $f \in L^2(\Gamma \backslash \mathfrak{H})$ . The action of  $\mathbf{T}_n$  is given explicitly by the formula

$$\mathbf{T}_n f(z) = \frac{1}{\sqrt{n}} \sum_{\substack{ad=n \\ 0 \leq b < d}} f\left(\frac{az+b}{d}\right).$$

**Theorem 4.13.** *Let  $\langle \cdot, \cdot \rangle$  denote the  $L^2$  (Peterson) inner product on  $L^2(\Gamma \backslash \mathfrak{H})$ . The Hecke operators  $\mathbf{T}_n$  are self-adjoint with respect to  $\langle \cdot, \cdot \rangle$ ; i.e.*

$$\langle \mathbf{T}_n f, g \rangle = \langle f, \mathbf{T}_n g \rangle \quad \forall f, g \in L^2(\Gamma \backslash \mathfrak{H}). \quad \blacksquare$$

**Definition 4.14.** *Define the involution  $\mathbf{T}_{-1}$  on  $L^2(\Gamma \backslash \mathfrak{H})$  by*

$$\mathbf{T}_{-1} f\left(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\right) := f\left(\begin{pmatrix} y & -x \\ 0 & 1 \end{pmatrix}\right).$$

**Theorem 4.15** ([20, Theorem 3.12.6, p. 82]). *The Hecke operators  $\{\mathbf{T}_n\}_{n \in \mathbf{N}}$  commute with each other, with the operator  $\mathbf{T}_{-1}$ , and with the Laplacian  $\Delta = -y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$ .*

It follows in a standard way that the Hilbert space  $L^2(\Gamma \backslash \mathfrak{H})$  can be simultaneously diagonalized by the set of operators

$$\mathcal{T} := \{ \mathbf{T}_n : n = -1, n = 1, 2, \dots \} \cup \{ \Delta \}.$$

Therefore, we may consider Maaß forms that are simultaneous eigenfunctions of  $\mathcal{T}$ . We say such a form  $f$  is even if  $\mathbf{T}_{-1} f = f$ , odd if  $\mathbf{T}_{-1} f = -f$ . We let  $\lambda_n$  denote the eigenvalues of  $f$  with respect to the Hecke operators; i.e.

$$\mathbf{T}_n f = \lambda_n f.$$

In some sense, the ‘point’ of the Hecke operators is that they provide additional structure on the space of Maaß forms, and tell us something arithmetic about the Fourier coefficients of Hecke-Maaß forms. Namely, we have the following essential theorem.

**Theorem 4.16** (Multiplicativity of Fourier coefficients [20, Theorem 3.12.8, p.83]). *Consider*

$$f(z) = \sum_{n \neq 0} a(n) \sqrt{2\pi y} \cdot K_{\nu-\frac{1}{2}}(2\pi|n|y) \cdot e(nx),$$

*a Maaß form of type  $\nu$  as in Proposition 4.9 which is also an eigenfunction of all the Hecke operators. If  $a(1) = 0$ , then  $f \equiv 0$ . Assume  $f \neq 0$  is normalized so that  $a(1) = 1$ . Then*

$$T_n f = a(n) \cdot f \quad \forall n = 1, 2, \dots$$

*We have the following multiplicativity relations*

$$a(m)a(n) = a(mn) \quad \text{if } (m, n) = 1,$$

$$a(m)a(n) = \sum_{d|(m,n)} a\left(\frac{mn}{d^2}\right),$$

$$a(p^{r+1}) = a(p)a(p^r) - a(p^{r-1}) \quad \forall p \text{ prime, } \mathbf{Z} \ni r \geq 1. \quad \blacksquare$$

As already alluded to, this multiplicative structure is necessary to develop the Euler product associated to the L-function  $L(s, f)$ . (This statement is a theorem of Hecke.) The presence of the Hecke ring  $\mathcal{H}_{\Gamma, \Delta}$  also has allowed for the resolution of the quantum unique ergodicity conjecture in the arithmetic case; this is elaborated upon in §4.5. Put neatly, the Hecke operators can be considered ‘arithmetic symmetries.’

**4.2.3. Existence of Maaß forms & nodal domains.** Before moving on to a discussion of the Ramanujan conjecture in the next section, we stop to make a short detour to illustrate how different Maaß forms are from their holomorphic brethren. Among other things, while it is completely classical to write down cuspidal modular forms in terms of a  $q$ -expansion, up to now no one has found a single example of a Maaß form for  $\mathrm{PSL}(2, \mathbf{Z})$ , though, as discussed in §4.4, Selberg used the trace formula that bears his name to prove that there are actually infinitely many. Maaß [43] did write down some examples for discrete subgroups other than  $\mathrm{PSL}(2, \mathbf{Z})$ .

Another example of the difference in difficulty between the two kinds of cusp forms on  $\mathrm{GL}(2)$  is found in the nature of their zeros on  $\Gamma \backslash \mathfrak{H}$ . As a consequence of Duke’s equidistribution theorem [12], which we take up in detail in §4.6, the zeros of holomorphic cusp forms on the modular surface, which are topologically discrete points, become equidistributed in the appropriate limit. Since Maaß forms are actually waveforms, eigenfunctions on the modular surface, which has dimension two as Riemannian manifold, the nodes of a Maaß waveform  $f$  are actually a collection of nodal lines – classically, lines on the modular surface that are ‘motionless.’ As Ernst Chladni developed a technique of picturing these ‘nodal lines’ on a vibrating membrane, we can inquire after what the nodal lines of a Maaß wave eigenform look like on  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ ; i.e. we can try to make Chladni figures for Maaß forms. In some sense, this would be a ‘picture’ of a Maaß form.

In some sense, such a study would be a good study of the line between high-energy eigenstates in a classically chaotic system, and classical chaos. For, one might hope to be able to

study how a chaotic system passes from ‘classical chaos’ to ‘quantum chaos.’ One way to do this is to study high-energy eigenwaveforms. These are the Maaß forms on  $\Gamma \backslash \mathfrak{H}$ . The benefit of the arithmetic setting is that the additional number-theoretic symmetries afforded by the Hecke operators make studying and computing with high-energy eigenwaveforms possible (or at least easier).

The existence of such waveforms in a more general setting than the modular surface, say, the one considered previously where we have a quotient of the upper half-plane by a possibly nonarithmetic discrete cofinite Fuchsian group with exactly one cusp, is difficult to ascertain. There is a general philosophy due to Phillips and Sarnak, explicated in [48], which, inspired by work of Colin de Verdière [9] on deformations or perturbations of a hyperbolic surface, relates the existence of cusp forms to the critical zeros of the Selberg zeta-function  $Z_\Gamma(s)$  associated to a cofinite, possibly non-arithmetic, discrete group  $\Gamma$ , and in turn to the vanishing of an associated Rankin-Selberg L-function on the critical line. The Phillips-Sarnak theory would suggest that, in fact, there are only finitely many waveforms on a general hyperbolic surface, unless there is some additional structure and symmetry, such as is available in the case  $\Gamma = \mathrm{PSL}(2, \mathbf{Z})$ .

Putting aside this question, which is in itself interesting, we proceed to describe some results of Hejhal and Rackner [24] and Ghosh, Reznikov, and Sarnak [19] on nodal domains of Maaß waveforms in the arithmetic case of  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ . We refer the reader to [24, Fig. 6–8, pp. 285–287] for several remarkable ‘pictures’ of nodal domains attached to various Maaß waveforms for  $\mathrm{PSL}(2, \mathbf{Z})$ . The expectation is that the nodal lines will become increasingly ‘chaotic’ or complex as the energy (eigenvalue) of the waveform gets large. Indeed, an understanding of the interface between the classical and quantum chaos, and the distribution of mass associated to waveforms on a general hyperbolic surface, is the subject of so-called quantum ergodicity statements, taken up in §4.5.

We now summarize some quantitative results of Ghosh, Reznikov, and Sarnak [19]. Let  $\mathbf{X}$  denote the modular surface  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ . We consider even Maaß forms on  $\mathbf{X}$ , and write the eigenvalue associated to such a form  $f$  by  $\lambda_f = \frac{1}{4} + t_f^2$ . Let  $Z_f \subset \mathbf{X}$  denote the nodal line of  $f$  given by  $\{z \in \mathbf{X} : f(z) = 0\}$ . It consists of a finite union of real-analytic curves. The connected components of  $\mathbf{X} \setminus Z_f$  are the nodal domains of  $f$ ; we denote their number by  $\mathcal{N}_f$ . For any subset  $Y \subset \mathbf{X}$ , let  $\mathcal{N}_f^Y$  denote the number of nodal domains associated to  $f$  that have nonempty intersection with  $Y$ . Questions of interest to Ghosh-Reznikov-Sarnak include (a) estimating  $\mathcal{N}_f$  and  $\mathcal{N}_f^Y$ , and (b) determining whether  $Z_f$  is non-singular; i.e. whether or not the nodal line intersects itself.

Using a variety of heuristics arising from physical models, Bogolmony and Schmit conjecture [6] a precise asymptotic for  $\mathcal{N}_f$  in terms of the numbering of waveforms when ordered by eigenvalue/energy. Let  $\kappa_f$  denote the natural number associated to  $f$  by such an ordering.



**Conjecture 1** (Bogomolny-Schmit).

$$\mathcal{N}_f \sim \frac{2}{\pi}(3\sqrt{3} - 5)\kappa_f, \quad \kappa_f \rightarrow \infty.$$

Let  $\sigma$  denote the orientation-reversing isometry on  $\mathbf{X}$  induced by the reflection of  $\mathfrak{H}$  given by the involution  $\sigma(x + iy) = -x + iy$ . The isometry  $\sigma$  maps the nodal domains of  $f$  bijectively to themselves; a given nodal domain may either be unchanged after applying  $\sigma$ , or it could be different. Ghosh-Reznikov-Sarnak call a domain split or inert, respectively. We denote the number of inert nodal domains by  $\dot{\mathcal{N}}_f, \ddot{\mathcal{N}}_f$ , respectively. By definition, we have

$$\mathcal{N}_f = \dot{\mathcal{N}}_f + \ddot{\mathcal{N}}_f.$$

It is not known whether there are any split nodal domains. Let  $d^*z$  denote the arclength measure  $\frac{\sqrt{dx^2 + dy^2}}{y}$ . We now state some theorems of Ghosh-Reznikov-Sarnak.

**Theorem 4.17** ([19, Theorem 1.1]). *Let  $\mathcal{C}$  be a closed horocycle in  $\mathbf{X}$ . Then for  $\varepsilon > 0$*

$$t_f^{-\varepsilon} \ll_{\varepsilon} \int_{\mathcal{C}} f^2(z) d^*z \ll_{\varepsilon} t_f^{\varepsilon}. \quad \blacksquare$$

This theorem clearly implies that  $\mathcal{C}$  is not a part of  $Z_f$ . Let  $\delta = \{z \in \mathbf{X} : \sigma(z) = z\}$ . Then  $\delta$  is an arc composed of three piecewise-analytic geodesics, denoted by  $\delta_1, \delta_2, \delta_3$ . Let  $\delta_1$  be the sub-arc that is given by the projection to  $\mathbf{X}$  of  $\{z = x + iy \in \mathfrak{H} : x = 0, y \geq 1\}$ , and finally let  $\delta_2$  denote the sub-arc given by the projection of the geodesic  $\{z \in \mathfrak{H} : x = 1, y \geq \frac{\sqrt{3}}{2}\}$ .

**Theorem 4.18** ([19, Theorem 1.3]). *If  $\beta$  is a long enough but fixed compact subsegment in  $\delta_1$  or  $\delta_2$ , then for  $\varepsilon > 0$*

$$1 \ll_{\beta} \int_{\beta} f^2(z) d^*z \leq \int_{\delta} f^2(z) d^*z \ll_{\varepsilon} t_f^{\varepsilon}. \quad \blacksquare$$

**Theorem 4.19** ([19, Theorem 1.4]). *Let  $\mathcal{C}$  be a fixed closed horocycle in  $\mathbf{X}$ . Then for  $\varepsilon > 0$*

$$t_f^{\frac{1}{12} - \varepsilon} \ll_{\varepsilon} |Z_f \cap \mathcal{C}| \ll t_f. \quad \blacksquare$$

Note that the random wave model of Bogomolny and Schmit predicts

$$|Z_f \cap \mathcal{C}| \sim \frac{\text{length}(\mathcal{C})}{\pi} t_f.$$

**Theorem 4.20** ([19, Theorem 1.6]). *Fix  $\beta \subset \delta$  a sufficiently long compact geodesic segment on  $\delta_1$  or  $\delta_2$ , and assume the Lindelöf Hypothesis for the L-functions  $L(s, f)$ . Then*

$$|Z_f \cap \beta| \gg_{\varepsilon} t_f^{\frac{1}{12} - \varepsilon}. \quad \blacksquare$$

**Theorem 4.21** ([19, Theorem 1.8]). *With the same assumptions as in Theorem 4.20,*

$$\mathcal{N}_f^\beta \gg_\varepsilon t_f^{\frac{1}{12}-\varepsilon},$$

and in particular that  $\mathcal{N}_f^\beta$  goes to infinity as  $\kappa_f$  does. ■

The proofs of these theorems use the arithmetic quantum unique ergodicity theorem of Lindenstrauss [38] and Soundararajan [61].

**4.3. The Ramanujan Conjecture.** Srinivasa Ramanujan was interested in the arithmetic function  $\tau : \mathbf{N} \rightarrow \mathbf{Z}$  defined by the identity

$$\sum_{n \geq 1} \tau(n)q^n = q \prod_{n \geq 1} (1 - q^n)^{24} = \eta(z)^{24} = \Delta(z),$$

where  $q = e(z)$  with  $\Im z > 0$ ,  $\eta$  is the Dedekind eta function, and  $\Delta$  is the first cusp form; i.e. the unique weight 12 holomorphic cusp form for the full modular group. It is known as the discriminant modular form. From the modern viewpoint, Ramanujan was investigating the Fourier coefficients of a holomorphic cusp form. But, by emphasizing  $\tau$  as an arithmetic function, Ramanujan observed a pair of properties of  $\tau$  that prefigured the Hecke theory, and also conjectured that  $|\tau(p)| \leq 2p^{\frac{11}{2}}$  on primes  $p$ . This growth estimate is now known as the Ramanujan conjecture for holomorphic cusp forms. In the modern setting, we usually normalize the coefficients by taking out a factor of  $n^{(k-1)/2}$ ; denoting a normalized  $\tau$  function with  $\tilde{\tau} = \tau/n^{11/2}$ , Ramanujan's original conjecture would be simply  $|\tilde{\tau}(p)| \leq 2$ .

Let  $g \in S_k^*(\mathbf{N})$  be a holomorphic cuspidal newform of weight  $k$  and level  $\mathbf{N}$ .  $g$  admits a Fourier expansion

$$g(z) = \sum_{n > 0} a_n n^{(k-1)/2} e(nz).$$

It is now classical that  $a_n = n^{o(1)}$ , thanks to Deligne's proof of the Weil conjectures, including the Riemann hypothesis for function fields, and his earlier work to connect the Weil conjectures to the growth of Fourier coefficients of holomorphic cusp forms. Remarkably, the Satake parameters  $\alpha_p, \beta_p$  associated to the coefficient  $a_p$  for  $p$  prime by  $\alpha_p + \beta_p = a_p$  (the coefficients at primes determine the others completely via multiplicativity and the Hecke relations), have an interpretation as eigenvalues of Frobenius acting on  $\ell$ -adic cohomology groups. The statement  $|\alpha_p| = 1 = |\beta_p|$  (for unramified primes  $p$ ) then follows from the purity theorem for eigenvalues of Frobenius. In turn, this gives the Deligne divisor bound

$$a_n \leq d(n),$$

where  $d$ , the divisor function, counts the number of divisors of  $n$ . Note that  $d(n) = n^{o(1)}$ . This achieves the Ramanujan conjecture for these holomorphic cusp forms.

It would be reasonable to hope that the general growth bound  $a_n = n^{o(1)}$  holds more generally for (appropriately-normalized) automorphic forms on  $GL(2)$ . The appropriate generalization for other Lie groups, including groups of higher rank, was initially unclear; Satake proposed [54] such a generalization, but Kurokawa [34] and Howe and Piatetski-Shapiro [30] found that Satake's generalization did not hold. A subsequent refinement of the conjecture by Piatetski-Shapiro [49] is the current formulation of the generalized Ramanujan conjecture. For an excellent discussion of the generalized Ramanujan conjecture, see Sarnak [53].

A problem alluded to previously is establishing the automorphy of higher symmetric power L-functions  $L(s, \text{sym}^r f)$  for Maaß forms on  $GL(2)$ . Should the automorphy be obtained for all  $r \geq 1$ , the Ramanujan conjecture would follow. This is also known as establishing the functoriality of symmetric powers of automorphic representations of  $GL(2)$ , since, in loose terms, the principle of functoriality dictates that natural operations on automorphic L-functions carry over to operations on the side of automorphic representations. Establishing that taking symmetric powers on the level of L-functions corresponds to an automorphic object would be a functorial statement for symmetric powers of  $GL(2)$ . It is then perhaps no surprise that the current record for the best bound towards Ramanujan for  $GL(2)$  Maaß forms were obtained by Kim and Sarnak as a consequence of Kim's proof [33] of the automorphy of the symmetric fourth; see the second appendix. They prove that the Satake parameters  $\alpha_p, \beta_p$  associated to a Maaß form on  $GL(2)$  satisfy

$$|\alpha_p|, |\beta_p| \leq p^{\frac{7}{64}}.$$

Evidently, for those familiar with  $GL(2)$  L-theory, the bound on Fourier coefficients translates into a statement about the abscissa of absolute convergence of the Dirichlet series associated to  $L(s, f)$  for  $f$  a fixed Maaß cusp form.

**4.4. Weyl's Law.** As alluded to previously, no one has written down an example of a Maaß form for  $PSL(2, \mathbf{Z})$ . Any reasonable person might find this startling fact difficult to reconcile with how much has been proved about these forms. Indeed, one might expect that automorphic forms are too special to exist, and worry that all these results are vacuous! To the contrary, it is a classical theorem of Selberg that, in fact, there are *infinitely many* Maaß forms for  $SL(2, \mathbf{Z})$ . The gulf between this fact, proven using the Selberg trace formula, and the fact that no one has been able to write down a concrete example, indicates that it is likely that Maaß forms are quite curious creatures, and do not lend themselves to neat expression in terms of, say, a compact  $q$ -expansion, which is usually how modular forms are written down.

We refer in part to the exposition of Müller [47]. By way of introduction, let us first return to the case of a Riemann surface, taken up in §2, or, more generally, a smooth, compact Riemannian manifold  $M$  of dimension  $n$  with smooth, possibly empty boundary  $\partial M$ . The spectrum of the Laplacian on such an  $M$  is discrete, as we have already discussed in the case  $n = 2$ . The only accumulation point is  $\infty$ , and each eigenvalue occurs with finite multiplicity.

Ordering the eigenvalues of  $\Delta$  on  $M$  like

$$0 \leq \lambda_0 \leq \lambda_1 \leq \cdots \rightarrow \infty,$$

we can count the number of eigenvalues less than some number  $R$  with multiplicity. Let

$$\mathcal{N}(R) := \# \left\{ j : \sqrt{\lambda_j} \leq R \right\}$$

denote such a count. An obvious fundamental question is then, how does  $\mathcal{N}(R)$  grow with  $R$ ? The Weyl law in this setting states a precise asymptotic.

$$\mathcal{N}(R) \sim \frac{\text{vol}(M)}{(4\pi)^{n/2} \Gamma\left(\frac{n}{2} + 1\right)} R^n.$$

It was proved by Weyl [68] in the case of a bounded domain in  $\mathbf{R}^3$ , Raleigh [50] for a cube, Garding [16] for a general elliptic operator on a domain in  $\mathbf{R}^n$ , and by Minakshisundaram and Pleijel [46] in the case of a closed Riemannian manifold. A natural question then follows: what about the lower order terms in an asymptotic expansion of  $\mathcal{N}(R)$ ? To address this, form the remainder term

$$\mathcal{R}(R) := \mathcal{N}(R) - \frac{\text{vol}(M)}{(4\pi)^{n/2} \Gamma\left(\frac{n}{2} + 1\right)} R^n.$$

We then ask, what is the optimal bound on the remainder? For a closed Riemannian manifold, Avakumović [4] proved the Weyl law with a power savings in the error term, and showed this bound of  $O(R^{n-1})$  for the error was optimal for the sphere. This result was subsequently extended by Hörmander [29]. Weyl [69] then conjectured the exact form of a next-to-leading-order term in the case of a bounded domain  $\Omega \subset \mathbf{R}^3$ ; precisely, that

$$\mathcal{N}(R) = \frac{\text{vol}(\Omega)}{6\pi^2} R^3 - \frac{\text{vol}(\partial \Omega)}{16\pi} R^2 + o(R^2), \quad R \rightarrow \infty.$$

This statement was proved by Ivrii [31] and Melrose [45] for manifolds with boundary assuming a certain condition on the period billiard trajectories.

We now pass to considering the analogous question in the case of  $X = \text{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ . Using the trace formula he developed, Selberg [55, (9.33), p.668] proved the following version of Weyl's law for an arbitrary lattice  $\Gamma$  in  $\text{SL}(2, \mathbf{R})$ .

$$\mathcal{N}_\Gamma(R) + \mathcal{M}_\Gamma(R) \sim \frac{\text{area}(\Gamma \backslash \mathfrak{H})}{4\pi} R^2, \quad R \rightarrow \infty.$$

Here,  $\mathcal{N}_\Gamma(R)$  counts the number of eigenvalues less than  $R$  with multiplicity, while  $\mathcal{M}_\Gamma$  measures the contribution from the continuous spectrum. More precisely, we have the statement already stated in Theorem 3.19 (i). Combining it with item (ii) of the same theorem, we may conclude

$$(4.7) \quad \# \{0 \leq r_n \leq R\} + \frac{1}{4\pi} \int_{-R}^R \frac{\phi'}{\phi} \left(\frac{1}{2} + ir\right) dr \sim \frac{\mu(F)}{4\pi} R^2 \quad R \rightarrow \infty.$$

Here,  $\phi$  is as given in (3.13). Note that this statement is valid in the case of a *singular*  $\chi = 1$ . This is precisely (up to normalization) Selberg's statement [55, (9.33), p. 668].

The basic issue with the Weyl law for  $\Gamma \backslash \mathfrak{H}$  as stated in (4.7) is that the contribution from the discrete spectrum is not isolated. So, we may have that  $\mathcal{N}_\Gamma(R) \ll 1$ , and all the contribution comes from  $\mathcal{M}_\Gamma(R)$ . Quoting Selberg [55, p. 668], only in some special cases can the function  $\phi(s)$  'be expressed in terms of functions that are known from analytic number theory,' and in every one of these special cases,  $\mathcal{M}_\Gamma(R) \ll R \log R$ . Crucially for our interests, these 'special cases' include the case of  $\mathrm{PSL}(2, \mathbf{Z})$  and congruence subgroups. This shows that there are infinitely many Maaß forms for  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$  and the arithmetic congruence subgroups we care about.

Weyl's law for  $\Gamma \backslash \mathfrak{H}$ , given in (4.7), constitutes an application that demonstrates the phenomenal power of the Selberg trace formula.

Though we *can* separate out the contribution from the continuous spectrum and show that it does not contribute to leading order in the arithmetic cases we care about, a natural question is, could we somehow *deform* the arithmetic manifolds we care about, and does the Weyl law still hold in such a 'deformed' setting? Any such investigation would be directly motivated by the pioneering work of Phillips and Sarnak [48] as discussed in §4.2.3, who study the real-analytic deformation of discrete groups in  $\mathrm{PSL}(2, \mathbf{R})$ . The general Phillips-Sarnak philosophy is that the arithmetic structure present, in, for example, the Hecke structure on  $\Gamma \backslash \mathfrak{H}$  that we discussed in §4.2.2, is *delicate*, in some rigorous sense. That is to say, it does not survive under deformation. Following Phillips-Sarnak, this structure is connected to the zeros of the Selberg zeta-function along the critical line, which is in turn related to the vanishing of an associated Rankin-Selberg L-function at these same points. If these quantities varied continuously under deformation of the underlying manifold, then the special properties we observe in the arithmetic case would be expected to break under deformation. This is essentially what Luo [40] shows. Luo makes the notion of deformation rigorous in the following sense. Let  $Q \in S_4(\Gamma_0(p))$  be a fixed newform, where  $\Gamma_0(p)$  is the usual congruence subgroup and  $p$  is prime. And let  $T(\Gamma_0(p))$  be the associated Teichmüller space. Then, he considers the deformation  $\Gamma_\tau$  generated by the quadratic differential  $Q(z) dz^2$  at  $\Gamma_0(p)$ . By showing that a positive proportion of Rankin-Selberg L-functions *do not vanish* at critical zeros obtained from eigenvalues attached to an orthonormal basis of Hecke-Maaß cusp forms for the full modular group  $\Gamma_0(1)$ , Luo is able to prove in his Theorem 2 that the Weyl law

$$\mathcal{N}_{\Gamma_\tau}(R) \sim \frac{\mathrm{area}(\Gamma_\tau \backslash \mathfrak{H})}{4\pi} R^2$$

*fails* for generic deformations  $\Gamma_\tau$ , under the assumption that the eigenvalue multiplicities of the Laplacian on  $\Gamma_0(p) \backslash \mathfrak{H}$  are bounded (or a slightly weaker technical assumption). This remarkable result indicates that the arithmetic structure on  $\Gamma_0(N) \backslash \mathfrak{H}$  is *delicate* in a rigorous sense.

**4.5. Quantum Ergodicity.** In §4.2.3, we broached some of the remarkable properties of Maaß forms on  $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ . In particular, the nodal lines of Maaß forms suggested some kind of chaotic behavior as the energy of the waveform gets large. We refer the reader to [41, 52, 62] for some critical developments in this area that inspire the following exposition.

This section is devoted to the phenomenon of so-called quantum chaos or quantum ergodicity. The study of quantum ergodicity is motivated by a desire to probe the boundary between classical and quantum models for a chaotic (i.e. regular or ergodic) system. In particular, the goal is to understand how a quantum understanding of such a system is reflected in how the system appears from a classical perspective.

The above discussion seems like it might be of interest to physicists, but what about it would interest the number theorist? The notion of arithmetic quantum chaos or arithmetic quantum ergodicity, can be rephrased as a desire to understand the distribution of  $(L^2)$  mass of eigenfunctions of the Laplacian on arithmetic manifolds. We pause to compare the two cases, following [52].

Suppose  $\Sigma_g = \pi_1(\Sigma_g) \backslash \mathfrak{H}$  is a compact hyperbolic surface, where  $\pi_1(\Sigma_g) \subset \mathrm{PSL}(2, \mathbf{R})$  is discrete and cocompact. The geodesic flow on the unit cotangent bundle is generated by a Hamiltonian and is known to be ergodic, Anosov, and appears chaotic. The quantization of this Hamiltonian is the Laplace-Beltrami operator on  $\Sigma_g$ , which we denote  $\Delta$ . Let  $\phi_j$  denote the eigenfunctions (Maaß forms) on  $\Sigma_g$ . It would be natural to try to describe the distributional properties of the  $\phi_j$ . Heller [25] found that, in fact, there was some structure associated to these waveforms; there was a concentration of  $L^2$  mass in certain states related to a finite union of periodic, unstable orbits. He called this phenomenon ‘scarring.’

In an influential paper, Rudnick and Sarnak [52] claim that this ‘scarring’ is not generic. More precisely, they define the probability measures  $\mu_j$  on  $\Sigma_g$  by

$$d\mu_j = \left| \phi_j(z) \right|^2 d\mathrm{vol}(z).$$

These probability measures have the quantum-mechanical interpretation the probability densities of finding a particle in the state  $\mu_j$  at the the point  $z$ . Rudnick and Sarnak then conjecture

**Conjecture 2** (Quantum Unique Ergodicity). *Let  $X$  be a compact manifold of negative curvature. Then the measures  $\mu_j$  converge to  $d\mathrm{vol}$ .*

The strength of this conjecture is that there is no need to pass to a subsequence of measures  $\mu_{j_k}$ . Rudnick-Sarnak conjecture that there is *no* exceptional subsequence of measures. As they explain, the physical interpretation is remarkable; it is that ‘at the quantum level and in the semi-classical limit, there is little manifestation of chaos’ ([52, p. 196]). The uniqueness of the limit implies that there is only one quantum limit, while classical unique ergodicity, which is the uniqueness of the invariant measure for the Hamiltonian flow, is never satisfied for chaotic systems.

At the time that Rudnick and Sarnak conjectured quantum unique ergodicity, Schnirelman [60], Colin de Verdière [10], and Zelditch [72] had proven that the quantum analogue of the geodesic flow on a compact Riemannian manifold  $Y$  is ergodic. More specifically, they localize the measures  $\mu_j$  to  $S_1^*(Y)$ , the unit cotangent bundle, and show that if the geodesic flow on  $S_1^*(Y)$  is ergodic, there is a full-density subsequence  $\lambda_{j_k}$  (where  $\lambda_j$  are the (ordered) eigenvalues of the spectrum of  $\Delta$  on  $Y$ ) for which  $\mu_{j_k}(A) \rightarrow \text{vol}(A)/\text{vol}(Y)$  for all sufficiently ‘nice’ sets  $A$  such as geodesic balls. (A full-density subsequence is one satisfying  $\sum_{\lambda_{j_k} \leq \lambda} 1 \sim \sum_{\lambda_j \leq \lambda} 1$ .) Zelditch [73] then extended this result to some noncompact surfaces such as  $\mathbf{X} = \text{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ . As described by [41], he shows that if  $b \in C_{00}^\infty(Y)$  and  $\int_{\mathbf{X}} b(x) \, d\text{vol}(x) = 0$ , then

$$\sum_{\lambda_j \leq \lambda} \left| \langle b, \mu_j \rangle \right|^2 \ll_b \frac{\lambda}{\log \lambda}.$$

Combining this with Selberg’s result [55] that  $\sum_{\lambda_j \leq \lambda} 1 \sim \frac{\lambda}{12}$  establishes quantum ergodicity in this setting. The startling nature of the quantum unique ergodicity conjecture is that it maintains there is no need to pass to a subsequence.

In 1995, Luo and Sarnak [41] proved an analogue of quantum unique ergodicity for Eisenstein series on  $\mathbf{X}$ . That is, in analogy with the definition of the  $\mu_j$ , they formed the natural measures<sup>2</sup>

$$\mu_t := \left| E \left( z, \frac{1}{2} + it \right) \right|^2 \, d\text{vol}(z),$$

and then showed that these ‘Eisenstein’ measures become individually distributed in the full-sequence limit.

**Theorem 4.22** ([41, Theorem 1.1, p. 208]). *Let  $A, B$  be compact Jordan-measurable subsets of  $\mathbf{X}$ . Then*

$$\lim_{t \rightarrow \infty} \frac{\mu_t(A)}{\mu_t(B)} = \frac{\text{vol}(A)}{\text{vol}(B)}. \quad \blacksquare$$

In the general setting, QUE remains open, though Anantharaman [1] has showed that any limiting measure must have positive entropy. In the case  $\text{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ , however, due to work of Lindenstrauss [38] and Soundararajan [61], the conjecture is proved. We follow the exposition of [62].

Lindenstrauss’s approach to the problem is to consider microlocalizations of the Hecke-Maaß form measures  $\mu_j$  to  $S_1^*(\mathbf{X}) = \text{SL}(2, \mathbf{Z}) \backslash \text{SL}(2, \mathbf{R})$ ; these lifts are approximately invariant under the geodesic flow from the results of Schnirelman [60], Colin de Verdière [10], and Zelditch [72], as described to above. Lindenstrauss’s major contribution is to use results from measure rigidity to show that the only possible limiting measures are of the form

<sup>2</sup>Note that  $\mu_t(\mathbf{X}) = \infty$ , so there is no natural normalization.

$\frac{3}{\pi}c \frac{dx dy}{y^2}$ , where  $0 \leq c \leq 1$ . This shows that the measures become equidistributed unless some of the  $L^2$  mass ‘escapes’ into the cusp. Soundararajan’s major contribution was to show that such ‘escape’ was impossible, proving  $c = 1$  in the statement above.

**Theorem 4.23** ([62, Theorem 1, p. 359]). *For any sequence of  $L^2$ -normalized Hecke-Maaß eigenforms  $\phi_j$ , the measures  $|\phi_j|^2 \frac{dx dy}{y^2}$  tend weakly to the measure  $\frac{3}{\pi} \frac{dx dy}{y^2}$  as  $\lambda_j \rightarrow \infty$ . ■*

It is important to mention that the arithmetic structure afforded by the presence of the Hecke ring on  $\mathbf{X}$  was a crucial ingredient in the proof of Theorem 4.23. This is a sterling example of the value of the Hecke structure alluded to in previous sections. Subsequently, Holowinsky and Soundararajan [28] established a natural analogue of QUE for holomorphic Hecke eigencuspforms.

A subconvexity bound for a particular degree-six L-function implies Theorem 4.23 via Watson’s explicit triple-product formula [67]. This subconvexity bound follows in turn from the Lindelöf Hypothesis, which is a consequence of the Generalized Riemann Hypothesis. See Zhao [75] for more details. In that paper, Zhao studies the quantum variance as another angle on QUE. More particularly, he considers the sum

$$\sum_{\lambda_j \leq \lambda} \left| \mu_j(\Psi) \right|^2,$$

where  $\Psi \in C_{00}^\infty(\mathbf{X})$ . First introduced by Zelditch [74], the quantum variance describes the variation in a quantum observable  $\langle \text{Op}(\Psi)\mu_j, \mu_j \rangle =: \mu_j(\Psi)$ , where  $\mu_j$  is the measure formed from a Hecke-Maaß form on  $\mathbf{X}$ . The quantum variance is related to the rate of convergence to the limiting measure  $\frac{3}{\pi} \frac{dx dy}{y^2}$  in Theorem 4.23. Luo and Sarnak [42] consider the analogous question for holomorphic cusp forms on  $\mathbf{X}$ . As a striking consequence, it can be shown that the zeros of holomorphic Hecke eigencuspforms of weight  $k$  become equidistributed on the modular surface  $\mathbf{X}$  as  $k \rightarrow \infty$ ; see [51].

**4.6. Duke’s Theorem.** In this final section, the author indulges in discussing a favorite theorem of his. The setting is again the modular surface  $\mathbf{X} = \text{PSL}(2, \mathbf{Z}) \backslash \mathfrak{H}$ . We partially follow the exposition of [13], who give an alternate proof of Duke’s theorem using ergodic theory. We describe Duke’s original proof, which is closer in spirit to the focus of this exposition, which is the spectral theory of automorphic forms and harmonic analysis.

Duke’s theorem [12, Theorem 1] is a statement about distribution of Heegner points and geodesics on  $\mathbf{X}$ . We reproduce some of his background from [12, p. 75]. Denote by  $h(d)$  the number of proper  $(\text{SL}(2, \mathbf{Z})$ -inequivalent) classes of primitive irreducible integral binary quadratic forms  $Q = Q(x, y) = ax^2 + bxy + cy^2$  with discriminant  $d = b^2 - 4ac$ .  $h(d)$  is the class number. If  $d > 0$ , Pell’s equation is  $x^2 - dy^2 = 4$ . Let  $(x_d, y_d)$  with  $x_d, y_d > 0$  be the fundamental solution and set  $\varepsilon_d = (x_d + \sqrt{d}y_d)/2$ . Let  $\Gamma = \text{PSL}(2, \mathbf{Z})$  and let  $F$  denote the



fundamental domain for  $\Gamma \backslash \mathfrak{H}$ . If  $d < 0$ , the  $h(d)$  Heegner points are given by

$$\Lambda_d = \left\{ z_Q = \frac{-b + \sqrt{d}}{2a} : b^2 - 4ac = d, z_Q \in \mathbb{F} \right\}.$$

If  $d < 0$  is fundamental, then these points correspond to ideal classes in  $\mathbf{Q}(\sqrt{d})$ . If  $d > 0$ , the points  $(-b \pm \sqrt{d})/2a$  specify the endpoints of a geodesic in  $\mathfrak{H}$  with respect to the usual hyperbolic metric  $ds^2 = (dx^2 + dy^2)/y^2$ ; these endpoints induce a unique primitive, positively-oriented closed geodesic in  $\Gamma \backslash \mathfrak{H}$  of length  $\log \varepsilon_d$  or  $2 \log \varepsilon_d$ , according as whether  $Q$  is or is not equivalent to  $-Q$ . If  $d > 0$ , let  $\Lambda_d$  denote the set of all such distinct geodesics. The total length of geodesics in  $\Lambda_d$  is  $h(d) \log \varepsilon_d$ , and every primitive, positive-oriented closed geodesic in  $\Gamma \backslash \mathfrak{H}$  occurs in exactly one  $\Lambda_d$ . We recall that  $d \in \mathbf{Z}$  is a fundamental discriminant if it is  $\equiv 1 \pmod{4}$  and squarefree or equal to  $4m$ , where  $m \equiv 2$  or  $3 \pmod{4}$  is squarefree. Let  $d\mu(z) = \frac{3}{\pi} \frac{dx dy}{y^2}$  be the familiar measure, normalized so that  $\mu(\mathbb{F}) = 1$ . Duke proves

**Theorem 4.24** ([12, Theorem 1, p. 75]). *Suppose  $d$  is a fundamental discriminant and  $\Omega \subset \mathbb{F}$  is convex (in the hyperbolic sense) and  $\partial\Omega$  is piecewise-smooth. Then for some  $\delta > 0$  depending only on  $\Omega$*

- (i)  $\frac{\#\Lambda_d \cap \Omega}{\#\Lambda_d} = \mu(\Omega) + O(|d|^{-\delta})$  as  $d \rightarrow -\infty$ , and
- (ii)  $\frac{\sum_{\mathcal{C} \in \Lambda_d} |\mathcal{C} \cap \Omega|}{\sum_{\mathcal{C} \in \Lambda_d} |\mathcal{C}|} = \mu(\Omega) + O(d^{-\delta})$  as  $d \rightarrow +\infty$

where  $|\mathcal{C}|$  is the (non-Euclidean) length of  $\mathcal{C}$  and the implied constants depend only on  $\delta$  and  $\Omega$ , though ineffectively.

Part (i) with error term  $O(\log^{-A} |d|)$  for some  $A > 0$  with additional conditions on the fundamental discriminant was proved by Linnik [39] using ergodic methods. Log savings are common in theorems proved using ergodic techniques, though power savings like those Duke obtains are obviously preferable. Chelluri [8] subsequently extended Theorem 4.24 to the unit cotangent bundle  $S_1^*(\mathbf{X})$ ; he obtained the following. There is a geodesic orbit associated to each geodesic class in  $\Lambda_d$ , which can be lifted to the unit tangent bundle of  $\mathfrak{H}$  and then projected to a geodesic orbit on  $S_1^*(\mathbf{X})$ . Let  $\mathcal{G}_d$  denote the image of  $\Lambda_d$  on  $S_1^*(\mathbf{X})$ .  $\mathcal{G}_d$  is a collection of compact orbits of the geodesic flow, and it carries a natural probability measure  $\mu_d$  that is invariant under this geodesic flow. Chelluri proves the following extension of Theorem 4.24.

**Theorem 4.25** ([8]). *As  $d \rightarrow +\infty$  amongst the positive fundamental discriminants, the set  $\mathcal{G}_d$  becomes equidistributed with respect to the Liouville (Haar) probability measure  $\mu_{\mathbf{L}}$  on  $S_1^*(\mathbf{X})$ : for any*

$\Psi \in C_{00}^\infty(S_1^*(\mathbf{X}))$ ,

$$\int_{\Lambda_d} \Psi(t) \, d\mu_d(t) \rightarrow \int_{S_1^*(\mathbf{X})} \Psi(u) \, d\mu_L(u). \quad \blacksquare$$

Duke's theorem, Theorem 4.24, is a beautiful statement about the equidistribution of closed geodesics on  $\mathbf{X}$ . Note that if Gauss' conjecture that  $h(d) = 1$  for infinitely many fundamental  $d > 0$  holds, then by (ii) of Theorem 4.24, we have that in fact *individual* geodesics become equidistributed in the limit of large  $d$ .

The proof of Theorem 4.24 is a beautiful application of Selberg's spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H})$ , Weyl's equidistribution criterion, and a subconvexity estimate for Fourier coefficients of particular Maaß forms of half-integral weight. We sketch it.

*Sketch of proof of Theorem 4.24.* By Weyl's equidistribution criterion and the spectral decomposition of  $L^2(\Gamma \backslash \mathfrak{H})$ , the theorem is proved if the following two 'Weyl sums' can be shown to decay in  $|d|$ . The sums are

$$W_{\text{Eis}}(d, t) = \begin{cases} \frac{1}{\#\Lambda_d} \sum_{z \in \Lambda_d} \mathbf{E}\left(z, \frac{1}{2} + it\right) & d < 0 \\ \frac{1}{\sum |\mathcal{C}|} \sum_{\mathcal{C} \in \Lambda_d} \int_{\mathcal{C}} \mathbf{E}\left(z, \frac{1}{2} + it\right) \, ds & d > 0, \end{cases}$$

and

$$W_{\text{cusp}}(d, t) = \begin{cases} \frac{1}{\#\Lambda_d} \sum_{z \in \Lambda_d} u(z) & d < 0 \\ \frac{1}{\sum |\mathcal{C}|} \sum_{\mathcal{C} \in \Lambda_d} \int_{\mathcal{C}} u(x) \, ds & d > 0. \end{cases}$$

Here,  $\mathbf{E}(z, s)$  is an Eisenstein series, and  $u(z)$  is a certain automorphic eigenfunction of the Laplace-Beltrami operator on a matrix space. There is a theta lift of  $u$  due to Maaß that yields a Maaß form of half-integral weight. Let  $\theta(z)$  denote Siegel's theta function for an indefinite quadratic form, as in [12, Theorem 3, p. 81].

**Theorem 4.26** ([12, Theorem 4, p. 84]). *Let  $S[x]$  be an integral ternary quadratic form of signature  $(1, 2)$ . There is a subset of the integers  $\mathbf{D}_S \subset \mathbf{Z}$  depending on  $S$  such that, for  $f(z) = y^{3/4} \langle u(\cdot), \theta(z, \cdot) \rangle$  and  $d \in \mathbf{D}_S$ , the  $d$ th Fourier coefficient of  $f(z)$  (at the cusp) is given by*

$$\zeta(d) \frac{\pi^{-\text{sgn}(d)/4}}{\sqrt{2}} |d|^{-3/4} M_u(d),$$

where

$$M_u(d) = \begin{cases} \sum'_{z \in \Lambda_d^+} u(z) & d < 0 \\ \sum_{\mathcal{C} \in \Lambda_d^+} \int_{\mathcal{C}} u(z) \, ds & d > 0. \end{cases}$$

Here,  $\Lambda_d^+$  is a certain collection of points in  $\Gamma \backslash \mathfrak{H}$ . The tick in the first sum indicates that  $u(z)$  is divided by the order of the stabilizer of  $z$  in  $\Gamma$ . ■

The point is that we can now use a subconvexity bound for Fourier coefficients of Maaß forms that Duke proves using an estimate of Iwaniec for a certain sum of Kloosterman sums over varying levels together with Proskurin's generalization of the Kuznetsov sum formula. Concretely, he proves

**Theorem 4.27** ([12, Theorem 5, p. 85]). *Let  $\{\zeta(n)\}$  be the Fourier coefficients of a spectral Maaß form  $f(z)$  of weight  $k = 1/2 + \ell$  and (even) discriminant  $D$  for  $\Gamma_0(N)$ , where  $\ell \in \mathbf{Z}$  and  $N \equiv 0 \pmod{D}$ , with eigenvalue  $\lambda = 1/4 + t^2$ . We have the estimate*

$$\zeta(n) \ll_{k,D,\varepsilon} |\lambda|^A \cosh(\pi t/2) |n|^{-2/7+\varepsilon} \quad \text{as } |n| \rightarrow \infty,$$

provide  $n$  is squarefree or a fundamental discriminant. We may take  $A = 5/4 - k/4 \operatorname{sgn}(n)$ . ■

Combining this estimate for Fourier coefficients of half-integral weight Maaß forms with Maaß's explicit theta correspondance, we can estimate the Weyl sums  $W_{\text{cusp}}(d, t)$ . We also need Siegel's (ineffective) estimate, in the forms

$$\begin{aligned} \#\Lambda_d &\gg_{\varepsilon} |d|^{1/2-\varepsilon} \quad \text{as } d \rightarrow -\infty, \text{ and} \\ \sum_{\mathcal{C} \in \Lambda_d} |\mathcal{C}| &\gg_{\varepsilon} |d|^{1/2-\varepsilon} \quad \text{as } d \rightarrow +\infty. \end{aligned}$$

Combining these estimates, Duke shows for  $d$  a fundamental discriminant,

$$W_{\text{Eis}}(d, t) \ll_{\varepsilon} |t|^A |d|^{-1/4+\varepsilon} L\left(\frac{1}{2} + it, \chi_d\right) \quad \text{as } |d| \rightarrow \infty,$$

where  $A > 0$  is a constant that may change from line to line. Theorem 4.27 then yields the estimate

$$W_{\text{Eis}}(d, t) \ll_{\varepsilon} |t|^A |d|^{-1/28+\varepsilon}.$$

On the cuspidal side, combining the above theorems and estimates, we have for fundamental  $d$

$$W_{\text{cusp}}(d, t) \ll_{\varepsilon} |t|^A |d|^{-1/28+\varepsilon}.$$

The theorem now follows in a standard way from the Weyl equidistribution criterion. ■

## REFERENCES

- [1] N. Anantharaman, *Entropy and localization of eigenfunctions*, Ann. of Math. (2) **168** (2008), 435–475.
- [2] J. Arthur, *Eisenstein series and the trace formula*, (Oregon State University, Corvallis, OR, 1977), Automorphic forms, representations, and L-functions. 1, Proc. Sympos. Pure Math., vol. XXXIII, Amer. Math. Soc., Providence, RI, 1979, pp. 253–274.
- [3] ———, *The work of Ngô Bao Châu*, Proc. International Math. Congr. Hyderabad (2010), 1–14.
- [4] V. G. Avakumović, *Über die Eigenfunktionen auf geschlossenen Riemannschen Mannigfaltigkeiten*, Math. Z. **65** (1956), 327–344.
- [5] V. Blomer, *Applications of the Kuznetsov formula on  $GL(3)$* , Invent. math. **194** (2013), 673–729.
- [6] E. Bogomolny and C. Schmit, *Percolation model for nodal domains of chaotic wave functions.*, Physical Review Letters **88** (2002).
- [7] D. Bump, *Spectral theory of  $\Gamma \backslash SL(2, \mathbf{R})$*  (2001), lecture notes.

- [8] T. Chelluri, *Equidistribution of roots of quadratic congruences*, 2004, Rutgers The State University of New Jersey, New Brunswick, Ph.D. Thesis.
- [9] Y. Colin de Verdière, *Pseudo-Laplacians*. I **32**, 275–286; II **33**, 87–113, *Ann. Inst. Fourier* (1983).
- [10] ———, *Ergodicité et fonctions propre du laplacian*, *Commun. Math. Phys.* **102**, 497–502.
- [11] P. Deligne, *La conjecture de Weil II*, *Inst. Haute Études Sci. Publ. Math* **52** (1980), 137–252.
- [12] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, *Invent. math.* **92** (1988), 73–90.
- [13] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, *The distributino of closed geodesics on the modular surface, and Duke’s theorem*, *L’Enseignement Mathématique* (2) **58** (2012), 249–313.
- [14] N. Elkies, *The Riemann zeta function and its functional equation (and a review of the Gamma function and Poisson summation)*, lecture notes, available at [www.math.harvard.edu/~elkies/M259.02/zeta1.pdf](http://www.math.harvard.edu/~elkies/M259.02/zeta1.pdf).
- [15] P. Garabedian, *Partial Differential Equations*, John Wiley, 1964.
- [16] L. Garding, *Dirichlet’s problem for linear elliptic partial differential equations*, *Math. Scand.* **1** (1953), 55–72.
- [17] P. Garrett, *Continuous spectrum for  $SL_2(\mathbf{Z}) \backslash \mathfrak{H}$*  (2011), lecture notes, available at [http://www.math.umn.edu/~garrett/m/mfms/notes\\_c/cont\\_afc\\_spec.pdf](http://www.math.umn.edu/~garrett/m/mfms/notes_c/cont_afc_spec.pdf).
- [18] ———, *Continuous spectrum for  $SL_2(\mathbf{Z}) \backslash \mathfrak{H}$*  (2014), lecture notes, available at [http://www.math.umn.edu/~garrett/m/mfms/notes\\_2013-14/13\\_1\\_cont\\_afc\\_spec.pdf](http://www.math.umn.edu/~garrett/m/mfms/notes_2013-14/13_1_cont_afc_spec.pdf).
- [19] A. Ghosh, A. Reznikov, and P. Sarnak, *Nodal Domains of Maass Forms I*, *Geom. Funct. Anal.* **23** (2013), 1515–1568.
- [20] D. Goldfeld, *Automorphic Forms and L-Functions for the Group  $GL(n, \mathbf{R})$* , Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, UK, 2006.
- [21] G. Halász, *Selberg’s trace-formula: a generalization of the Poisson summation formula*, lecture notes, available at [www.math-inst.hu/~major/articles/selberg1.pdf](http://www.math-inst.hu/~major/articles/selberg1.pdf).
- [22] D. Hejhal, *The Selberg Trace Formula for  $PSL(2, \mathbf{R})$* , Lecture Notes in Mathematics, vol. 1, Springer-Verlag, Berlin-Heidelberg-New York, 1976.
- [23] ———, *The Selberg Trace Formula for  $PSL(2, \mathbf{R})$* , Lecture Notes in Mathematics, vol. 2, Springer-Verlag, Berlin-Heidelberg-New York, 1983.
- [24] D. Hejhal and B. Rackner, *On the Topography of Maass Waveforms for  $PSL(2, \mathbf{Z})$* , *Exper. Math.* **1** (1992), no. 4.
- [25] E. J. Heller, *Chaos and Quantum Physics* (Les Houches, 1989) (M. J. Giannoni, A. Voros, and J. Zinn-Justin, eds.), Amsterdam: North-Holland, 1991, pp. 549–661.
- [26] E. Hewitt and K. Ross, *Abstract Harmonic Analysis*, Vol. 1, Springer-Verlag, Berlin, 1963.
- [27] D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, B.G. Teubner, Leipzig, 1912.
- [28] R. Holowinsky and K. Soundararajan, *Mass equidistribution for Hecke eigenforms*, *Ann. of Math.* (2) **172** (2010), 1517–1528.
- [29] L. Hörmander, *The spectral function of an elliptic operator*, *Acta Math.* **121** (1968), 193–218.
- [30] R. Howe and I. I. Piatetski-Shapiro, *A counterexample to the “generalized Ramanujan conjecture” for (quasi-) split groups*, (Oregon State Univ., Corvallis, Ore., 1977), Automorphic forms, representations and L-functions. 1 (A. Borel and W. Casselman, eds.), *Proc. Sympos. Pure Math.*, vol. XXXIII, Amer. Math. Soc., Providence, RI, 1979, pp. 315–322.
- [31] V. Ivrii, *The second term of the spectral asymptotics for a Laplace-Beltrami operator on manifolds with boundary*, *Funktional. Anal. i Prilozhen.* **14** (1980), no. 2, 24–34.
- [32] H. Iwaniec, *Spectral Methods of Automorphic Forms*, 2nd ed., Graduate Studies in Mathematics, vol. 53, Amer. Math. Soc. and Revista Matemática Iberoamericana, Providence, R.I. and Madrid, 2002.
- [33] H. Kim, *Functoriality for the exterior square of  $GL_4$  and symmetric fourth of  $GL_2$* , with an appendix by D. Ramakrishnan and an appendix by H. Kim and P. Sarnak, *J. Amer. Math. Soc.* **16** (2002), no. 1, 139–183.
- [34] N. Kurokawa, *Examples of eigenvalues of Hecke operators on Siegel cusp forms of degree two*, *Invent. Math.* **49** (1978), no. 2, 149–165.

- [35] E. Lapid, *Introductory notes on the trace formula*, Automorphic Forms and the Langlands Program (L. Ji, K. Liu, S.-T. Yau, and Z.-J. Zheng, eds.), Advanced Lectures in Mathematics, vol. IX, 2010, pp. 135–175.
- [36] R. P. Langlands, *Eisenstein series*, Algebraic Groups and Discontinuous Subgroups, Proc. Symp. Pure Math., vol. IX, Amer. Math. Soc., Providence, RI, 1966, pp. 235–252.
- [37] ———, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Mathematics, Springer-Verlag, 1976.
- [38] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (1) **163** (2006), 165–219.
- [39] Y. V. Linnik, *Ergodic properties of algebraic fields*, Springer, Berlin-Heidelberg-New York, 1968.
- [40] W. Luo, *Nonvanishing of L-values and the Weyl law*, Ann. Math. **154** (2001), 477–502.
- [41] W. Luo and P. Sarnak, *Quantum ergodicity of eigenfunctions on  $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}^2$* , Inst. Haute Études Sci. Publ. Math **81** (1995), 207–237.
- [42] ———, *Quantum variance for Hecke eigenforms*, Ann. Sci. École Norm. Sup. (4) **37** (2004), 769–799.
- [43] H. Maaß, *Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Ann. Math. **122** (1949), 141–183.
- [44] ———, *Die Differentialgleichungen in der Theorie der elliptischen Modulfunktionen*, Math. Ann. **125** (1953), 235–263.
- [45] R. B. Melrose, *Weyl’s conjecture for manifolds with concave boundary*, Geometry of the Laplace operator, Proc. Sympos. Pure Math., vol. XXXVI, Amer. Math. Soc., Providence, R.I., 1980, pp. 257–274.
- [46] S. Minakshisundaram and A. Pleijel, *Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds*, Canadian J. Math. **1** (1949), 242–256.
- [47] W. Müller, *Weyl’s law in the theory of automorphic forms*, Groups and Analysis: The Legacy of Hermann Weyl (K. Tent, ed.), London Mathematical Society Lecture Note Series, London Mathematical Society, London, 2008, 2007, available at [arxiv:0710.2319](https://arxiv.org/abs/0710.2319) [math.SP].
- [48] R. S. Phillips and P. Sarnak, *Invent. Math.* **80** (1985), 339–364.
- [49] I. I. Piatetski-Shapiro, *Multiplicity one theorems*, (Oregon State Univ., Corvallis, Ore., 1977), Automorphic forms, representations and L-functions. 1 (A. Borel and W. Casselman, eds.), Proc. Sympos. Pure Math., vol. XXXIII, Amer. Math. Soc., Providence, RI, 1979, pp. 209–212.
- [50] Lord Raleigh, *The dynamical theory of gases and of radiation*, Nature **72** (1905), 54–55, 243–244.
- [51] Z. Rudnick, *On the asymptotic distribution of zeros of modular forms*, Int. Math. Res. Not. **34** (2005), 2059–2074.
- [52] Z. Rudnick and P. Sarnak, *The Behavior of Eigenstates of Arithmetic Hyperbolic Manifolds*, Commun. Math. Phys. **161** (1994), 195–213.
- [53] P. Sarnak, *Notes on the Generalized Ramanujan Conjectures*, Clay Math. Proc. **4** (2005).
- [54] I. Satake, *Spherical functions and Ramanujan conjecture*, (Boulder, Colo., 1965), Algebraic Groups and Discontinuous Subgroups (A. Borel and G. Mostow, eds.), Proc. Sympos. Pure Math., vol. IX, Amer. Math. Soc., Providence, RI, 1966, pp. 258–264.
- [55] A. Selberg, *Harmonic analysis* (1954), preprint, Universität Göttingen; reprinted in A. Selberg, *Collected Papers*, Vol. 1, Springer-Verlag, Berlin-Heidelberg-New York, 1989.
- [56] ———, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series.*, J. Ind. Math. Soc. **20** (1956), 47–87.
- [57] ———, *Automorphic functions and integral operators*, Seminars on Analytic Functions, Vol. 2, Institute for Adv. Study, 1957, pp. 152–161.
- [58] ———, *On discontinuous groups in higher dimensional symmetric spaces*, International Colloquium on Function Theory, Tata Institute, Bombay, 1960, pp. 147–164.
- [59] ———, *Discontinuous groups and harmonic analysis*, Proc. International Math. Congr. Stockholm (1962), 177–189.
- [60] A. Schnirelman, *Usp. Mat. Nauk* (6) **29** (1974), 181–182.

- [61] K. Soundararajan, *Quantum unique ergodicity for  $SL_2(\mathbf{Z}) \backslash \mathfrak{H}$* , *Annals of Math* (2) **172** (2010), 1529–1538.
- [62] ———, *Quantum Unique Ergodicity and Number Theory*, *Proc. International Math. Congr. Hyderabad* (2010), 357–382.
- [63] D. I. Wallace, *Maximal parabolic terms in the Selberg trace formula for  $SL(3, \mathbf{Z}) \backslash SL(3, \mathbf{R}) / SO(3, \mathbf{R})$* , *J. Number Theor.* (2) **29** (1988), 101–117.
- [64] ———, *Terms in the Selberg trace formula for  $SL(3, \mathbf{Z}) \backslash SL(3, \mathbf{R}) / SO(3, \mathbf{R})$  associated to Eisenstein series coming from a maximum parabolic subgroup*, *Proc. Amer. Math. Soc* (4) **106** (1989), 875–883.
- [65] ———, *Terms in the Selberg trace formula for  $SL(3, \mathbf{Z}) \backslash SL(3, \mathbf{R}) / SO(3, \mathbf{R})$  associated to Eisenstein series coming from a minimal parabolic subgroup*, *Trans. Amer. Math. Soc.* (2) **327** (1991), 781–793.
- [66] ———, *The Selberg trace formula for  $SL(3, \mathbf{Z}) \backslash SL(3, \mathbf{R}) / SO(3, \mathbf{R})$* , *Trans. Amer. Math. Soc.* (1) **345** (1994), 1–36.
- [67] T. Watson, *Central Value of Rankin Triple L-function for Unramified Maass Cusp Forms*, Princeton Ph.D. thesis, 2004.
- [68] H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)*, *Math. Ann* **71** (1912), 441–479.
- [69] ———, *Über die Randwertaufgabe der Strahlungstheorie und asymptotische Spektralgesetze*, *Math. Ann* **71** (1913), 441–479.
- [70] K. Yosida, *Functional Analysis*, 2nd ed., Springer-Verlag, Berlin, 1968.
- [71] D. Zagier, *Eisenstein series and the Selberg trace formula I*, *Automorphic Forms, Representation Theory and Arithmetic*, Springer-Verlag, Berlin-Heidelberg-New York, 1981, pp. 275–301.
- [72] S. Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, *Duke Math. J.* **55** (1987), 919–941.
- [73] ———, *Selberg trace formulae and equidistribution theorems*, *Memoirs of the AMS* **96** (1992), no. 465.
- [74] ———, *On the rate of quantum ergodicity*, *Commun. Math. Phys.* **160** (1994), 81–92.
- [75] P. Zhao, *Quantum Variance of Maass-Hecke Cusp Forms*, *Commun. Math. Phys.* **297** (2010), 475–514.