

SPECTRA OF HYPERBOLIC SURFACES

PETER SARNAK

ABSTRACT. These notes attempt to describe some aspects of the spectral theory of modular surfaces. They are by no means a complete survey.

CONTENTS:

- (1) Introduction.
- (2) Existence.
- (3) Low-Energy Spectrum.
- (4) High-Energy Spectrum.

§1. INTRODUCTION

Harmonic analysis on \mathbb{R}/\mathbb{Z} , that is to say the spectral theory of the translation invariant operator $D = \frac{d}{dx}$ on periodic functions, is an important first step in understanding the Riemann Zeta Function. In more detail, the Poisson Summation formula asserts that if $f \in \mathcal{S}(\mathbb{R})$ (that is a smooth function which together with its derivatives is rapidly decreasing) and $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e(-\xi x)dx$, $e(z) := e^{2\pi iz}$, then

$$(1) \quad \sum_{n \in \mathbb{Z}} f(n) = \sum_{m \in \mathbb{Z}} \hat{f}(m).$$

This is proven by expanding the periodic function

$$(2) \quad F(x) = \sum_{n \in \mathbb{Z}} f(n+x)$$

in a Fourier series.

Recall that the zeta function $\zeta(s)$ is defined for $\Re(s) > 1$ by

$$(3) \quad \zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_p (1 - p^{-s})^{-1},$$

the product being over the prime numbers and the identity being equivalent to unique factorization of integers into primes.

Received by the editors March 6, 2003.

2000 *Mathematics Subject Classification*. Primary 11F03, 11N75, 11R42, 35P30.

This paper is based on notes for the Colloquium Lecture presented at the Joint Mathematics Meetings in Baltimore in January 2003.

Applying Poisson summation to $\sum_{n \in \mathbb{Z}} f(nx)$ with f even and $f(0) = \hat{f}(0) = 0$ in the relation

$$\frac{1}{2} \int_0^\infty \left(\sum_{n \in \mathbb{Z}} f(nx) \right) x^s \frac{dx}{x} = \zeta(s) \int_0^\infty f(x) x^s \frac{dx}{x}$$

leads to Riemann's analytic continuation and functional equation for $\zeta(s)$ (see [Bo] for a recent historical account). The functional equation is the identity

$$(4) \quad \Lambda(s) := \pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \Lambda(1-s)$$

where

$$(5) \quad \Gamma(s) = \int_0^\infty e^{-x} x^s \frac{dx}{x}.$$

The modern theory of automorphic forms is concerned in part with spectral problems associated with quotients of more general (nonabelian) groups, their homogeneous and symmetric spaces and the formation of related zeta functions.

In this article we will discuss only the case of the upper-half plane. This case is plenty interesting and challenging and still offers quite striking applications. However, it will become clear that to fully understand even this special case, more general groups are needed and are used. Let $\mathbb{H} = \{z = x + iy | y > 0\}$ be the upper half-plane. It comes with a complex as well as a Riemannian structure, the line element being

$$(6) \quad ds = \frac{|dz|}{y}.$$

The group $G = SL(2, \mathbb{R})$ of 2×2 real matrices of determinant equal to 1 acts on \mathbb{H} by linear fractional transformations. For

$$(7) \quad g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, z \longrightarrow gz = \frac{az + b}{cz + d}.$$

This action preserves both the complex and Riemannian structures on \mathbb{H} . With ds , \mathbb{H} has curvature $K \equiv -1$ and is a *hyperbolic* surface (the universal such surface which is simply connected). In these coordinates the area element for (\mathbb{H}, ds) takes the form

$$(8) \quad dA(z) = \frac{dx dy}{y^2},$$

and the Laplacian $\Delta := \operatorname{div} \operatorname{grad}$ is given by

$$(9) \quad \Delta = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

Δ commutes with the action of G ; that is if $R_g f(z) = f(gz)$, then

$$(10) \quad R_g \Delta = \Delta R_g, \quad \text{for } g \in G.$$

Next, we need a discrete subgroup Γ of G . For us the most important subgroups are the modular group

$$(11) \quad SL(2, \mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G \mid a, b, c, d \in \mathbb{Z} \right\}$$

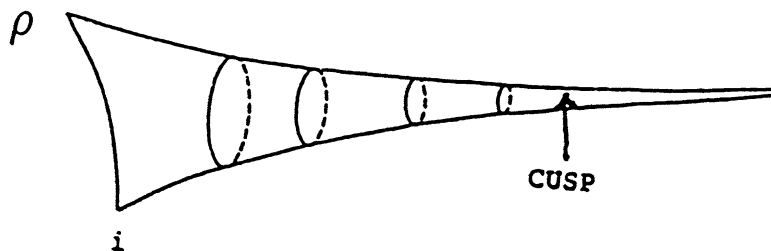


FIGURE 1.

and its congruence subgroups. For $N \geq 1$ the principal congruence subgroup of level N is

$$\Gamma(N) = \{ \gamma \in SL(2, \mathbb{Z}) \mid \gamma \equiv I \pmod{N} \} .$$

A congruence subgroup Γ of $\Gamma(1)$ is a subgroup for which there is M such that $\Gamma \supset \Gamma(M)$. The modular surface $X(N)$ is defined as the quotient $\Gamma(N) \backslash \mathbb{H}$. It is a finite area, non-compact, hyperbolic surface. Of course with its complex structure $X(N)$ is a Riemann surface (or a curve, as an algebraic geometer would call it) whose genus is roughly N^3 when N gets large. $X(N)$ is also a parameter space (moduli space) of elliptic curves with suitable structures. For example, as is well known from complex function theory $X(1)$ parametrizes the space of elliptic curves E over \mathbb{C} , while the set of E 's together with points of order N on each E gives rise to meromorphic functions which generate the field of such functions on $X(N)$; see [Shi1]. All the above realizations of $X(N)$ are important.

As a quotient space the modular surface $X(1) = \Gamma(1) \backslash \mathbb{H}$ looks like Figure 1. This can be seen from the standard fundamental domain $\mathcal{F}(1)$ for the action of $\Gamma(1)$ (see Figure 2). To see this, note that we may identify \mathbb{H} with G/K through the association $g \rightarrow gi$ which has stabilizer $K = SO(2)$. Since $\Gamma(1)$ is discrete in G and K is compact, $\Gamma(1)$ acts on \mathbb{H} without limit points. Using the transformations $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ which generate $\Gamma(1)$, we can reduce any $z \in \mathbb{H}$ to $\mathcal{F}(1)$ as follows: First use T^m for suitable $m \in \mathbb{Z}$ to move z to z' with $y(z') = y(z)$ and $-\frac{1}{2} \leq x(z') < \frac{1}{2}$. If $|z'| \geq 1$, then $z' \in \mathcal{F}(1)$; otherwise apply S . Then $y(Sz') \geq y(z')$ and repeat the above process. It must terminate after a finite number of steps with z having been reduced to $\mathcal{F}(1)$, for otherwise $\Gamma(1)z$ would have a limit point in \mathbb{H} . From Figure 2 and (8) it is clear that $\text{Area}(X(1)) < \infty$. In fact, using hyperbolic geometry one has

$$(12) \quad \text{Area}(X(1)) = \frac{\pi}{3} .$$

We can now formulate the fundamental spectral problem. We seek non-zero, square integrable solutions to

$$(13) \quad \left. \begin{aligned} \Delta\phi + \lambda\phi &= 0 \\ \phi(\gamma z) &= \phi(z), \gamma \in \Gamma(N) \\ \int_{X(N)} |\phi(z)|^2 dA(z) &< \infty \end{aligned} \right\} .$$

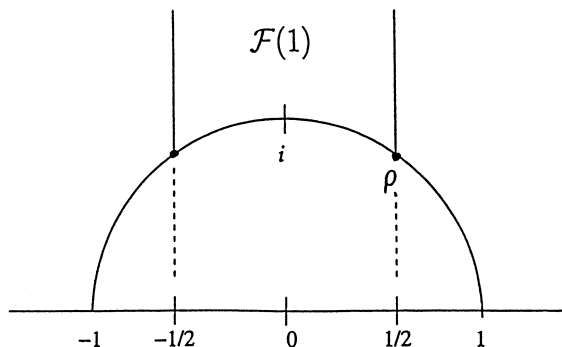


FIGURE 2.

The numbers¹ $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$ for which (13) has a solution turn out to be discrete and form the (discrete) spectrum of $X(N)$. The corresponding eigenfunctions $\phi_\lambda(z)$ are also of much interest. The only obvious eigenfunction is the constant function for which $\lambda = \lambda_0 = 0$. We call a solution to (13) a *Maass form* after Maass, who first introduced them.² Their existence is by no means obvious; see Section 2.

Intensive numerical investigations [He1] [St] have determined the first 10,000 eigenvalues for $X(1)$. The first few are: 0, 91.14..., 148.43..., 190.13..., 206.16... They are respectively even, odd, odd, even and odd, with respect to the symmetry about $x = 0$. The method used to do these computations is outlined in Appendix 7. The profiles of the eigenfunctions $\phi_1, \phi_{10}, \phi_{17}$ and ϕ_{33} corresponding to $\lambda_1, \lambda_{10}, \lambda_{17}$ and λ_{33} are pictured in Figure 3; they were computed in [G-S] and [Str].

Why the interest in these special vibrating membranes? To answer this we describe some applications of this spectral theory. We begin with algebraic number theory. Let K be a Galois extension of the rational number field \mathbb{Q} . Let $\rho : \text{Gal}(K/\mathbb{Q}) \rightarrow \text{GL}(2, \mathbb{C})$ be an irreducible two-dimensional complex representation of the finite group $\text{Gal}(K/\mathbb{Q})$. To each prime p unramified in K one can associate a conjugacy class Frob_p in $\text{Gal}(K/\mathbb{Q})$; see Appendix 2. Following Artin define the L -functions $L(s, \rho)$ by

$$\begin{aligned}
 L(s, \rho) &= \prod_p \det(I - \rho(\text{Frob}_p)p^{-s})^{-1} \\
 (14) \quad &= \prod_p (1 - \text{trace} \rho(\text{Frob}_p)p^{-s} + \det \rho(\text{Frob}_p)p^{-2s})^{-1} \\
 &:= \sum_{n=1}^{\infty} \lambda_\rho(n)n^{-s}
 \end{aligned}$$

(care must be taken in defining the local Euler factors at ramified primes [Lan2]). Artin conjectured that $L(s, \rho)$ extends to an entire function of s . For our discussion let's assume that ρ is unramified over \mathbb{R} (see Appendix 1 for definitions); then it

¹An integration by parts shows that these must be non-negative.

²For $N > 1$ he gave some explicit examples; see Appendix 1.

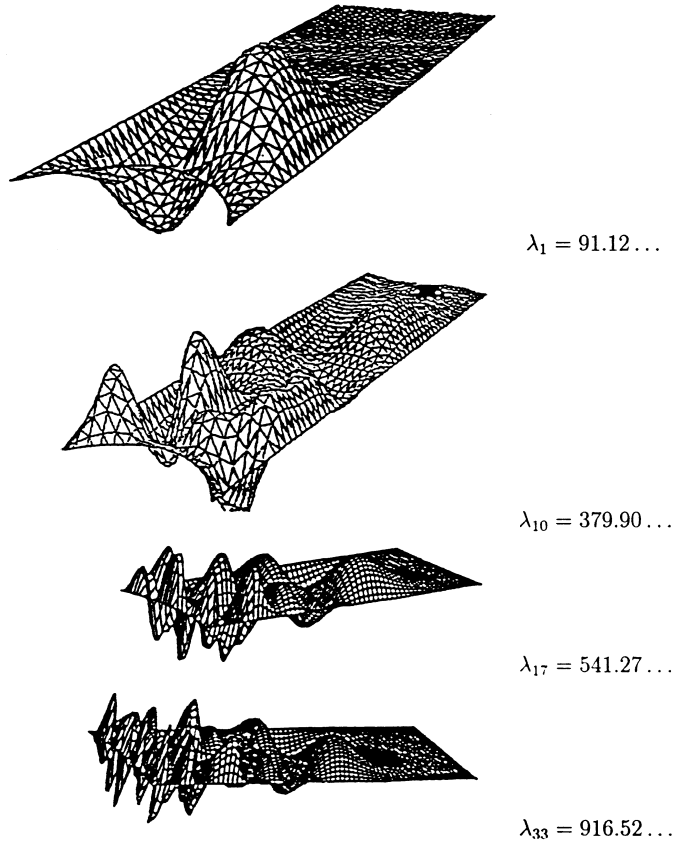


FIGURE 3. The 1st, 10th, 17th and 33rd eigenfunctions for the modular group. They are all odd with respect to the symmetry $z \rightarrow -\bar{z}$.

was shown very recently [Boo]³ that if $L(s, \rho)$ is entire, then

$$(15) \quad \phi(z) = \sum_{n=1}^{\infty} \lambda_{\rho}(n) y^{1/2} K_0(2\pi n y) \cos(2\pi n x)$$

is a Maass form for $X(N)$ with eigenvalue $\lambda_{\phi} = \frac{1}{4}$! Here N is the conductor of ρ , (Appendix 2), and $K_0(y)$ is the Bessel function. The latter may be defined by

$$(16) \quad K_{\nu}(y) = \int_0^{\infty} e^{-y \cosh t} \cosh(\nu t) dt,$$

and it satisfies

$$(17) \quad K_{\nu}'' + \frac{1}{y} K_{\nu}' + \left(1 - \frac{\nu^2}{y^2}\right) K_{\nu} = 0.$$

Observe that if

$$(18) \quad \psi(z) = \sum_{n=1}^{\infty} a(n) y^{1/2} K_{it}(2\pi n y) \cos(2\pi n x)$$

³This constitutes a strengthening of [Wei] in that no twists are needed.

for any coefficients $a(n)$, then

$$(19) \quad \Delta\psi + \left(\frac{1}{4} + t^2\right) \psi(z) = 0.$$

The important feature in (15) is the $\Gamma(N)$ invariance.

Thus, according to Artin’s Conjecture even Galois representations (Appendix 2) give rise to Maass forms with eigenvalue $\frac{1}{4}$. Similarly, odd ones give rise to *holomorphic* forms of weight 1; see [Se]. If the image of ρ in $PGL(2, \mathbb{C})$ (which being finite must according to Klein [KI] be one of the following: dihedral, tetrahedral, octahedral or icosahedral) is not icosahedral, then the Artin Conjecture is true. The most difficult cases are the tetrahedral and octahedral ones which were established in [La1] and [Tu].⁴ The proof makes crucial use of the spectral theory via use of the trace formula (Appendix 3) to establish cyclic base change. The latter gives a precise relation between the automorphic (in particular Maass) spectrum over a number field L and that of a cyclic extension K of L .

It is believed that conversely any Maass form ϕ with eigenvalue $\lambda_\phi = \frac{1}{4}$ must correspond to an even Galois representation as above. In [Sa1] it is shown that if $\phi(z)$ is a Maass form for some $X(N)$ and has integer coefficients in its Fourier expansion (18), then in fact $\lambda_\phi = \frac{1}{4}$ and ϕ comes from a Galois representation of dihedral or tetrahedral type. The proof of this result relies on recent advances [K-S1] on the functorial lifts $\text{sym}^3 : GL(2) \rightarrow GL(4)$; see Appendix 1.

We turn to some applications of this spectral theory to problems in analytic number theory. In these it is the entire spectrum that usually enters. For example, let $\lambda_\rho(n)$ be the coefficients in (14) or for that matter the coefficients of any holomorphic or Maass form. For integers $\nu_1, \nu_2, h \geq 1$ consider the Dirichlet series

$$(20) \quad D(s, \nu_1, \nu_2, h) = \sum_{\nu_1 n - \nu_2 m = h} \lambda_\rho(n) \overline{\lambda_\rho(m)} (\nu_1 n + \nu_2 m)^{-s}.$$

The series converges absolutely for $\Re(s) > 1$. As was first noted in [Sel2] and is explained in Appendix 6, D has an analytic continuation to $\Re(s) \geq \frac{1}{2}$ with possible poles at $s = \frac{1}{2} + it_\phi$ where $0 \neq \lambda_\phi = \frac{1}{4} + t_\phi^2$ is an eigenvalue of Δ on $X(\nu_1 \nu_2 N)$. Notice that if $\lambda_\phi \geq \frac{1}{4}$ (see Section 3), then in fact $D(s, \nu_1, \nu_2, h)$ is analytic in $\Re(s) > \frac{1}{2}$. The latter represents a substantial (“square root”) cancellation in the following smooth sums: If $\psi \in C_0^\infty(0, \infty)$ and $\epsilon > 0$, there is $C_{\epsilon, \psi}$ such that

$$(21) \quad \left| \sum_{\nu_1 n - \nu_2 m = h} \psi\left(\frac{\nu_1 n + \nu_2 m}{Y}\right) \lambda_\rho(n) \overline{\lambda_\rho(m)} \right| \leq C_{\epsilon, \psi} Y^{\frac{1}{2} + \epsilon}, \text{ as } Y \rightarrow \infty.$$

Cancellation in such and related arithmetical sums is at the heart of many of the applications of the Maass form spectral theory. We mention a couple.

(1) EQUIDISTRIBUTION OF ROOTS:

Let $f(x) \in \mathbb{Z}[x]$ be an irreducible polynomial over \mathbb{Q} . If K is the splitting field for f , then Frob_p and (14) are concerned with how $f(x)$ factors mod p , for different primes p . Let $0 \leq x_j(p) \leq p - 1$, $j = 1, 2, \dots, \nu_p$, $\nu_p \leq \deg f$ be the roots of $f(x) \equiv 0(p)$ if there are any. Numerical experiments suggest that $\{x_j(p)/p\}$, $j = 1, \dots, \nu_p$, $p \leq X$, become equidistributed

⁴For recent progress for ρ odd and icosahedral see [B-D-S-T].

in $[0, 1]$ as $X \rightarrow \infty$. In [D-F-I] and [To] it is shown that this is indeed the case when f is of degree 2. That is for $0 \leq \alpha \leq \beta \leq 1$,

$$(22) \quad \frac{\#\{p \leq X, j \leq \nu_p \mid \frac{x_j(p)}{p} \in [\alpha, \beta]\}}{\#\{p \leq X, j \leq \nu_p\}} \rightarrow \beta - \alpha, \quad \text{as } X \rightarrow \infty.$$

(2) HILBERT'S ELEVENTH PROBLEM:

This asks about the representations of integers in a number field K (respectively of elements in K) by an integral (respectively K rational) quadratic form $F(x_1, x_2, \dots, x_n)$ in n -variables. For the case of representability of members of K by a form F with coefficients in K , this was resolved by Hasse [Ha]. He showed that $F(x) = n$ has a solution $x \in K^n, m \in K$ iff $F(x) = m$ has a solution over K_v for every completion K_v of K . This is called a local to global principle. The case of integral representations is apparently more difficult. After works of Minkowski, Siegel and others, an appropriate local to global principle for integral forms in four or more variables was established in [Kne]. In the most interesting case where the form is definite the local to global principle applies when m is large. For two variables there is in general no such local to global principle. The case of three variables was resolved recently in [Co-PS-S], where a local to global principle is proven. There is an added caveat of the result being ineffective and that for three variables there may be a finite number of quadratic exceptional sequences [D-SP], [SP]. An interesting application of the above is the determination of which integers m in K are a sum of three squares of integers of K . An important ingredient in [Co-PS-S] is the analysis of the Maass form spectrum and especially the low-energy eigenvalues for Hilbert modular manifolds (for an example of these see the end of Section 4), which are the natural generalizations of the $X(N)$'s for K (see [Cog] for an outline of the set up and of the methods).

We hope that the above examples convince you of the central role that the spectrum of $X(N)$ plays in number theory. In the analytic aspects it is the low-energy spectrum that is critical. In Section 3 we discuss this aspect of the spectrum. The study of the large eigenvalues for a given X is also of interest, especially as a problem in mathematical physics. The limit $\lambda \rightarrow \infty$ is the so-called semi-classical limit. In our case of a hyperbolic surface X , we are dealing with a quantization of a classically chaotic Hamiltonian, and for these (unlike the case of a completely integrable Hamiltonian) the relation in the semi-classical limit between the classical and quantum mechanics is not well understood. We discuss these issues as well as some recent decisive breakthroughs for the surfaces $X(N)$ in Section 4.

§2. EXISTENCE

We first recall a fundamental result of Weyl. Let $\Omega \subset \mathbb{R}^2$ be a compact planar Euclidian domain with smooth boundary $\partial\Omega$. The eigenvalue problem for the usual Laplacian $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ on Ω with Dirichlet boundary conditions is

$$\Delta\phi(z) + \lambda\phi(z) = 0 \quad \text{for } z \in \Omega$$

$$\phi|_{\partial\Omega} = 0.$$

Let $N_\Omega(R)$ be the number of such eigenvalues λ counted with multiplicity, with $\lambda \leq R$. Weyl's result, known as Weyl's law, asserts that

$$N_\Omega(R) \sim \frac{\text{Area}(\Omega)}{4\pi} R, \quad \text{as } R \longrightarrow \infty.$$

The result has been generalized to Riemannian manifolds of any dimension. For many, the favored modern means of proving this law is by analyzing the small time asymptotics of the heat kernel on $\mathbb{R} \times \Omega$ [Mc-Si]. The sharpest forms of the remainder terms for such Weyl asymptotics are gotten by analyzing the propagation of singularities for the wave kernel [Du-Gu].

We return to our setting of finite area hyperbolic surfaces. Since the spaces $X(N)$ are not compact it is not at all clear that there are any solutions to (13) with $\lambda > 0$. The discrete spectrum that we seek lies embedded in the continuous spectrum, making these eigenvalues very difficult to isolate analytically. The theory of Eisenstein series and their analytic continuation developed in [Sel1] for a general hyperbolic surface $X_\Gamma = \Gamma \backslash \mathbb{H}$ furnishes the continuous spectrum. The latter consists of the interval $[\frac{1}{4}, \infty)$ with multiplicity the number of cusps of X_Γ . The constant term in the Fourier expansion of the Eisenstein series (see (43)), $\phi_\Gamma(s)$ (called the determinant of the scattering matrix in [L-P] or the intertwining operator in [Sh]) is meromorphic in \mathbb{C} . Its only poles in $\Re(s) \geq \frac{1}{2}$ are in $(\frac{1}{2}, 1]$, and the residues at these poles furnish solutions to (13), called the residual spectrum of X . The poles of $\phi_\Gamma(s)$ in $\Re(s) < \frac{1}{2}$ yield resonances for problem (13).

The orthogonal complement in $L^2(X_\Gamma)$ of the continuous and residual spectrum is the cuspidal space $L^2_{\text{cusp}}(X_\Gamma)$. It is invariant under Δ , and the resolvent $(\lambda - \Delta)^{-1}$ is compact when restricted to $L^2_{\text{cusp}}(X_\Gamma)$. L^2_{cusp} has a simple description in terms of periods of closed horocycles associated with the cusps C_1, \dots, C_ℓ of X_Γ (see Figure 3a). For example for $\Gamma(1)$ the modular group

$$(23) \quad L^2_{\text{cusp}}(X(1)) = \left\{ f \in L^2(X(1)) \mid \int_0^1 f(z) dx = 0 \text{ for almost all } y > 0 \right\}.$$

A Maass form (as in (13)) which also lies in L^2_{cusp} is called a Maass *cusp* form. These are the building blocks (the fundamental particles) of the theory of automorphic forms. Their existence in this setting is tied to the size of $L^2_{\text{cusp}}(X(1))$. Whether $L^2_{\text{cusp}}(X) \neq \{0\}$ for a general hyperbolic X is by no means obvious. An interesting discussion in terms of integral geometry is given in [Lax].

One of the early triumphs of the trace formula (Appendix 3) developed in [Sel1] was the proof that the modular surfaces $X(N)$ carry an abundance of Maass cusp forms. For these surfaces the functions $\phi_{\Gamma(N)}(s)$ may be expressed in terms of Dirichlet L -functions (Appendix 1). For example for $\Gamma(1)$

$$(24) \quad \phi_{\Gamma(1)}(s) = \frac{\Lambda(2s-1)}{\Lambda(2s)}$$

with $\Lambda(s)$ as in (4). In particular, $\phi_{\Gamma(N)}(s)$ has no poles in $(\frac{1}{2}, 1)$ so that in these cases there is no residual spectrum (besides $\lambda = 0$) and any solution of (13) with $\lambda > 0$ is automatically a cusp form. For the general X_Γ , the trace formula provides a Weyl-like law for counting asymptotically the sum of the cuspidal spectrum and the continuous spectrum, the latter through the winding of the unitary quantity $\phi_\Gamma(s)$ for $\Re(s) = \frac{1}{2}$ (Appendix 3). In the case of a modular surface the expression of $\phi_{\Gamma(N)}(s)$ in terms of L -functions allows one to show that the contribution of the



FIGURE 3a.

continuous spectrum to this Weyl law is negligible. That is, for $X(N)$ it is shown in [Sel1] (see Appendix 3) that

$$(25) \quad N_{\Gamma(N)}^{\text{cusp}}(\lambda) := \sum_{0 < \lambda_j \leq \lambda} 1 \sim \frac{\text{Area } X(N)}{4\pi} \lambda, \text{ as } \lambda \rightarrow \infty.$$

Thus, solutions to (13) exist and in abundance! We call a surface X essentially cuspidal if (25) holds.

It is interesting⁵ from many points of view to understand when solutions to (13) exist, for the more general hyperbolic surface X . Contrary to early beliefs it appears now that essential cuspidality is limited to special arithmetic surfaces! We review briefly these developments. In the papers [P-S1],[P-S2],[P-S3], the behavior of the discrete spectrum is studied when Γ undergoes a deformation. Fix N and let $T(\Gamma(N))$ be the deformation space (Teichmüller space) of continuous deformations of $\Gamma(N)$ as a discrete subgroup of cofinite area in $SL(2, \mathbb{R})$, that is as a hyperbolic surface. The cotangent space to $T(\Gamma(N))$ at $\Gamma(N)$ may be naturally identified with the space of holomorphic quadratic differentials on $X(N)$ (here the analytic structure of $X(N)$ is exploited) with suitable behavior at the punctures [Be]. Using the uniformization $X(N) = \Gamma(N) \backslash \mathbb{H}$ this cotangent space can be realized as the space of holomorphic *cusp* forms Q of weight 4 for $\Gamma(N)$. That is:

$$(26) \quad \left. \begin{aligned} &\bullet \quad Q(z) \text{ is holomorphic in } \mathbb{H} \\ &\bullet \quad Q\left(\frac{az+b}{cz+d}\right) = (cz+d)^4 Q(z) \text{ for } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma(N) \\ &\bullet \quad Q \text{ vanishes at the cusps} \end{aligned} \right\}.$$

⁵This is especially relevant in higher dimensions where the only understanding of $\phi_{\Gamma}(s)$ comes from spectral theory [La3].

Unlike the issue of Maass cusp forms, the dimension of the space of holomorphic cusp forms of weight 4 on any surface X is determined purely in terms of the topology of X via the Riemann-Roch formula.

Let X_t be a real analytic curve in $T(\Gamma(N))$ with $X_0 = X(N)$ and with tangent vector at $t = 0$ given by $Q(z)$ as above. In order to investigate the behavior of the spectrum under deformation, define the singular set $\sigma(X)$ to be the numbers $\frac{1}{2} + it_j$ (with multiplicities) if $\lambda_j = \frac{1}{4} + t_j^2$ is an eigenvalue of Δ on X_Γ , together with the poles ρ_j (again with order) of $\phi_\Gamma(s)$ in $\Re(s) < \frac{1}{2}$. The multiplicity of the point $s = \frac{1}{2}$ as a member of $\sigma(X)$ requires a special definition [P-S3]. Note that $\sigma(X) \cap \{s | \Re(s) \geq \frac{1}{2}\}$ consists of points located in $\{\Re(s) = \frac{1}{2}\} \cup (1/2, 1]$. In [P-S3] it is shown that $\sigma(X_t)$ is an algebroid function of t . That is, branches $\rho_j(t)$ may be chosen so as to be analytic in t or at worst to have algebraic singularities locally. Now, suppose that $\frac{1}{2} + it_j$, $t_j > 0$, is a simple point in $\sigma(X(N))$ corresponding to a Maass form ϕ_j . Let $\rho_j(t)$ be its corresponding deformation. Either $\Re(\rho_j(t)) \equiv \frac{1}{2}$ or $\rho_j(t)$ moves into $\Re(s) < \frac{1}{2}$. In the latter case the cusp form ϕ_j is dissolved under the deformation into a pole of $\phi_{\Gamma t}(s)$. In [P-S3] the following ‘‘Fermi Golden rule’’⁶ is proven:

$$(27) \quad \left. \frac{d^2}{dt^2} \Re(\rho_j(t)) \right|_{t=0} = -c(t_j) \left| L\left(\frac{1}{2} + it_j, Q \times \phi_j\right) \right|^2.$$

Here $c(t_j) > 0$ and $L(s, Q \times \phi_j)$ is the Rankin-Selberg L -function of Q and ϕ_j (see Appendix 1 for a definition). The proof of (27) relies in a crucial way on the scattering theory (specifically the semi-group $Z(t)$) developed in [L-P]. A similar formula for the movement of $s_j = \frac{1}{2}$ (ie $t_j = 0$ above), when movement to the right is also possible, is developed in [Pe].

The dissolving condition (27) boils down to a vanishing question about a special value of a zeta function. Note that the special value is on the critical line for this function so that the vanishing does not violate the corresponding Riemann Hypothesis (Appendix 1). Starting with [D-I] there have been a series of results concerning the non-vanishing of these numbers $L(\frac{1}{2} + it_j, Q \times \phi_j)$, $|t_j| \leq T$. In [Lu] it is shown that a positive proportion of these numbers are not zero as $T \rightarrow \infty$. Note that once a Maass form ϕ_j is dissolved it is a pole (or resonance) for all but countably many values of t . Thus generically in $T(\Gamma(N))$ it is dissolved into a pole. There is a technical difficulty in this analysis in the case that $\frac{1}{2} + it_j$ is a multiple eigenvalue. Specifically degenerate perturbation theory leads to a much less tractable (in terms of L -functions) formula for the Golden Rule. This leads us to the issue of the possible multiplicities of the eigenvalues of $X(N)$. This has proven to be an embarrassingly difficult problem; see Section 4. It is believed that the multiplicity $m(\lambda)$ of an eigenvalue λ of $X(N)$ should be uniformly bounded (here N is fixed). Assuming this, then the above analysis leads to the conclusion that for the generic $\Gamma \in T(\Gamma(N))$, X_Γ is not essentially cuspidal. The dissolving condition (27) led to the following conjectures [Sa2]:

- Conjecture 1.** (a) *The generic Γ in a given Teichmüller space of finite area hyperbolic surfaces is not essentially cuspidal.*
 (b) *Except for the Teichmüller space of the once punctured torus, the generic Γ has only a finite number of discrete eigenvalues.*

⁶See [Si] for a discussion of this terminology in the context of helium-like atoms.

The reason for omitting the once punctured torus is that this Teichmüller space has a persistent symmetry of order 2 which leaves the continuous spectrum invariant. The functions which are odd with respect to this symmetry constitute half of $L^2(X_\Gamma)$, and these functions are all cuspidal. In the context of deforming X_Γ in the infinite dimensional space of all Riemannian metrics which are conformal to X_Γ , part (b) is known and was established in [Co2].

Remarkable further progress on the above conjecture was made in the series of papers [Wol1], [Wol2]. Instead of considering regular deformations in $T(\Gamma)$, Wolpert follows special deformations Γ_t to the boundary of $T(\Gamma)$. This is a formidable task, being an analysis of eigenvalues embedded in the continuum for a very singular perturbation. The payoff of going to the boundary is well worth it, the point being that the key dissolving condition (27) takes a similar form with $L(\frac{1}{2} + it_j, Q \times \phi_j)$ being replaced by $L(\frac{1}{2} + it_j, E_4 \times \phi_j)$, where E_4 is the Eisenstein series:

$$(28) \quad E_4(z) = \sum_{\gamma \in \Gamma_\infty \setminus \Gamma(N)} (cz + d)^{-4}$$

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & Nm \\ 0 & 1 \end{pmatrix} \middle| m \in \mathbb{Z} \right\}.$$

This holomorphic weight 4 Eisenstein series generates a singular deformation at the boundary of T . The theory of such singular deformations (of infinite energy) was developed in [W]. The advantage of E_4 over Q is that the degree 4 L -function $L(s, E_4 \times \phi_j)$ factors into the two degree 2 L -functions (Appendix 1) $L(s + 1, \phi_j) \cdot L(s - 1, \phi_j)$. Thus evaluation at $s = \frac{1}{2} + it_j$ ensures the non-vanishing of this product since L -functions can only vanish in $0 < \Re(s) < 1$ or at the trivial zeroes! Unfortunately, there is still the caveat of the multiplicity issue. In fact, the analysis needed to handle this very singular perturbation requires the strongest multiplicity conjecture (see Section 4).

Theorem 1 ([Wol2]). *Assume that the cuspidal Maass spectrum (of new and old forms) for $X(2)$ is simple; then part (a) of Conjecture 1 is true.*

One may ask if there is a characterization of those Γ which have many Maass cusp forms. In [Sa2] the question of the relation to arithmeticity is raised. The simplest case to explain this phenomenon is the one for a hyperbolic triangle (see Figure 4). The conjecture is that there will be infinitely many solutions to $\Delta u + \lambda u = 0$, $u \in L^2(\mathcal{F}_q)$, $\partial_n u \Big|_{\partial \mathcal{F}_q} = 0$ (that is Neumann boundary conditions) iff $q = 3, 4$ or 6 . $q = 3$ corresponds to even Maass cusp forms for $\Gamma(1) = SL(2, \mathbb{Z})$; $q = 4$ and 6 correspond to other congruence subgroups of $SL(2, \mathbb{Z})$. All other integer values of q give, via reflections in the sides of the triangle, non-arithmetic subgroups of $SL(2, \mathbb{R})$ [Ta]. If $q \notin \mathbb{Z}$, then reflections in the sides of \mathcal{F}_q do not generate a discrete subgroup of $SL(2, \mathbb{R})$, but the above eigenvalue problem for the triangle makes sense. This conjecture about these triangles has been checked numerically in [He2]. For example for $q = 5, 7$ and 8 no eigenvalues were found for $0 < \lambda < 3600$. In [Ju] the dissolving condition (27) at “ $q = \infty$ ” is developed using a similar singular perturbation method to [Wol1]. Furthermore, in [Ju] it is shown that subject to the same hypothesis as in Theorem 1, \mathcal{F}_q has only a finite number of eigenvalues for all but a countable set of q .

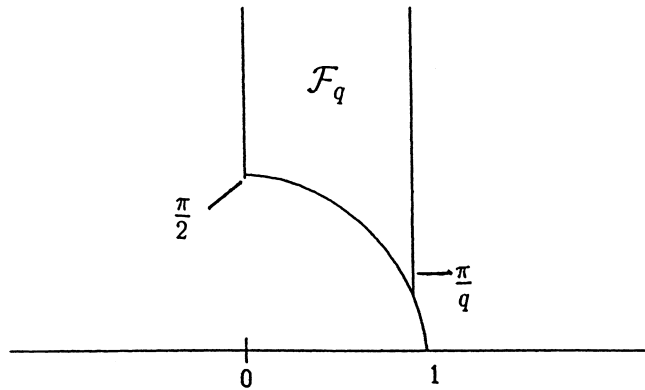


FIGURE 4.

By looking at more general surfaces X_Γ we have learned something about the Maass forms ϕ for $X(N)$. They are very fragile and special objects, and even their existence is tied to the arithmetic of $\Gamma(N)$. We end this section by pointing out that some higher rank cases of essential cuspidality have been established recently. For $X = SL(3, \mathbb{Z}) \backslash SL(3, \mathbb{R}) / SO(3)$ this is done in [Mi] and for the general congruence quotient of $SL(n, \mathbb{R}) / SO(n)$ in [Mu].

§3. LOW-ENERGY SPECTRUM

We mentioned in Section 1 in connection with Galois representations the importance of the eigenvalue $\lambda = \frac{1}{4}$ for $X(N)$. It was also noted that in analytic applications an eigenvalue $0 < \lambda_\phi < \frac{1}{4}$ would have a drastic impact on the cancellations in the sums (21). The question of the smallest non-zero eigenvalue for these surfaces is one of the major problems in the subject. Let $\lambda_1(X)$ denote the next to smallest (the smallest is $\lambda_0 = 0$) eigenvalue of Δ on $L^2(X)$.

Conjecture 2 ([Sel1]). For $N \geq 1$, $\lambda_1(X(N)) \geq \frac{1}{4}$.

We call this the Selberg-Ramanujan Conjecture since it was first formulated in [Sel1], but today we understand it as part of the general Ramanujan Conjectures [La2].

Comments.

- (1) As noted in Section 2 there is no residual spectrum for $X(N)$, so we could take for $\lambda_1(X(N))$ the smallest eigenvalue of a Maass *cuspidal* form for $X(N)$. The continuous spectrum is the interval $[\frac{1}{4}, \infty)$, so we may also formulate the conjecture variationally as follows:

For any $f \in C_0^\infty(X(N))$ that is smooth and compactly supported and for which

$$\int_{X(N)} f(z) dA(z) = 0$$

we have

$$(29) \quad \int_{X(N)} |\nabla_H f(z)|^2 dA(z) \geq \frac{1}{4} \int_{X(N)} |f(z)|^2 dA(z).$$

- (2) Another reason for the relevance of the number $\frac{1}{4}$ is that the spectrum of Δ on the universal covering $L^2(\mathbb{H})$ is $[\frac{1}{4}, \infty)$. To see that $\lambda_0(\mathbb{H}) \geq \frac{1}{4}$ let $f \in C_0^\infty(\mathbb{H})$. Following [Mc] we have

$$\begin{aligned} \frac{1}{2} \int_0^\infty f^2(z) \frac{dy}{y^2} &= \int_0^\infty f_y(z) f(z) \frac{dy}{y} \\ &\leq \left(\int_0^\infty (f_y(z))^2 dy \right)^{1/2} \left(\int_0^\infty f^2(z) \frac{dy}{y^2} \right)^{1/2}. \end{aligned}$$

Hence

$$\frac{1}{4} \int_0^\infty f^2(z) \frac{dy}{y^2} \leq \int_0^\infty |f_y(z)|^2 dy \leq \int_0^\infty |\nabla f|^2 dy.$$

Integrating with respect to x yields

$$\frac{1}{4} \int_{\mathbb{H}} f^2(z) dA(z) \leq \int_{\mathbb{H}} |\nabla f|^2 dx dy = \int_{\mathbb{H}} |\nabla_H f|^2 dA(z).$$

- (3) It is not difficult to show that the cuspidal spectrum of Δ becomes dense in $[\frac{1}{4}, \infty)$ as $N \rightarrow \infty$. Thus Conjecture 2 if true is sharp.
- (4) Conjecture 2 is somewhat surprising. After all the surfaces $X(N)$ have their area and genus going to infinity with N . This might lead one to expect that the low “overtone” $\lambda_1(X(N))$ should go to zero. That this is not the case is combinatorially very powerful. The optimally highly connected but sparse Ramanujan Graphs [L-P-S] and [M] are constructed via similar congruence quotients of the p -adic groups $PGL(2, \mathbb{Q}_p)$.
- (5) The assumption that Γ is a congruence subgroup of $SL(2, \mathbb{Z})$ cannot be dropped. In [Sel1] it is shown that given $\epsilon > 0$ there is a cyclic cover $X_{\Gamma'}$ of $X(2)$ for which $\lambda_1(X_{\Gamma'}) < \epsilon$. In view of the results below towards Conjecture 2 it follows that for ϵ small Γ' cannot contain $\Gamma(M)$ for any M . These Γ' are finite index subgroups of $SL(2, \mathbb{Z})$ which are not congruence subgroups.
- (6) Conjecture 2 is known for $N \leq 17$ [Hu]. The combinatorial topological method used there when successful establishes that $\lambda_1^{\text{cusp}}(X(N)) > \frac{1}{4}$. In view of the presence of Galois representations this method breaks down for large N .

In [Sel2] the bound $\lambda_1(X(N)) \geq \frac{3}{16}$ was established. It is already very useful in many applications. By a completely different method, it is shown in [Ge-Ja] that $\lambda_1(X(N)) > \frac{3}{16}$. Their approach is based on the symmetric square $\text{sym}^2 : GL(2) \rightarrow GL(3)$ functorial lift (see Appendix 1). They then invoke local representation theoretic bounds concerning generic representations of $GL(3, \mathbb{R})$ [Ja-Sh]. In [L-R-S] global bounds towards the general Ramanujan Conjectures on $GL(m, \mathbb{A}_{\mathbb{Q}})$, $m \geq 2$, are established using families of L -functions. Combining this for the case of $GL(3)$ together with sym^2 as above leads to $\lambda_1(X(N)) \geq \frac{21}{100}$. Recently, the functorial lifts $\text{sym}^3 : GL(2) \rightarrow GL(4)$ and $\text{sym}^4 : GL(2) \rightarrow GL(5)$ were established [K-S1] [K]. Combining these with the methods from families of L -functions, one obtains

Theorem 2 ([Ki-Sa]). $\lambda_1(X(N)) \geq \frac{975}{4096} = 0.238\dots$

This is getting close to $\frac{1}{4}$, but it is also close to the limit of these methods. The functorial lifts sym^3 and sym^4 are based on the continuous spectrum (Eisenstein series) on exceptional groups including E_8 . What can be done this way terminates with the finite list of exceptional groups. We see clearly here that other groups and symmetric spaces play a central role in understanding $GL(2)$ (that is essentially the upper half-plane). We note that if the general Functoriality Conjectures concerning the automorphic lifts $\text{sym}^k : GL(2) \rightarrow GL(k+1)$, $k \geq 1$, are valid, then so is Conjecture 2 [La2].

Conjecture 2 is concerned with the bottom of the spectrum of the Laplacian Δ on $L^2_{\text{cusp}}(X(N))$, that is the spectrum at the “archimedean” place. There is a similar conjecture for each prime p for the corresponding Hecke operator T_p . The surfaces $X(N)$ carry algebraic correspondences which give rise to the family of Hecke operators. For $(n, N) = 1$ define T_n by

$$(30) \quad T_n f(z) = \frac{1}{\sqrt{n}} \sum_{\substack{ad=n \\ b \pmod{d}}} f\left(\frac{az+b}{d}\right).$$

These are closely related to the cosets of the finite index subgroups

$$\begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} \Gamma(N) \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix}^{-1} \cap \Gamma(N) \text{ of } \Gamma(N).$$

The linear transformation T_n maps $L^2(X(N)) \rightarrow L^2(X(N))$. Moreover,

(31) the T_n 's are normal, commute with each other and with Δ , and leave $L^2_{\text{cusp}}(X(N))$ invariant.

The Ramanujan Conjectures for these T_p 's asserts the following:

$$(32) \quad \left\| T_p \Big|_{L^2_{\text{cusp}}(X(N))} \right\| \leq 2.$$

One can show that the above norm is at least 2 so that (32) is equivalent $\| T_p |_{L^2_{\text{cusp}}(X(N))} \| = 2$. The trivial bound (from (29)) for $\| T_p \|$ is

$$(33) \quad \| T_p \| \leq p^{\frac{1}{2}} + p^{-\frac{1}{2}}.$$

The functorial techniques [K] apply to these finite places as well, and if one combines them with the L -function techniques of [Du-Iw], one obtains the following very useful bounds [Ki-Sa]:

$$(34) \quad \| T_p |_{L^2_{\text{cusp}}(X(N))} \| \leq p^{\frac{7}{64}} + p^{-\frac{7}{64}}.$$

A unification and clarification of the theory of the spectrum of Δ and of T_p are provided by the representation theoretic description of the subject, that is the spectral decomposition of the right regular action of $GL(2, \mathbb{A}_{\mathbb{Q}})$ on $GL(2, \mathbb{Q}) \backslash GL(2, \mathbb{A}_{\mathbb{Q}})$ where $\mathbb{A}_{\mathbb{Q}}$ is the adèle group of \mathbb{Q} [Ge]. This language is indispensable for many purposes. We have avoided it in order to keep our discussion as self-contained and elementary as possible. A recent discussion of these adelic spectral problems for more general groups can be found in [Cl].

§4. HIGH-ENERGY SPECTRUM

In this section we examine the solutions (13) for λ large. For this limit the case of $X(1)$ already contains the key features, so for the most part we will concentrate on $X(1)$. Probably the simplest and most basic question concerning the eigenvalues of $X(1)$ was raised in [Ca], where the first attempts to numerically compute the eigenvalues were carried out.

Conjecture 3 ([Ca]). The (cuspidal) spectrum of Δ on $X(1)$ is simple.

The main evidence for this consists of the numerical computations [St] which confirm the conjecture for the first 10,000 eigenvalues. Let $m_{X(1)}(\lambda)$ be the multiplicity of the eigenvalue λ . The best known bounds for $m(\lambda)$ are rather poor. They are derived from the trace formula by analyzing the remainder term in the Weyl asymptotics (25). The difficulty in all questions involving the large λ limit is that when localizing a test function in the trace formula (Appendix 3) near λ , one is faced on the dual side (via Fourier Transform) with an exponential in λ number of terms involving closed geodesics. It is then very problematic to establish suitable cancellations in these sums over the closed geodesics. All that seems possible is to treat these sums trivially, and after optimizing the choice of test functions one obtains a bound for $m_X(\lambda)$ for the general hyperbolic surface X :

$$(35) \quad \overline{\lim}_{\lambda \rightarrow \infty} \frac{m_X(\lambda) \log \lambda}{\sqrt{\lambda}} \leq \frac{\text{Area}(X)}{2}.$$

For $X(1)$ one can obtain square root cancellation in the relevant sums by using the relative trace formula [Ku] which involves Kloosterman sums instead of closed geodesics. This leads to a modest improvement over (35) for $X(1)$ [Sa3]:

$$(36) \quad \overline{\lim}_{\lambda \rightarrow \infty} \frac{m_{X(1)}(\lambda) \log \lambda}{\sqrt{\lambda}} \leq \frac{\text{Area}(X(1))}{4} = \frac{\pi}{12}.$$

It is an important problem to improve this bound (36). In my youth I would have been impressed with the bound $m_{X(1)}(\lambda) = O(\lambda^{\frac{1}{2}-\delta})$ for some $\delta > 0$. Today $m_{X(1)}(\lambda) = o(\sqrt{\lambda}/\log \lambda)$ looks good.

An interesting phenomenological fact about the spectrum of $X(1)$ was noted in [St]. Consider the numbers $\tilde{\lambda}_j = \lambda_j/12$ where λ_j are the eigenvalues of $X(1)$. According to (25) the mean spacing between the numbers $\tilde{\lambda}_j$ is 1. The numerical experiments in [St] suggest that the consecutive spacings and for that matter any local spacing statistic behave like random numbers would, or as is often said, the local spacing statistics are ‘‘Poissonian’’. More precisely, the consecutive spacings apparently obey the following law: For $0 \leq \alpha < \beta < \infty$,

$$(37) \quad \text{as } N \rightarrow \infty, \quad \frac{\#\{j \leq N \mid \tilde{\lambda}_{j+1} - \tilde{\lambda}_j \in [\alpha, \beta]\}}{N} \rightarrow \int_{\alpha}^{\beta} e^{-x} dx.$$

This random behavior of the eigenvalues of $X(1)$ is unexpected. To explain why we recall the interest that the spectrum of a hyperbolic surface has generated in mathematical physics. Given $x \in X$ and $\xi \in T_x(X)$ a unit tangent vector and $t \in \mathbb{R}$, let $(x(t), \xi(t))$, $\xi(t) \in T_{x(t)}(X)$ be the position and tangent vector after flowing for arc-length t from x in the direction ξ along the corresponding geodesic. This flow $\mathcal{G}_t : S(X) \rightarrow S(X)$ on the unit circle bundle to X is the geodesic flow.

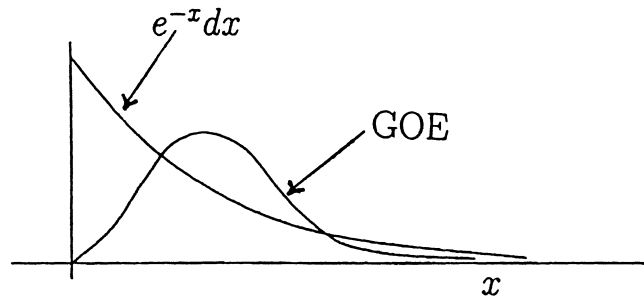


FIGURE 5.

It is a Hamilton flow. In terms of the uniformization $X = \Gamma \backslash \mathbb{H}$ we can identify $S(X)$ with $\Gamma \backslash G$, $G = PSL(2, \mathbb{R})$, and \mathcal{G}_t takes the form

$$(38) \quad \Gamma g \xrightarrow{\mathcal{G}_t} \Gamma g \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}.$$

\mathcal{G}_t clearly preserves the Haar measure dg .

It is well known that the geodesic flow of a manifold of negative curvature is ergodic,⁷ Anosov, has positive Lyapunov exponents In short, it is a “chaotic” Hamiltonian. Now the eigenvalue problem for Δ on X_Γ corresponds to a quantization of this classically chaotic dynamics. The large λ limit for this quantization is the same as the semi-classical limit ($\hbar \rightarrow 0$) of the quantum system. Unlike the case of the quantization of a completely integrable Hamiltonian where the impact of classical invariant tori on the spectrum in the semi-classical limit is well understood [Laz], very little can be said analytically in the chaotic case. The study of the semi-classical limit, specifically the relation between the classical and quantum mechanics of a classically chaotic Hamiltonian, goes by the name of Quantum Chaos.

One of the interesting suggestions that has emerged from many numerical investigations is that the local spacing statistics of the eigenvalues of the quantization of a classically chaotic Hamiltonian are modeled by the local spacing laws of the eigenvalues of a random matrix in a suitable matrix ensemble. These laws are very different from the Poissonian laws that the spectrum of $X(1)$ exhibits. This then is the sense in which the spectrum of $X(1)$ has unexpected statistics. No doubt the reason for this anomaly is that Δ commutes with the geometrically defined Hecke operators T_n in (30) (see also [Sa4]). Indeed, for the spectrum for the Dirichlet problem (which has discrete spectrum only) for the triangles \mathcal{F}_q in Figure 4, it is found numerically that for $q \neq 3, 4, 6$ (in which case there are no Hecke operators) the local spacings follow the local “GOE laws” [B-S]. GOE stands for the Gaussian Orthogonal Ensemble and is concerned with the eigenvalues of a random real symmetric matrix of large size [Me]. For example, the scaled consecutive spacing distribution for the latter is depicted in Figure 5. The density of the GOE consecutive spacing law vanishes to first order at $x = 0$, indicating that the eigenvalues of a random matrix repel each other and that the corresponding spectrum is rigid. In contrast there are near degeneracies in the Poissonian laws. This phenomenology about the spectrum of $X(1)$ is fascinating, but very little can be proven.

⁷The first case to be studied was in fact $X(1)$, by Artin [Ar].

We turn to the behavior of the eigenfunctions ϕ_λ for $X(1)$ for which much is now known. The basic issue is whether as $\lambda \rightarrow \infty$ these eigenstates can localize, or do they spread out evenly. For reasons similar to those mentioned above in connection with the question of multiplicities, it is very difficult to analyze the high-energy eigenstates of the quantization of a classically chaotic Hamiltonian. We next state the basic conjectures asserting the non-localization of ϕ_λ for a general hyperbolic X (possibly compact) as $\lambda \rightarrow \infty$.

Conjecture 4 ([I-S1]). Fix X , $K \subset X$ a compact (nice set),⁸ $2 < p \leq \infty$ and $\epsilon > 0$. There is $c = c(p, K, \epsilon)$ such that

$$\left(\int_K |\phi_\lambda(z)|^p dA(z) \right)^{\frac{1}{p}} \leq c\lambda^\epsilon \left(\int_K |\phi_\lambda(z)|^2 dA(z) \right)^{\frac{1}{2}}.$$

Remarks. (1) This conjecture asserts that the L^p norms on a nice compact set $K \subset X$ (if X is compact, then take $K = X$) grow slower than any power of λ times the L^2 norm. It quantifies the lack of localization of ϕ_λ . For the case of $p = \infty$ this conjecture, if true, is quite deep since it implies the classical Lindelof Hypothesis for the Riemann Zeta Function (Appendix 1). Also, for $p = \infty$ it implies that $m_X(\lambda) = O_\epsilon(\lambda^\epsilon)$.

(2) The λ^ϵ factor is necessary when $p = \infty$ since in [I-S1] it is shown that $\|\phi_\lambda\|_\infty$ is not uniformly bounded for certain compact X . Here (and henceforth) we will normalize ϕ_λ so that $\|\phi_\lambda\|_2 = 1$.

There are some general interpolation bounds for the L^p norms of eigenfunctions ϕ_λ on a (fixed) general compact Riemannian surface X . In [So] it is shown that

$$(39) \quad \|\phi_\lambda\|_p \ll \lambda^{\delta(p)},$$

where $\delta(p) = \frac{1}{4} - \frac{1}{p}$, for $6 \leq p \leq \infty$ and $\delta(p) = \frac{1}{8} - \frac{1}{4p}$, for $2 \leq p \leq 6$. The proof of (39) uses a construction of a parametrix for the wave equation via Fourier Integral Operators, combined with ideas from restriction theorems in Fourier Analysis [Ste]. The bound (39) is local in that it is derived on pieces of X directly. The global aspects of the geodesic motion do not enter. As a local bound (39) cannot be improved since it is sharp on S^2 with the round metric [So].

Another means to analyze the localization question is to examine the probability measures

$$(40) \quad \mu_\phi = |\phi_\lambda(z)|^2 dA(z).$$

Quantum mechanically this gives the probability distribution on X associated to being in the state ϕ_λ . One can form a probability measure ν_ϕ on $\Gamma \backslash G$, the micro-local lift of μ_ϕ to $S(X)$ (Appendix 4), which projects to μ_ϕ and measures how ϕ_λ is distributed in phase space. For $a \in C^\infty(\Gamma \backslash G)$, $\nu_\phi(a)$ measures the quantum observable $Op(a)$ (see Appendix 4) when in the state ϕ .

Conjecture 5 ([R-S] Quantum Unique Ergodicity). The measures ν_ϕ become equidistributed with respect to dg as $\lambda \rightarrow \infty$. Precisely, if $f \in C_0(\Gamma \backslash G)$, then

$$\lim_{\lambda \rightarrow \infty} \int_{\Gamma \backslash G} f(g) d\nu_\phi(g) = \int_{\Gamma \backslash G} f(g) dg := \bar{f}$$

where dg is a Haar measure on G normalized so that $\text{Vol}(\Gamma \backslash G) = 1$.

⁸For example K could be the closure of a non-empty geodesic ball.

Comments.

- (1) While this conjecture seems reasonable enough, we point out that it contradicts some suggestions that eigenstates in chaotic quantizations might concentrate on unstable periodic orbits, a phenomenon called scarring [Hel].
- (2) The name “quantum unique ergodicity” stems from there being in this context an analogue at the quantum level of ergodicity. Let ϕ_j be any orthonormal basis of $L^2(X)$ (if X is not compact, then we assume that $X = X(N)$ and that ϕ_j is an orthonormal basis of $L^2_{\text{cusp}}(X)$). In [Shn], [Co], [Ze] it is shown that if $f \in C^\infty(\Gamma \backslash SL(2, \mathbb{R}))$, then as $\lambda \rightarrow \infty$

$$(41) \quad \sum_{\lambda_j \leq \lambda} |\nu_{\phi_j}(f) - \bar{f}|^2 = o\left(\sum_{\lambda_j \leq \lambda} 1\right).$$

In particular it follows that almost all, in the sense of density of the number of eigenvalues, of the ν_ϕ 's become equidistributed as $\lambda_j \rightarrow \infty$. Recall that \mathcal{G}_t being ergodic means that almost all orbits of the flow become equidistributed as $t \rightarrow \infty$. Thus the above is the quantum analogue of the geodesic flow \mathcal{G}_t being ergodic. For \mathcal{G}_t however there are many singular invariant measures (the most singular being arclength on an unstable periodic orbit). Thus, Conjecture 5 asserts that at the quantum level things are quite different in that all eigenstates become equidistributed (a flow for which all orbits become equidistributed is called uniquely ergodic).

- (3) We call any weak limit of the measures ν_{ϕ_λ} a quantum limit. Using some standard results about propagation of singularities in the theory of Fourier Integral Operators, one can show that any quantum limit is \mathcal{G}_t invariant; see Appendix 4. Conjecture 5 is equivalent to the statement that the only quantum limit is dg .

For the general hyperbolic surface X little has been proven towards Conjectures 4 and 5. However, for $X(1)$ or more generally $X(N)$ there has been some decisive progress. We restrict our discussion to $X(1)$. In view of (30) we can simultaneously diagonalize Δ and the operators T_n , $n \geq 1$. Henceforth we assume that our ϕ_λ is also a Hecke eigenform:

$$(42) \quad T_n \phi_\lambda = \lambda_\phi(n) \phi_\lambda.$$

Note that if Conjecture 3 is true, then (42) is automatic. In any case it is the Maass-Hecke eigenform that is of interest.

All these questions about the ϕ_λ 's for $X(1)$ make sense for the continuous spectrum as well. Explicitly the Eisenstein series $E(z, s)$ for $X(1)$ is defined as

$$(43) \quad E(z, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma(1)} (y(\gamma z))^s \quad \text{for } \Re(s) > 1.$$

$E(z, s)$ extends meromorphically to \mathbb{C} and is analytic on $\Re(s) = \frac{1}{2}$. The continuous spectrum for $X(1)$ is furnished by the generalized eigenfunctions $E(z, \frac{1}{2} + it)$, $t \geq 0$,

$$(44) \quad \Delta E\left(z, \frac{1}{2} + it\right) + \left(\frac{1}{4} + t^2\right) E\left(z, \frac{1}{2} + it\right) = 0$$

and of course

$$E(\gamma z, s) = E(z, s) \quad \text{for } \gamma \in \Gamma(1).$$

Concerning Conjecture 4 for the case $p = \infty$ a sub-convex (or sub-interpolation) bound for $X(1)$ is established in [I-S1]:

$$(45) \quad \|\phi_\lambda\|_\infty \ll \lambda^{5/24}.$$

A similar bound for $E(z, \frac{1}{2} + it)$ is also proven.

In [L-S1] and [Ja] Conjecture 5 is proven for the continuous spectrum of $X(1)$. Precisely, let μ_t be the corresponding measure (of infinite mass)

$$(46) \quad \mu_t = \left| E\left(z, \frac{1}{2} + it\right) \right|^2 dA(z).$$

Let ν_t be its micro-local lift to $\Gamma(1)\backslash G$. Then for K_1 and K_2 (nice) compact subsets of $\Gamma(1)\backslash G$ we have that

$$(47) \quad \lim_{t \rightarrow \infty} \frac{\nu_t(K_1)}{\nu_t(K_2)} = \frac{\text{Vol}_{dg}(K_1)}{\text{Vol}_{dg}(K_2)}.$$

For the measures ν_ϕ , Conjecture 5 has proven to be much more difficult to attack via the methods of [L-S1]. In [R-S] it is shown that no quantum limit ν on $\Gamma(1)\backslash G$ can be supported on a finite union of closed geodesics. In particular this strong form of scarring is not possible for $X(1)$. In [B-L] a significant extension of this is given; the authors show that if ν is a quantum limit, then it must have positive entropy for the geodesic flow \mathcal{G}_t .

An identity is derived in [Wa] which allows one to unify and make explicit the relation between triple products and special values of L -functions [H-K], [L-S1]. This allows one to convert some of these questions concerning eigenfunctions to ones concerning the size of L -functions at special points on their critical lines. The precise identity is as follows:

Let ϕ_1, ϕ_2, ϕ_3 be three Maass (Hecke) eigenforms on $X(1)$ normalized as before so that $\|\phi_j\|_2 = 1$. Then [Wa]

$$(48) \quad 216 \cdot \left| \int_{X(1)} \phi_1(z) \phi_2(z) \phi_3(z) dA(z) \right|^2 = \frac{\pi^4 \cdot \Lambda\left(\frac{1}{2}, \phi_1 \times \phi_2 \times \phi_3\right)}{\Lambda(1, \text{sym}^2 \phi_1) \Lambda(1, \text{sym}^2 \phi_2) \Lambda(1, \text{sym}^2 \phi_3)}.$$

Here $\Lambda(s, \phi_1 \times \phi_2 \times \phi_3)$ is the (completed) degree 8 L -function associated with ϕ_1, ϕ_2, ϕ_3 , and $s = \frac{1}{2}$ is its central value (Appendix 1). $\Lambda(s, \text{sym}^2 \phi)$ is the degree 3 L -function (Appendix 1), and $s = 1$ is at the edge of the critical strip.

Using (48) one can reformulate Conjecture 5 in terms of estimates for L -functions. For example the Lindelof Hypothesis (see [I-S2] and Appendix 1), which is a consequence of the Generalized Riemann Hypothesis for these degree 8 L -functions, implies the following strong form of Conjecture 5: For a fixed $f \in C_0^\infty(\Gamma(1)\backslash G)$ and $\epsilon > 0$

$$(49) \quad |\nu_{\phi_\lambda}(f) - \bar{f}| \ll_{\epsilon, f} \lambda^{-1/4+\epsilon}.$$

In fact, a “sub-convex” (Appendix 1) estimate for these L -functions would already establish Conjecture 5. While there has been much progress on the general sub-convexity problem for L -functions, the general case at hand remains out of reach at the present time. We also note that the Lindelof Hypothesis for these degree 8 L -functions implies Conjecture 4 with $p = 4$.

Recently, there have been a series of breakthroughs which has led to the solution of parts of Conjectures 4 and 5. These results are still being written up, so they

should be regarded with appropriate caution until they have been independently confirmed.

Theorem 3 ([Sa-Wa], [Sp]). *Fix $\epsilon > 0$; then*

- (a) $\|\phi_\lambda\|_4 \ll_{\epsilon} \lambda^\epsilon$.
- (b) For $K \subset X(1)$ compact

$$\left(\int_K \left| E\left(z, \frac{1}{2} + it\right) \right|^4 dA(z) \right)^{1/4} \ll_{\epsilon, K} (1 + |t|)^\epsilon \left(\int_K \left| E\left(z, \frac{1}{2} + it\right) \right|^2 dA(z) \right)^{1/2}.$$

These establish Conjecture 4 for $X(1)$ for $2 < p \leq 4$, for both the discrete and continuous spectrum. Combining these sharp L^p bounds with (45) and interpolating yields subconvex bounds for all $2 \leq p \leq \infty$. In Figure 6 the exponent $\delta(p)$ of λ in these bounds is graphed against $1/p$. The solid lines correspond to the convex bound (39) and the dashed lines to the subconvex bounds. A random wave model for the eigenfunctions of the quantization of a classically chaotic Hamiltonian is put forth in [Be]. In [H-R] this is tested numerically for $X(1)$ as far as the behavior of the value distribution of $\phi_\lambda(z)$ and $E(z, \frac{1}{2} + it)$ as the energy goes to infinity. In particular they find a Gaussian behavior. Thus we expect the odd moments of ϕ_λ and E to go to zero and the even moments (at least in the form (b)) to remain bounded. For special ϕ_λ 's (the dihedral ones) on $X(N)$ one can prove this uniform boundedness of the L_4 norms. Moreover, (48) applied to the case $\phi_1 = \phi_2 = \phi_3 = \phi_\lambda$ together with known sub-convexity bounds for degree 2 L -functions [Iw1] shows that the third moment $\int_{X(1)} \phi_\lambda^3 dA$ goes to zero as $\lambda \rightarrow \infty$ [Wa].

Theorem 4 ([Li1]). *Let ν be a quantum limit for the ϕ_λ 's on $\Gamma(1)\backslash G$; then $\nu = c.dg$ with $0 \leq c \leq 1$.*

This essentially establishes Conjecture 5 for $X(1)$. The only missing point is that c may be less than 1 if for example the mass of the ν_ϕ 's escape into the cusp of $X(1)$ as $\lambda \rightarrow \infty$. In the case of a compact arithmetic quotient X_Γ of \mathbb{H} (and ϕ_λ an eigenfunction of the corresponding Hecke operators) the last difficulty does not arise and Conjecture 5 is completely established for X_Γ [Li1].

Finally, for $X(1)$ one can determine the main term for the asymptotics in the quantum ergodicity sums (introduced in [Ze]) of equation (41).

Theorem 5 ([L-S2]). *There is quadratic form $B(f)$ on $C_0^\infty(\Gamma(1)\backslash\mathbb{H})$ such that*

- (a) $\sum_{\lambda_j \leq \lambda} |\mu_{\phi_j}(f) - \bar{f}|^2 = B(f) \sqrt{\lambda} + o(\sqrt{\lambda})$ as $\lambda \rightarrow \infty$.
- (b) *The polarization of B satisfies*

$$B(f_1, \Delta f_2) = B(\Delta f_1, f_2).$$

(c) *Defining the operator A by*

$$\langle Af_1, f_2 \rangle = B(f_1, f_2)$$

extends A to a non-negative self-adjoint operator which commutes with Δ and is diagonalized by the $E(z, \frac{1}{2} + it)$ and ϕ_λ 's. Moreover, the eigenvalue of

⁹At present the proof of (a) which is involved uses Conjecture 2 and (32) freely. We expect in the end to get around this.

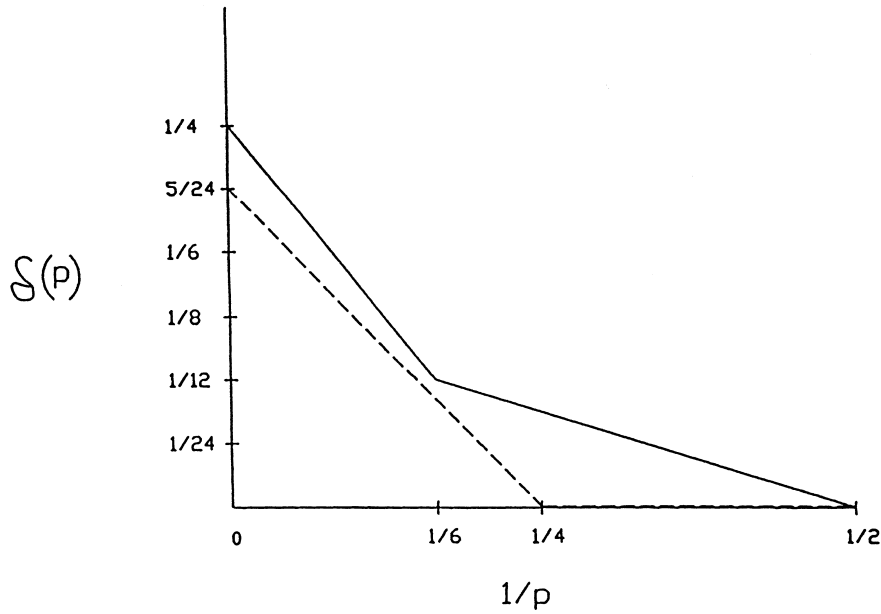


FIGURE 6.

A corresponding to ϕ_λ is essentially $L(\frac{1}{2}, \phi_\lambda)$, where $L(s, \phi)$ is the standard L -function associated with ϕ (Appendix 1).

Thus form B provides a non-negative self-adjoint operator on a Hilbert space whose eigenvalues are special values of a family of L -functions! In particular, $L(\frac{1}{2}, \phi) \geq 0$, a fact that was known by other means (see [L-R] for the most general such non-negativity result). Since it is known that many of the values $L(\frac{1}{2}, \phi)$ are non-zero, it follows that for most f , $B(f) > 0$. In particular, this shows that the decay rate for Conjecture 1 predicted by the Lindelof Hypothesis in (49) is sharp (at least up to the exponent ϵ). In Appendix 5 we give a comparison of (a) with the variation of f along the geodesic flow. The discrepancy between the classical and quantum fluctuations is given by the arithmetic factor $L(\frac{1}{2}, \phi)$!

We make some brief comments about the methods used to prove these recent results. Theorem 3 is approached by using Parseval's identity to express $\|\phi_\lambda\|_4^4$ as follows:

$$(50) \quad \|\phi_\lambda\|_4^4 = \sum_j |\langle \phi_\lambda^2, \phi_j \rangle|^2 + \text{similar term from continuous spectrum.}$$

Now apply (48) to the terms $\langle \phi_\lambda^2, \phi_j \rangle$, which converts the j sum to a sum of degree 8 L -functions. The recent advances in the theory of families of L -functions (see [I-S2]) and especially the methods for averaging over such families using various trace formulae can be applied. However, being of degree 8, there are a number of new difficulties that need to be overcome. One useful technical device that we mention and which is used is the recent $GL(3)$ Voronoi Summation Formula [M-S]. In any case, suffice it to say that the key techniques used to prove Theorems 3 and 5 are those from the theory of L -functions.

The approach in [Li1] to Theorem 4 is very different. We know that ν is \mathcal{G}_t invariant but that this is far from sufficient to identify ν . The idea is to try to use the fact that the ϕ_λ 's are also eigenfunctions of the Hecke operators T_p . Rather than describe the case of $X(1)$, consider a higher dimensional case which is conceptually simpler. Let $\Gamma \leq G \times G$ be an irreducible lattice such as $SL(2, \mathbb{Z}[\sqrt{2}])$ embedded diagonally in $G \times G$ via $\gamma \longrightarrow (\gamma, \gamma')$, γ' being the Galois conjugate of γ . There is a similar theory of Maass cusp forms $\phi_{\lambda_1, \lambda_2}(z_1, z_2)$ on the Hilbert modular manifold $X = \Gamma \backslash \mathbb{H} \times \mathbb{H}$. Such a form is an eigenfunction of Δ_{z_j} with eigenvalue λ_j , $j = 1, 2$. The Laplacian on X is $\Delta_{z_1} + \Delta_{z_2}$. One can construct a micro-local lift ν_ϕ on $\Gamma \backslash G \times G$, of $|\phi(z_1, z_2)|^2 dv(z_1, z_2)$; see [Li2]. Being an eigenfunction of both Δ_z and Δ_{z_2} (which commute), one can show that a quantum limit, that is a weak limit of the ν_ϕ 's, is invariant under the two parameter Cartan action on $\Gamma \backslash G \times G$:

$$(51) \quad \Gamma g \longrightarrow \Gamma g \left(\left(\begin{array}{cc} e^{t_1/2} & 0 \\ 0 & e^{-t_1/2} \end{array} \right), \left(\begin{array}{cc} e^{t_2/2} & 0 \\ 0 & e^{-t_2/2} \end{array} \right) \right).$$

Unlike the case of the geodesic flow \mathcal{G}_t there are fewer measures invariant under such two (or higher parameter) actions. This is the so-called measure rigidity phenomenon that has seen many advances recently [Ru], [Ka-Sp], [E-K]. The flow (51) does not fall into the setting of these works, and in [Li1] a substitute theory is developed. One point worth noting is that progress on these measure rigidity questions has only been possible by assuming that such an ergodic invariant measure has positive entropy. As mentioned earlier for $X(1)$ any quantum limit has been shown to have positive entropy.

It is interesting that the problems discussed in these notes can be approached from such different points of view. Moreover, having recast the problem in different terms (for example the Hilbert problem on page 11 as a sub-convexity problem for automorphic L -functions (see [I-S2]), or Conjecture 5 as a measure rigidity problem), one finds that the solution demands a significant advance in the corresponding theory. Thus both sides are enriched.

To end, we point the reader to books [Ve], [He2] and [Iw2], which treat the basic material concerning hyperbolic surfaces, the trace formula and Eisenstein series. Also, the books [Lan1] and [Bor] give introductions to the approach via representation theory of $SL(2, \mathbb{R})$, and [L-P] via scattering theory for the corresponding wave equation. An earlier review article of some of the material discussed in the lectures can be found in [Sa4].

APPENDICES

These appendices are meant to give brief descriptions and definitions of various objects that were mentioned in the text. Detailed treatments can be found in the references.

Appendix 1: L -Functions. The analytic continuation and functional equation for the Riemann Zeta Function were mentioned in (3) and (4). The key tool used there, Poisson Summation, can also be used to analytically continue Dirichlet's L functions, $L(s, \chi)$, and their generalizations to number fields, Hecke's L -functions, $L(s, \lambda)$. A Dirichlet character χ is a function from \mathbb{Z} into \mathbb{C} which is periodic of (minimal) period $q \geq 1$ and which satisfies $\chi(mn) = \chi(m)\chi(n)$, $\chi(1) = 1$ and

$\chi(m) = 0$ if $(m, q) > 1$. The corresponding L -function is defined to be

$$(1) \quad L(s, \chi) = \sum_{n=1}^{\infty} \chi(n)n^{-s} = \prod_p (1 - \chi(p)p^{-s})^{-1}.$$

The completed L -function $\Lambda(s, \chi)$ is defined by:

$$(2) \quad \Lambda(s, \chi) = \pi^{-(s+a_\chi)/2} \Gamma\left(\frac{s+a_\chi}{2}\right) L(s, \chi),$$

where $a_\chi = \frac{1-\chi(-1)}{2}$. $\Lambda(s, \chi)$ is entire (if $\chi \neq 1$) and satisfies the functional equation [Da]:

$$(3) \quad \Lambda(s, \chi) = \frac{\tau(\chi)}{i^{a(\chi)}q^{1/2}} q^{\frac{1}{2}-s} \Lambda(1-s, \bar{\chi})$$

where $\tau(\chi)$ is the Gauss sum; q is called the conductor of χ .

The Hecke L -functions are defined in a similar way [Hecl]:

$$(4) \quad L(s, \lambda) = \sum_{\mathcal{A} \neq 0} \lambda(\mathcal{A})N(\mathcal{A})^{-s} = \prod_P (1 - \lambda(P)N(P)^{-s})^{-1}$$

where λ is a suitable character on the ideals of a number field K , \mathcal{A} ranges over the non-zero integral ideals, P over the prime ideals and $N(\mathcal{A})$ is the norm of \mathcal{A} . For us an interesting example is $K = \mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. The ring of integers of K , O_K is simply equal to $\{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$. The units in O_K are generated by $\epsilon_0 = 1 + \sqrt{2}$. For $\alpha \in K$ let α' denote its Galois conjugate. O_K happens to have class number one; that is every ideal is principal. For $0 \neq m \in \mathbb{Z}$ set

$$(5) \quad \lambda^m(\alpha) = \left| \frac{\alpha}{\alpha'} \right|^{\frac{i\pi m}{\log \epsilon_0}}.$$

Clearly, $\lambda^m(\epsilon\alpha) = \lambda^m(\alpha)$ for any unit ϵ . So λ is a character on the ideals of O_K . It is an example of a ‘‘Größencharakter’’ of Hecke in that the values assumed by λ as α varies are dense in the circle. Maass [Mas] showed how these may be used to construct Maass forms. If

$$(6) \quad \phi_m(z) = \sum_{\mathcal{A} \neq 0} \lambda^m(\mathcal{A})y^{1/2} K_{it_m}(2\pi N(\mathcal{A})y) \cos(2\pi N(\mathcal{A})x)$$

with $t_m = \frac{\pi m}{\log \epsilon_0}$, then ϕ_m satisfied (13) for $\gamma \in \Gamma(4)$ and with eigenvalue $\lambda_m = \frac{1}{4} + t_m^2$. In particular, $X(4)$ has this explicit subsequence of eigenvalues (no explicit eigenvalues are known or expected for $X(1)$).

The Dirichlet L -functions are Euler products of degree 1 (over \mathbb{Q}) while the Hecke L -functions $L(s, \lambda)$ are Euler products of degree 1 over K . Euler products of higher degree are constructed from modular forms, with modularity replacing Poisson Summation in the proof of the analytic continuation. To illustrate this, let ϕ be a Maass-Hecke eigenform as described in (30) and (13) for $X(1)$. That is, $\phi \in L_{\text{cusp}}^2(X(1))$ and

$$(7) \quad \left. \begin{aligned} \Delta\phi + \left(\frac{1}{4} + t_\phi^2\right)\phi &= 0 \\ T_n\phi &= \lambda_\phi(n)\phi \end{aligned} \right\}$$

(note we changed to a more convenient parameter t_ϕ where $\lambda_\phi = \frac{1}{4} + t_\phi^2$). The (standard) L -function associated to ϕ is denoted by $L(s, \phi)$ and is defined for $\Re(s)$ large by

$$(8) \quad L(s, \phi) = \sum_{n=1}^{\infty} \lambda_\phi(n) n^{-s} = \prod_p (1 - \lambda_\phi(p) p^{-s} + p^{-2s})^{-1}.$$

The sum to product formula follows from a similar relation that is satisfied by the T_n 's [Hec2]. It is convenient to introduce the roots $\alpha_\phi^{(1)}(p)$ and $\alpha_\phi^{(2)}(p)$ which are determined by

$$(9) \quad \alpha_\phi^{(1)}(p) \alpha_\phi^{(2)}(p) = 1, \quad \alpha_\phi^{(1)}(p) + \alpha_\phi^{(2)}(p) = \lambda_\phi(p).$$

Thus (8) is the degree 2 Euler product

$$(10) \quad L(s, \phi) = \prod_p \left[(1 - \alpha_\phi^{(1)}(p) p^{-s}) (1 - \alpha_\phi^{(2)}(p) p^{-s}) \right]^{-1}.$$

The modularity of $\phi(z)$ (in this case of $X(1)$) is equivalent to $L(s, \phi)$ extending to an entire function and satisfying the functional equation

$$(11) \quad \begin{aligned} \Lambda(s, \phi) &:= \pi^{-s} \Gamma\left(\frac{s+it_\phi}{2}\right) \Gamma\left(\frac{s-it_\phi}{2}\right) L(s, \phi) \\ &= \Lambda(1-s, \phi) \end{aligned}$$

(we have assumed here that ϕ is unramified at ∞ , that is that ϕ is even with respect to the isometry of $X(1)$, $z \rightarrow -\bar{z}$).

Thus these Maass forms give Euler products of degree 2 over \mathbb{Q} . Note that $L(s, \phi_m) = L(s, \lambda^m)$ in (4), (5) and (6) above; that is, the degree 1 Euler product over the quadratic extension K of \mathbb{Q} is a degree 2 Euler product over \mathbb{Q} and corresponds to a modular form. In general, any automorphic form π on $GL_2(\mathbb{A}_K)$ gives an Euler product of degree 2 over K which has an analytic continuation and functional equation relating s to $1-s$ and π to its contragredient $\tilde{\pi}$ [Go-Ja]. More generally, if π is automorphic and cuspidal on $GL_n(\mathbb{A}_K)$, its standard L -function, $L(s, \pi)$, is of degree n and is entire [Go-Ja]. In fact, one of the main interests in these automorphic cusp forms π is that it is believed that all L -functions (for example Hasse-Weil L -functions, Artin L -functions, ...) can be expressed as finite products and quotients of such standard L -functions.

Next we discuss the formation of tensor power L -functions from these π 's. For these much less is known. We restrict to the ϕ 's in (7). Let ϕ_1, \dots, ϕ_ℓ be ℓ such Maass forms. Define the degree 2^ℓ tensor power function $L(s, \phi_1 \times \phi_2 \times \dots \times \phi_\ell)$ by

$$(12) \quad L(s, \phi_1 \times \phi_2 \times \dots \times \phi_\ell) = \prod_p L_p(s, \phi_1 \times \phi_2 \times \dots \times \phi_\ell)$$

where

$$(13) \quad L_p(s, \phi_1 \times \dots \times \phi_\ell) = \prod_{\substack{\epsilon_j \in \{1, 2\} \\ j=1, \dots, \ell}} \left(1 - \alpha_{\phi_1}^{(\epsilon_1)}(p) \alpha_{\phi_2}^{(\epsilon_2)}(p) \cdots \alpha_{\phi_\ell}^{(\epsilon_\ell)}(p) p^{-s} \right)^{-1}.$$

The completed L -function $\Lambda(s, \phi_1 \times \phi_2 \times \dots \times \phi_\ell)$ is defined as usual by tacking on the corresponding product of 2^ℓ Gamma factors as in (11). It is believed that these $\Lambda(s, \phi_1 \times \phi_2 \times \dots \times \phi_\ell)$ have analytic continuations (except for possible poles at $s = 0$ and 1) and functional equations. This is known to be valid for $\ell = 2$ and 3. The case $\ell = 2$ is known as the Rankin-Selberg L -function, and its value for s on the

critical line (that is $\Re(s) = \frac{1}{2}$) arose in (27) (though in (27) one of the forms is holomorphic rather than a Maass form, but the theory is the same). The analytic continuation and functional equation for $\ell = 3$ is due to [Ga] and [PS-R]. The special value at $s = \frac{1}{2}$ of these L -functions is at the heart of the identity (48).

If $\phi_1 = \phi_2 = \dots = \phi_\ell$, one is led to form the symmetric tensor power L -functions. Define the degree $\ell + 1$, L -function $L(s, \phi, \text{sym}^\ell)$ as follows:

$$(14) \quad L(s, \phi, \text{sym}^\ell) = \prod_p L_p(s, \phi, \text{sym}^\ell)$$

where

$$(15) \quad L_p(s, \phi, \text{sym}^\ell) = \prod_{j=0}^{\ell} (1 - (\alpha_\phi^{(1)}(p))^j (\alpha_\phi^{(2)}(p))^{\ell-j} p^{-s})^{-1}.$$

As before, one forms the corresponding completed function, $\Lambda(s, \phi, \text{sym}^\ell)$. Again, it is conjectured that these have analytic continuations and functional equations. The *meromorphic* continuation and functional equation are known for these for $\ell \leq 9$ [Sh]. The recent developments [K], [K-S1] and [K-S2] mentioned in Sections 1 and 2 establish that $\Lambda(s, \phi, \text{sym}^\ell)$ is entire (except perhaps for poles at $s = 0$ and 1) for $\ell = 3$ and 4 (the case $\ell = 2$ is due to [Shi2].) Moreover, they show in these cases that there is an automorphic form π_ℓ on $GL_{\ell+1}(\mathbb{A}_\mathbb{Q})$ whose L -function, $L(s, \pi_\ell)$, is equal to $L(s, \phi, \text{sym}^\ell)$. This correspondence, $(\pi, \text{sym}^\ell) \rightarrow \pi_\ell$, is the $\text{sym}^\ell : GL(2) \rightarrow GL(\ell + 1)$ functorial lift. It is a special but quite striking and useful instance of the general functoriality conjecture [La2].

It is the analytic properties of the L -functions such as their size on $\Re(s) = \frac{1}{2}$ that is of most use to us. In this connection the Grand Riemann Hypothesis (GRH) is decisive. It asserts that the zeroes of any of the functions $\Lambda(s, \pi)$ mentioned above (here we are thinking of π being an automorphic form on GL_n) are on the line $\Re(s) = \frac{1}{2}$. A particular consequence of GRH is the Grand Lindelöf Hypothesis (GLH), which asserts that for π of fixed degree n say and $\epsilon > 0$,

$$(16) \quad L\left(\frac{1}{2} + it, \pi\right) \ll_{\epsilon, n} (C(\pi, t))^\epsilon,$$

where $C(\pi, t)$ is the “analytic conductor” defined in [I-S2]. For example, for our Maass forms ϕ on $X(1)$ with eigenvalue $\lambda_\phi = \frac{1}{4} + t_\phi^2$, $C(\phi, 0) = \lambda_\phi$. For a Dirichlet L -function $L(s, \chi)$, $C(\chi, t) = (|t| + 1)(q + 1)$, while for the L -functions $L(s, \phi_\lambda \times \phi_\lambda \times \phi)$ with ϕ fixed and $\lambda \rightarrow \infty$, $C(\phi_\lambda \times \phi_\lambda \times \phi, 0) = \lambda_\phi^2$. If the Ramanujan Conjectures (32) and their generalizations are true, then for $\delta > 0$, $L(s, \pi)$ is uniformly bounded for $\Re(s) \geq 1 + \delta$. This will not continue to hold even on $\Re(s) = 1$; however, the GLH asserts that it is almost true up to $\Re(s) = \frac{1}{2}$. That is, all L -functions are bounded by an arbitrary small power of their conductor in $\Re(s) \geq \frac{1}{2}$. It is this technical-looking feature that is very useful in the study of the eigenfunctions. In general, the only bound we have is the convexity bound (see [Har])

$$(17) \quad L\left(\frac{1}{2} + it, \pi\right) \ll_\epsilon [C(\pi, t)]^{\frac{1}{4} + \epsilon}.$$

A number of the problems discussed in these notes are resolved by establishing subconvex estimates for a suitable family of π 's. That is, a $\delta > 0$ is produced so that

$$(18) \quad L\left(\frac{1}{2}, \pi\right) \ll (C(\pi, 0))^{\frac{1}{4}-\delta},$$

for π in the family. See [F], [I-S2] for recent reviews.

Appendix 2: Frobenius Automorphisms. We review the definition of the Frobenius element. Details can be found in standard books on algebraic number theory, for example [Lan2] and [C-F]. Let K be a finite Galois extension of \mathbb{Q} and G the corresponding Galois group. The ring of integers of K , denoted O_K , is a Dedekind domain. Let p be a rational prime. The principal ideal (p) factors into a product of prime ideals $(p) = (\mathfrak{b}_1 \mathfrak{b}_2 \dots \mathfrak{b}_r)^e$. The integer e is the ramification index of p and is equal to 1 for all primes p not dividing the discriminant of K . We restrict attention to such unramified primes. If $\mathfrak{b}|p$ and $\sigma \in G$, then $\sigma(\mathfrak{b})|p$ and in fact G acts transitively on the primes \mathfrak{b} dividing p . For such \mathfrak{b} the decomposition group $G_{\mathfrak{b}}$ is the corresponding stabilizer of \mathfrak{b} , that is $\{\sigma \in G | \sigma(\mathfrak{b}) = \mathfrak{b}\}$. The different decomposition groups for $\mathfrak{b}|p$ are conjugate in G . Now $G_{\mathfrak{b}}$ acts in the obvious way as automorphisms of the finite field O_K/\mathfrak{b} , all of which fix the subfield $\mathbb{Z}/p\mathbb{Z}$. Denote the degree of the field extension $(O_K/\mathfrak{b})/(\mathbb{Z}/p\mathbb{Z})$ by f . Since we are assuming that $e = 1$, $fr = \deg(K/\mathbb{Q}) = n$. Also, since $e = 1$, $G_{\mathfrak{b}}$ is isomorphic to $\text{Gal}((O_K/\mathfrak{b})/(\mathbb{Z}/p\mathbb{Z}))$. By the theory of finite fields the latter is cyclic of order f and is generated by the Frobenius automorphism $x \rightarrow x^p$. We call the corresponding element of $G_{\mathfrak{b}}$, $\text{Frob}_{\mathfrak{b}}$. It satisfies the relation $\text{Frob}_{\mathfrak{b}}(\alpha) \equiv \alpha^p \pmod{\mathfrak{b}}$, for all $\alpha \in O_K$. The different elements $\text{Frob}_{\mathfrak{b}} \in G$ for $\mathfrak{b}|p$ are conjugate in G . In this way we obtain for each unramified prime p a conjugacy class Frob_p in G . If p is ramified in K , we can still define Frobenius elements $\text{Frob}_{\mathfrak{b}}$ in $G_{\mathfrak{b}}$, but they are only determined up to the subgroup of inertia, $\ker : G_{\mathfrak{b}} \rightarrow \text{Gal}((O_K/\mathfrak{b})/(\mathbb{Z}/p\mathbb{Z}))$.

Frob_p tells us how (p) factors in O_K . For example, if $\text{Frob}_p = \{1\}$ in G , then $(p) = \mathfrak{b}_1 \mathfrak{b}_2 \dots \mathfrak{b}_n$ and $O_K/\mathfrak{b}_j \cong \mathbb{Z}/p\mathbb{Z}$; that is p splits completely. If K is the splitting field of $f(x) \in \mathbb{Z}[x]$, then in this case f splits into linear factors over $\mathbb{Z}/p\mathbb{Z}$.

Using Brauer's theorem on characters of finite groups [Br] together with class field theory, one can show (Artin) that the Artin L -function defined in (14) can be expressed as a ratio of products of the Hecke L -functions $L(s, \lambda)$ for suitable finite order characters λ on suitable field extensions. In particular, this yields the meromorphicity and exact functional equation for $L(s, \rho)$ after completing it with an appropriate archimedean factor. The possible archimedean factors in this 2-dimensional case are $(2\pi)^{-s}\Gamma(s)$, $(\pi^{-s/2}\Gamma(s/2))^2$ or $(\pi^{-(s+1)/2}\Gamma(\frac{s+1}{2}))^2$. In the first case ρ is called odd while in the second and third cases it is called even. The cases odd or even can be characterized by whether $\det \rho(c)$ is -1 or 1 where c is a complex conjugation in $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. It is the even ρ which give rise to Maass forms with eigenvalue $\frac{1}{4}$. The fully unramified archimedean factor is the second case above, and it corresponds to a cosine series in (15). The integer appearing in the functional equation of $L(s, \rho)$ (in the same way as q appears in (3) of Appendix 1 for $L(s, \chi)$) is called the conductor of ρ , and it can be computed in terms of local ramification [C-F].

Appendix 3: Trace Formula. The trace formula for $X(1)$ reads as follows [Sel2]. Let $g \in C_0^\infty(\mathbb{R})$ be an even smooth function of compact support and let $h(\xi) = \hat{g}(\xi/2\pi)$. Note that h is an entire function:

$$\begin{aligned}
 & \sum_{t_\phi} h(t_\phi) - \frac{1}{2\pi} \int_{-\infty}^{\infty} h(t) \frac{\phi'_{\Gamma(1)}}{\phi_{\Gamma(1)}} \left(\frac{1}{2} + it \right) dt \\
 &= \frac{\text{Area}(X(1))}{2\pi} \int_{-\infty}^{\infty} \tanh(\pi t) t h(t) dt - \frac{1}{\pi} \int_{-\infty}^{\infty} h(t) \frac{\Gamma'}{\Gamma} (1 + it) dt \\
 &\quad - 2 \log 2g(0) + h(0) \\
 &\quad + \sum_{\{R\}} \sum_{1 \leq \nu \leq m-1} \frac{2}{m \sin \frac{\pi \nu}{m}} \int_{-\infty}^{\infty} \frac{h(r) e^{-\frac{\pi \nu}{m} r}}{1 + e^{-2\pi r}} dr \\
 (1) \quad &\quad + 2 \sum_{\{P\}} \sum_{k=1}^{\infty} \frac{\log N(P)}{N(P)^{k/2} - N(P)^{-k/2}} g(k \log N(P)).
 \end{aligned}$$

The t_ϕ 's run through the discrete spectrum of $X(1)$ ($\lambda_\phi = \frac{1}{4} + t_\phi^2$). The continuous spectrum contribution is through the winding number integral involving $\phi_{\Gamma(1)}(s)$ which is given explicitly in (24). The sum $\{P\}$ is over primitive hyperbolic conjugacy classes of $\Gamma(1)$. $\gamma \in \Gamma$ is primitive if $\gamma \neq \gamma_1^\nu$ for any $\gamma_1 \in \Gamma$ and $|\nu| \geq 2$; γ is hyperbolic if $|\text{trace}(\gamma)| > 2$. A hyperbolic $\gamma \in SL(2, \mathbb{R})$ can be conjugated (in $SL(2, \mathbb{R})$) into the form $\pm \begin{pmatrix} (N(\gamma))^{1/2} & 0 \\ 0 & N(\gamma)^{-1/2} \end{pmatrix}$ with $N(\gamma) > 1$. γ fixes a unique geodesic ℓ in \mathbb{H} , and ℓ modulo γ has length $\log N(\gamma)$. In this way the set $\{P\}$ corresponds to the set of primitive closed geodesics on $X(1)$. The sum $\{R\}$ is over elliptic conjugacy classes of which there are two, one of order $m = 2$ (in $PSL(2, \mathbb{Z})$) and the other of order $m = 3$.

The left-hand side of (1) is the spectral side of the trace formula being a sum over the discrete and continuous spectrum. The right-hand side being over the closed geodesics is called the geometric side (or the orbital integral side since in the derivation of the formula these terms arise as orbital integrals). The geometric side which in the most general setting [A] can be very complicated is nevertheless quite explicit. For the case at hand, the lengths of the primitive closed geodesics are precisely the numbers $2 \log \epsilon_d$, where $0 < d \equiv 0$ or $1 \pmod{4}$ is square-free and ϵ_d is the fundamental solution $\frac{t_0 + \sqrt{d}u_0}{2}$ to the Pell equation $t^2 - du^2 = 4$, with multiplicity the class number $h(d)$ of integral binary quadratic forms of discriminant d . The fact that the geometric side is explicit is at the heart of many modern applications of the general trace formula, one strategy being that one computes explicitly the geometric sides for quotients $\Gamma \backslash G$ and $\Gamma' \backslash G'$ of different (adele) groups G and G' . In some special but striking cases one can match the corresponding geometric sides. This leads to correspondences between the spectral sides, which then typically establishes a form of a functorial correspondence [La4]. The cyclic base change theorem for GL_2 mentioned in the introduction is proved this way using a Galois twisted version of the trace formula.

Returning to the case of $X(1)$, we apply (1) with $h_T(t) = H\left(\frac{t}{T}\right)$ for a fixed H and let $T \rightarrow \infty$. For T large enough the contribution to the hyperbolic conjugacy

classes is zero, and hence for any such H we have

$$\begin{aligned} \sum_{\phi} H\left(\frac{t_{\phi}}{T}\right) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H\left(\frac{t}{T}\right) \frac{\phi'_{\Gamma(1)}}{\phi_{\Gamma(1)}} \left(\frac{1}{2} + it\right) dt \\ &\sim \frac{\text{Area}(X(1))}{2\pi} \int_{-\infty}^{\infty} H\left(\frac{t}{T}\right) \tanh(\pi t) t dt. \end{aligned}$$

By an approximation argument this leads to

$$(2) \quad \sum_{|t_{\phi}| \leq T} 1 - \frac{1}{2\pi} \int_{-T}^T \frac{\phi'_{\Gamma(1)}}{\phi_{\Gamma(1)}} \left(\frac{1}{2} + it\right) dt \sim \frac{\text{Area}(X(1))}{2\pi} T^2.$$

In view of the expression of $\phi_{\Gamma(1)}(s)$ in terms of $\Lambda(s)$ (24) and in particular the fact that $\Lambda(s)$ is an entire function of order 1, one concludes that for $X(1)$

$$(3) \quad \sum_{|t_{\phi}| \leq T} 1 \sim \frac{\text{Area}(X(1))}{2\pi} T^2 \quad \text{as } T \rightarrow \infty.$$

This proves (25), that is that $X(1)$ is essentially cuspidal.

Appendix 4: Microlocal Lifts. We give the construction for a general compact Riemannian manifold X . Let Δ be the Laplacian on functions on X and ϕ_{λ} a normalized eigenfunction

$$(1) \quad \left. \begin{aligned} \Delta \phi_{\lambda} + \lambda \phi_{\lambda} &= 0 \\ \int_X (\phi_{\lambda}(x))^2 dv(x) &= 1 \end{aligned} \right\}.$$

The probability measure μ_{ϕ} on X corresponding to the state ϕ is given by $d\mu_{\phi} = \phi^2(x)dv(x)$. We want to define a microlocal lift ν_{ϕ} of μ_{ϕ} to the unit cosphere bundle S^*X . To do so let $a(x, \xi) \in C^{\infty}(S^*X)$, which we can think of as a symbol homogeneous of degree 0 in ξ , on the cotangent bundle T^*X . Let A be a zeroth order pseudo differential operator with principal symbol $\sigma_A = a(x, \xi)$. Using a Frederichs symmetrization, one can adjust A by a lower order operator to get a zeroth order pseudo differential operator $Op(a)$, with principal symbol a , such that $\langle Op(a)\psi, \psi \rangle \geq 0$ whenever $a \geq 0$ and $\psi \in C^{\infty}(X)$ [Ze]. In this way the Wigner distribution on $C^{\infty}(S^*X)$ defined by

$$(2) \quad a \longrightarrow \langle Op(a)\phi_{\lambda}, \phi_{\lambda} \rangle$$

is positive and hence defines a positive measure on S^*X . We denote it by $\nu_{\phi_{\lambda}}$. If $a(x, \xi) = a(x)$, then

$$(3) \quad \nu_{\phi_{\lambda}}(a) = \int_X a(x) d\mu_{\phi_{\lambda}} + o(1) \quad \text{as } \lambda \rightarrow \infty.$$

Hence, for the purposes of studying the large λ limits of $\mu_{\phi_{\lambda}}$ we see that $\nu_{\phi_{\lambda}}$ projects asymptotically to $\mu_{\phi_{\lambda}}$. We call $\nu_{\phi_{\lambda}}$ a microlocal lift of $\mu_{\phi_{\lambda}}$ to S^*X . In the case that we are discussing in these notes, that is $X = \Gamma \backslash \mathbb{H}$ and $S^*X = \Gamma \backslash PSL(2, \mathbb{R})$, one can give a canonical definition of this microlocal lift [Ze].

Fix $t \in \mathbb{R}$; then we have

$$(4) \quad \nu_{\phi_{\lambda}}(a) := \langle Op(a)\phi_{\lambda}, \phi_{\lambda} \rangle = \langle e^{-i\sqrt{\Delta}t} Op(a)e^{i\sqrt{\Delta}t} \phi_{\lambda}, \phi_{\lambda} \rangle.$$

By the propagation of singularities theorem for the wave equation [Eg] we have that the last inner product equals

$$(5) \quad \langle Op(a_o\mathcal{G}_t) \phi_\lambda, \phi_\lambda \rangle + o(1) \quad \text{as } \lambda \longrightarrow \infty.$$

Here \mathcal{G}_t is the geodesic flow on S^*X which is the bicharacteristic flow for the wave equation. Hence

$$(6) \quad \nu_{\phi_\lambda}(a) = \nu_{\phi_\lambda}(a_o\mathcal{G}_t) + o(1) \quad \text{as } \lambda \longrightarrow \infty.$$

It follows that if ν is a quantum limit on S^*X , that is to say a weak limit of the ν_{ϕ_λ} 's, then ν is \mathcal{G}_t invariant.

Appendix 5: Quantum versus Classical Fluctuations. This appendix is an outgrowth of discussions with Z. Rudnick, who pointed me to the physics literature and in particular to references [F-P] and [E-F-K-A-M-M]. In these, it is suggested and partially confirmed with some numerical experimentations that for strongly chaotic Hamiltonians the variance of the quantum observables $\langle Op(a)\phi_\lambda, \phi_\lambda \rangle$ should be the same as the variance of a along a typical orbit of the classical flow. For our case of $X = \Gamma \backslash G / K$, $G = SL(2, \mathbb{R})$, $K = SO(2)$ and $S^*X = \Gamma \backslash G$ this classical variance is (we assume that $\bar{a} = \int_{\Gamma \backslash G} a(g) dg = 0$)

$$(1) \quad V(T, a) = \int_{\Gamma \backslash G} \left| \int_0^T a(g\mathcal{G}_t) dt \right|^2 dg.$$

As $T \longrightarrow \infty$ we have $V(T, a) \sim V(a)T$, where

$$(2) \quad V(a) = \int_{-\infty}^{\infty} \int_{\Gamma \backslash G} a(g\mathcal{G}_t) \overline{a(g)} dg dt.$$

The geodesic flow on such a hyperbolic surface is rapidly mixing, and hence the t -integral in (2) converges absolutely. We may polarize the form V to get a bilinear form on smooth functions of mean zero on $\Gamma \backslash G$:

$$(3) \quad V(a, b) = \int_{-\infty}^{\infty} \langle U(t)a, b \rangle dt$$

where $U(t)a(g) = a(g\mathcal{G}_t)$ and $\langle \cdot, \cdot \rangle$ is the inner-product on $L^2(\Gamma \backslash G)$. Clearly V satisfies

- (i) $V(a, a) \geq 0$,
- (ii) $V(U(t_1)a, U(t_2)b) = V(a, b)$,
- (iii) $V(Da, b) = V(a, Db)$.

Here D is the Casimir element in the center of the enveloping algebra of $\text{Lie}(G)$. In particular V is diagonalized by the decomposition of $L^2(\Gamma \backslash G)$ into irreducibles under the right regular representation of G . In this notation the ‘‘conjectured’’ leading term in (41), that is for the quantum variance, is:

$$(4) \quad \sum_{\lambda_j \leq \lambda} |\langle Op(a)\phi_{\lambda_j}, \phi_{\lambda_j} \rangle|^2 \sim V(a)\sqrt{\lambda} \quad \text{as } \lambda \longrightarrow \infty \quad \text{[F-P].}$$

Certainly (4) does not hold in general. For example, if X has a reflection symmetry about a geodesic (such as $X(1)$ which has the symmetry $z \longrightarrow -\bar{z}$ in Figure 2), then the eigenfunctions ϕ_λ can be chosen to be even or odd with respect to this symmetry. Hence ϕ_λ^2 is even, and so if ψ is a function on X which is odd with

respect to this symmetry, then $\langle Op(\psi)\phi_\lambda, \phi_\lambda \rangle = \int_X \psi(z)\phi_\lambda^2(z)dA(z) = 0$. On the other hand, for such a ψ , $V(\psi) \neq 0$ so that (4) cannot hold.

We compare the quantum variance derived in Theorem 5 when $X = X(1)$ with the suggestion (4). Firstly, the order of magnitude $\sqrt{\lambda}$ is indeed correct. Secondly, for $a = \psi$ a function on $X(1)$, both of the quadratic forms $V(\psi)$ and $B(\psi)$ are diagonalized by the Maass-Hecke eigenfunctions ϕ_λ themselves! It therefore suffices to compare V and B on such an L^2 -normalized ϕ_λ . In this case the function $\langle \phi_\lambda, U(g)\phi_\lambda \rangle$ on G is a spherical function (that is a K -biinvariant eigenfunction of D) and has value 1 at $g = e$. As such it is uniquely determined by the eigenvalue $\lambda_\phi = \frac{1}{4} + t_\phi^2$. One can then calculate

$$(5) \quad V(\phi_\lambda) = \int_{-\infty}^{\infty} \langle \phi_\lambda, U \left(\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \right) \phi_\lambda \rangle dt = \frac{|\Gamma(\frac{1}{4} - \frac{it_\phi}{2})|^4}{4\pi |\Gamma(\frac{1}{2} - it_\phi)|^2}.$$

With this evaluation one finds that the more detailed version of Theorem 5(c) reads:

$$(6) \quad B(\phi_\lambda) = L\left(\frac{1}{2}, \phi_\lambda\right) V(\phi_\lambda).$$

Thus by inserting the arithmetical factor $L(\frac{1}{2}, \phi)$ as a correction to the classical variance we obtain the correct quantum variance! Note that if ϕ is odd with respect to the symmetry $z \rightarrow -\bar{z}$, then $\Lambda(s, \phi)$ is odd under $s \rightarrow 1-s$, so that $L(\frac{1}{2}, \phi) = 0$ and hence $B(\phi) = 0$, as it must be. The pleasing relation (6) between the quantum and classical variances when X is arithmetic suggests that perhaps (4) is true for the generic hyperbolic surface in Teichmüller space. It would be interesting to check this numerically.

The same physics papers above go on to conjecture that after normalization by the variance the observables $\langle Op(a)\phi_\lambda, \phi_\lambda \rangle$ become Gaussian as $\lambda \rightarrow \infty$ (the numerics for this are quite convincing). For the classical motion such a central limit theorem is valid [Ra]. However, for $X(1)$ such a Gaussian behavior for the quantum observables is unlikely to be true. At least such a behavior would violate certain conjectures. Consider again the case that $Op(a)$ is a multiplication operator by a fixed even Hecke-Maass form ϕ on $X(1)$ for which $L(\frac{1}{2}, \phi) \neq 0$ (which is expected to be the case for all such ϕ). Then according to (48) the distribution question is essentially one about the distribution of $L(\frac{1}{2}, \text{sym}^2\phi_\lambda \times \phi)$ as $\lambda \rightarrow \infty$. Now the set of L -functions $L(s, \text{sym}^2\phi_\lambda \times \phi)$, $\lambda_j \leq \lambda$, is a family of L -functions in the sense of [Ka-Sa]. It has an $SO(\text{even})$ symmetry (one can check this by examining the 1-level density of low-lying zeros which follow $W_{SO(\text{even})}$ density [Ka-Sa], at least for a restricted set of test functions). Now if we invoke the conjectures [Ke-Sn] concerning the moments of the special value at $s = \frac{1}{2}$ for this family, then we conclude that distribution of $L(\frac{1}{2}, \text{sym}^2\phi_\lambda \times \phi)$ cannot be Gaussian. The point is that $L(\frac{1}{2}, \text{sym}^2\phi_\lambda \times \phi)$ is large too often, rendering the moments to be infinite. On the other hand, by analogy with the distribution of $\log |\zeta(\frac{1}{2} + it)|$ as $t \rightarrow \infty$ [Sel3] and random matrix considerations [Ke-Sn] for this family, one might expect that for $X(1)$

$$\frac{\log (|\langle \phi_\lambda^2, \phi \rangle| \lambda^{1/4})}{(\log \log \lambda)^{1/2}}$$

becomes Gaussian about a drifting mean as $\lambda \rightarrow \infty$.

Appendix 6: Cancellation in Shifted Sums. We describe the connection between the sums (21) and the spectrum of $X(N)$. For definiteness we consider the congruence surfaces $X_0(N) = \Gamma_0(N) \backslash \mathbb{H}$ where

$$\Gamma_0(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : N|c \right\}.$$

For this group the subgroup stabilizing ∞ is $\Gamma_\infty = \left\{ \pm \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \mid m \in \mathbb{Z} \right\}$. Let f be a fixed holomorphic cusp form of even integral weight k for $\Gamma_0(N)$. The definition of such forms is as in (26), except that the weight 4 is allowed to be any positive even integer k and $\Gamma(N)$ is replaced by $\Gamma_0(N)$. We may expand f in a Fourier series

$$(1) \quad f(z) = \sum_{n=1}^{\infty} a_f(n) e(nz).$$

Let

$$(2) \quad \lambda_f(n) := a_f(n) n^{(1-k)/2}.$$

With this normalization the classical Ramanujan Conjecture (which is equivalent to the bound (33) for the Hecke operators in this setting), which was proved in [De], asserts that for any $\epsilon > 0$

$$(3) \quad \lambda_f(n) = O_\epsilon(n^\epsilon).$$

Fix ν_1, ν_2, h positive integers and set

$$(4) \quad D_f(s, \nu_1, \nu_2, h) = \sum_{\nu_1 m - \nu_2 n = h} \lambda_f(m) \overline{\lambda_f(n)} \left(\frac{\sqrt{mn\nu_1\nu_2}}{m\nu_1 + n\nu_2} \right)^{k-1} (m\nu_1 + n\nu_2)^{-s}.$$

In view of (3) this converges absolutely for $\Re(s) > 1$. In order to investigate the analytic properties of $D(s)$ define the Poincaré series $U_h(z, s)$ as follows

$$(5) \quad U_h(z, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma_0(\nu_1\nu_2 N)} y(\gamma z)^s e(-hx(\gamma z)).$$

The series (5) converges absolutely for $\Re(s) > 1$, and since $y^s e(-hx)$ is invariant under Γ_∞ it follows that $U_h(\gamma z, s) = U_h(z, s)$ for $\gamma \in \Gamma_0(N\nu_1\nu_2)$. Moreover, a direct calculation shows that

$$(6) \quad \Delta(y^s e(-hx)) + s(1-s)(y^s e(-hx)) = -4\pi^2 h^2 y^{s+2} e(-hx).$$

Hence

$$(7) \quad \Delta U_h(z, s) + s(1-s)U_h(z, s) = -4\pi^2 h^2 U_h(z, s+2).$$

Inverting this relation using the resolvent gives

$$(8) \quad U_h(z, s) = (\Delta + s(1-s))^{-1} U_h(z, s+2).$$

(There is a technical issue that $U_h(\cdot, s)$ is not in $L^2(X_0(N\nu_1\nu_2))$, but this difficulty can easily be overcome in what follows.) From (8) we see that since the series $U_h(z, s+2)$ is holomorphic in $\Re(s) > 0$, $U_h(z, s)$ is meromorphic in $\Re(s) > \frac{1}{2}$ with possible poles at $s(1-s) = \lambda_\phi$ where λ_ϕ is an eigenvalue of Δ on $X_0(\nu_1\nu_2 N)$. For $\Re(s) > \frac{1}{2}$ only eigenvalues $\lambda_\phi < \frac{1}{4}$ are relevant. The eigenvalue $\lambda_\phi = 0$ does not contribute a pole since $\langle U_h(z, s), 1 \rangle = 0$ (at least formally). Thus the main conjecture about the low-energy spectrum, that is Conjecture 2, implies that $U_h(z, s)$

is analytic in $\Re(s) > \frac{1}{2}$. Theorem 2 ensures that in any case $U_h(z, s)$ is analytic in $\Re(s) > \frac{39}{64}$.

To connect this with $D_f(s)$ let $V(z) = f(\nu_1 z) \overline{f(\nu_2 z)} y^k$. V is $\Gamma_0(N\nu_1\nu_2)$ invariant and is rapidly decreasing as z approaches the cusps of $X_0(\nu_1\nu_2N)$. Consider

$$(9) \quad I = \langle U_h, V \rangle = \int_{X_0(N\nu_1\nu_2)} U_h(z, s) V(z) dA(z).$$

Unfolding this integral according to the definition of U_h and the invariance of V yields

$$(10) \quad I = \int_0^\infty \int_0^1 V(z) y^s e(-hx) \frac{dx dy}{y^2}.$$

Now do the x -integral first and use (1) and then do the y -integral to get

$$(11) \quad I = (2\pi)^{-(s+k-1)} \Gamma(s+k-1) (\nu_1\nu_2)^{\frac{1-k}{2}} D_f(s, \nu_1, \nu_2, h).$$

That is

$$(12) \quad D_f(s, \nu_1, \nu_2, h) = \frac{(2\pi)^{s+k-1} (\nu_1\nu_2)^{\frac{k-1}{2}}}{\Gamma(s+k-1)} \langle U_h(\cdot, s), V \rangle.$$

From this and the previous discussion we may deduce that $D_f(s)$ is meromorphic in $\Re(s) > \frac{1}{2}$ and if Conjecture 2 is true that $D_f(s)$ is analytic in $\Re(s) > \frac{1}{2}$. With quite a bit more effort [Sa5] one can establish polynomial bounds for the growth of $D_f(s)$ in $\Re(s) > \frac{1}{2}$ when $|\Im(s)| \rightarrow \infty$. With this one can use Mellin inversion and standard contour shifts to deduce the cancellations claimed in (21); see [Sa5].

Appendix 7: Numerical Methods. The first attempts at the numerical computation of the spectrum of $X(1)$ were carried out in [Ca]. Their method yields the first few odd eigenfunctions, but it was not successful in picking up the elusive even eigenfunctions. There followed computations by others, and in one such computation the zeros of the Riemann Zeta Function appeared among the numbers $\frac{1}{2} + it_\phi$ (where as usual $\lambda_\phi = \frac{1}{4} + t_\phi^2$)! Eventually the reason for this was discovered in [He2]. The numerical method was faulty in that it allowed the eigenfunctions to have logarithmic singularities. These fake eigenfunctions were the ones that gave rise to eigenvalues which correspond to the zeros of $\zeta(s)$.

The method developed in [He1] to compute the spectrum is known as “collocation”. The eigenfunctions ϕ (cusp forms) in question have Fourier developments as in (18):

$$(1) \quad \begin{aligned} \phi^+(z) &= \sum_{n=1}^\infty \rho_\phi(n) y^{1/2} K_{it_\phi}(2\pi ny) \cos(2\pi nx), \\ \phi^-(z) &= \sum_{n=1}^\infty \rho_\phi(n) y^{1/2} K_{it_\phi}(2\pi ny) \sin(2\pi nx). \end{aligned}$$

Here $\phi^+(z)$ is even and $\phi^-(z)$ is odd. The numbers t_ϕ and $\rho_\phi(n)$ are what we want to compute. Note that the Bessel function $K_{it_\phi}(y)$ is exponentially decreasing for $y \gg |t_\phi|$, so that the series (1) is approximated to an accuracy of $O(e^{-2\pi M})$ for $y \geq \frac{\sqrt{3}}{2}$ when truncating the series at $n \leq M$ (here we are thinking of t_ϕ being of moderate size). What is lacking to ensure that $\phi(z)$ is a cusp form is the relation $\phi(-1/z) = \phi(z)$.

The first method which is good to get eigenvalues of $X(1)$ with $\lambda_\phi \leq 250000$ is as follows. Truncate (1) at $n \leq M$ and choose points $z_1, \dots, z_M \in \mathbb{H}$ which are evenly distributed in $\mathcal{F}(1)$. The equations for the truncated series $\phi^{(M)}(z)$,

$$(2) \quad \phi^{(M)}(z_j) = \phi^{(M)}(-1/z_j),$$

yield a homogeneous linear system of M equations with M unknowns. They are of the form

$$(3) \quad \sum_{n=1}^M \rho(n) I_n(z_j, t) = 0$$

where

$$(4) \quad I_n(z, t) = \sqrt{y^*} K_{it}(2\pi n y^*) \cos(2\pi n x^*) - \sqrt{y} K_{it}(2\pi n y) \cos(2\pi n x)$$

and $z^* = -1/z$.

One way to proceed at this point is to seek solutions $t \leq T$ of

$$(5) \quad \det(I_n(z_j, t))_{\substack{n=1, \dots, M \\ j=1, \dots, M}} = 0.$$

However, it is more expedient to choose a second set of points w_2, \dots, w_M and to solve the linear equations (by elimination and setting $\rho(1) = 1$):

$$(6) \quad \left. \begin{aligned} -I_1(z_j, t) &= \sum_{n=2}^M \rho^{(z)}(n) I_n(z_j, t) \\ -I_1(w_j, t) &= \sum_{n=2}^M \rho^{(w)}(n) I_n(w_j, t) \end{aligned} \right\}$$

for $j = 2, \dots, M$.

For a genuine eigenvalue $t = t_\phi$ we will have

$$(7) \quad \rho^{(z)}(n) = \rho^{(w)}(n), n = 2, \dots, M.$$

So one determines the t 's for which (7) is valid (actually one minimizes a related functional). This works well for $T \approx 500$. For larger T the system of linear equations becomes ill conditioned.

In order to deal with larger values of t , a somewhat different approach (inspired by [Sta]) is taken. The series (1) is truncated at say $M \approx 5T$ (with $t \approx T$) and yields a good approximation of ϕ for $z \in \mathcal{F}(1)$. On other hand, for y small

$$(8) \quad \rho(n) y^{1/2} K_{it}(2\pi n y) = \int_0^1 \phi^+(x + iy) \cos(2\pi n x) dx.$$

Replacing the integral by a sum (with Q large), we have

$$(9) \quad \rho(n) y^{1/2} K_{it}(2\pi n y) \cong \frac{1}{Q} \sum_{j=1}^Q \phi^+ \left(\frac{j}{Q} + iy \right) \cos \left(\frac{2\pi n j}{Q} \right).$$

By transforming the points $\frac{j}{Q} + iy$ into $\mathcal{F}(1)$ we can evaluate the right-hand side of (9) with our M term truncation. This yields (for a given y) a system of linear equations for $\rho(n)$, $n \leq M$ (and once these are determined many more coefficients can be computed from (8)). Proceeding as before with this linear system, set $\rho(1) = 1$ and solve for the coefficients with y and y' . Then choose the t 's so as to

minimize $|\rho^{(y)}(n) - \rho^{(y')}(n)|$. These linear equations turn out to be well conditioned for T as large as 11000 [He3]. To carry this out one also needs a fast routine to compute $K_{it}(y)$ with t large.

An excellent test as to whether the t_ϕ 's and $\rho_\phi(n)$'s are accurate is that if the spectrum is simple, then the coefficients must inherit the relations satisfied by the Hecke operators (30). That is $\rho_\phi(n)\rho_\phi(m) = \sum_{d|(n,m)} \rho_\phi\left(\frac{nm}{d^2}\right)$ must hold. Since

these relations were not imposed, their truth is convincing evidence that genuine cusp forms are being picked up. By examining average deviations in the Weyl count $|\{j|\lambda_j \leq \lambda\}|$ (see [St] for example) one can be convinced that no t_ϕ 's were omitted in the above calculations. The approach described above works equally well for computing the spectra of nonarithmetic triangle groups [He2].

The method used in [G-S] to obtain the plots in Figure 3 is quite different. They use the trace formula in Appendix 3 together with the explicit knowledge of the length spectrum for $X(1)$. While this method gives only the first few eigenvalues, it can be applied quite easily to $X(N)$ with (N moderate) as well.

Acknowledgements. I thank E. Lindenstrauss, Y. Petridis, J-P. Serre, F. Spinu and A. Strombergsson for their comments on a preliminary version of this paper. Also, thanks to G. Pecht for her excellent job in typing the manuscript.

REFERENCES

- [A] J. Arthur, The trace formula and Hecke operators, *Number Theory, Trace Formulas and Discrete Groups*, Academic Press (1989), 11-27. MR **90e**:11072
- [Ar] E. Artin, Abh. Math. Sem., Hamburg (1924), 170-175.
- [Ber] M. Berry, Regular and irregular semiclassical wavefunctions, *J. Phys.* **A10** (1977), 2083-2091. MR **58**:8961
- [Be] L. Bers, *Spaces of Riemann surfaces*, ICM Proceedings (Edin., 1958), Cambridge Univ. Press., New York (1960), 349-361. MR **23**:A1796
- [B-S] E. Bogomolny and C. Schmit, "Nonlinearity" **6** (1993), 523-547. MR **94i**:81027
- [Bo] E. Bombieri, Lectures on the zeta function (Italy, 2002).
- [Boo] A. Booker, "Poles of Artin L -functions and the strong Artin Conjecture," preprint (2002).
- [Bor] A. Borel, Automorphic forms on $SL_2(\mathbb{R})$, *Camb. Tracts in Math.* **130** (1977). MR **98j**:11028
- [B-L] J. Bourgain and E. Lindenstrauss, "Entropy of quantum limits," *Comm. Math. Phys.* **233** (2003), no. 1, 153-171.
- [Br] R. Brauer, On Artin's L -series with general group characters, *Annals of Math.*, Vol. 48, **2** (1947), 502-551. MR **8**:503g
- [B-D-S-T] K. Buzzard, M. Dickinson, N. Shepherd-Barron and R. Taylor, On icosahedral Artin representations, *D.M.J.*, Vol. 109 (2001), 283-317. MR **2002k**:11078
- [Ca] P. Cartier, *Computers in Number Theory*, ed. Atkin-Birch, Academic Press (1971), 37-48, MR **47**:3285
- [C-F] J. Cassels and A. Frohlich, eds., "Algebraic Number Theory," Academic Press, London; Thompson Book Co., Inc., Washington, D.C. (1967). MR **35**:6500
- [Cl] L. Clozel, "Spectral theory of automorphic forms", Lectures, Park City (2002).
- [Co-PS-S] J. Cogdell, I. Piatetsky-Shapiro and P. Sarnak, "Estimates for Hilbert modular L -functions and applications" (in preparation).
- [Cog] J. Cogdell, "On sums of 3 squares" (to appear), *J. Théor. Nombres Bordeaux* (2003).
- [Co] Y. Colin de Verdiere, Ergodicité et fonctions propres du laplacien, *Comm. Math. Phys.* **102** (1985), 497-502. MR **87d**:58145
- [Co2] Y. Colin de Verdiere, Pseudo-laplaciens. II. *Ann Inst. Fourier* **32** (1983), 275-286; and **33** (1983), 87-113. MR **84k**:58222

- [Da] H. Davenport, *Multiplicative number theory*, G.T.M., Vol. 74, Springer (1980). MR **82m**:10001
- [De] P. Deligne, *Formes modulaires et représentations l -adiques*, Lecture Notes in Math., Vol. 179, Springer (1971), 139-172.
- [D-I] J. Deshouillers and H. Iwaniec, The nonvanishing of Rankin-Selberg L -functions at special points, *Contemp. Math.* **53** (1986), 51-97. MR **88d**:11047
- [D-F-I] W. Duke, J. Friedlander and H. Iwaniec, Equidistribution of roots of a quadratic congruence to prime moduli, *Annals of Math.* **141** (1995), 423-441. MR **95k**:11124
- [Du-Iw] W. Duke and H. Iwaniec, Estimates for coefficients of L -functions. IV. *Amer. Jnl. Math.* **116** (1994), 207-217. MR **95k**:11073
- [D-SP] W. Duke and R. Schulze-Pillot, Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids, *Invent. Math.* **99** (1990), 49-57. MR **90m**:11051
- [Du-Gu] H. Duistermaat and V. Guillemin, The spectrum of positive elliptic operators and periodic bicharacteristics, *Invent. Math.* **29** (1975), 39-79. MR **53**:9307
- [E-F-K-A-M-M] B. Eckhardt, S. Fishman, J. Keating, O. Agam, J. Main, K. Muller, Approach to ergodicity in quantum wave functions, *Phys. Rev. E*, Vol. 52, No. 6 (1995), 5893-5903.
- [Eg] Y. Egorov, The canonical transformations of pseudodifferential operators, *Uspehi. Mat. Nauk* **24** (1969) **5** (149), 235-236. MR **42**:657
- [E-K] M. Einsiedler and A. Katok, "Invariant measures on G/Γ for split simple Lie groups G ", *Comm. Pure Appl. Math.*, Vol. LVI, no. 8 (2003), 1184-1221.
- [F-P] M. Feingold and A. Peres, Distribution of matrix elements of chaotic systems, *Phys. Rev. A*, Vol. 34, No. 1 (1986), 591-595. MR **87i**:81055
- [F] J. Friedlander, Bounds for L -functions, *Proc. ICM* (Zurich 1994), Birkhäuser (1995), 363-373. MR **97e**:11108
- [Ga] P. Garrett, Decomposition of Eisenstein series: Rankin triple products, *Annals of Math.* (2) **125** (1987), 209-235. MR **88m**:11033
- [Ge] S. Gelbart, Automorphic forms on adèle groups, *Ann. Math. Studies* **83** (1975). MR **52**:280
- [Ge-Ja] S. Gelbart and H. Jacquet, A relation between automorphic representations of $GL(2)$ and $GL(3)$, *Ann. Sci. École Norm. Sup.* (4) **11** (1978), 471-552. MR **81e**:10025
- [Go-Ja] R. Godement and H. Jacquet, Zeta functions of simple algebras, *L.N.M.* **260**, Springer-Verlag, Berlin-New York (1972). MR **49**:7241
- [G-S] V. Golovshanski and N. Smotrov, preprint, *Inst. App. Math. Khabarovsk* (1982).
- [Har] G. Harcos, Uniform approximate functional equation for principal L -functions, *IMRN* (2002) **18**, 923-932. MR **2003d**:11074
- [H-K] M. Harris and S. Kudla, The central critical value of a triple product L -function, *Ann. of Math.* **133** (1991), 605-672. MR **93a**:11043
- [Ha] H. Hasse, *J. reine angew. Math.* **153** (1924) 113-130.
- [Hecl] E. Hecke, *Math. Zeit.*, Vol. 1 (1918), 357-376.
- [Hec2] E. Hecke, *Math. Ann.*, Vol. 114 (1937), 1-28.
- [He1] D. Hejhal, Eigenvalues of the Laplacian for Hecke triangle groups, *Memoirs of the AMS* **97** (1992), no. 469. MR **93f**:11043
- [He2] D. Hejhal, The Selberg trace formula for $PSL(2, R)$. Vol. I. *S. L. Notes*, Vol. 548 (1976); and Vol. 1001, Springer-Verlag, Berlin-New York (1980). MR **55**:12641
- [He3] D. Hejhal, On eigenfunctions of the Laplacian for Hecke triangle groups, "Emerging Applications of Number Theory," *IMA Series* **109**, Springer, New York (1999), 291-315. MR **2000f**:11063
- [H-R] D. Hejhal and Rackner, On the topography of Maass waveforms for $PSL(2, Z)$, *Exp. Math.*, **1** (1992), 275-305. MR **95f**:11037
- [Hel] R. Heller, Bound-state eigenfunctions of classically chaotic Hamiltonian systems: scars of periodic orbits, *Phys. Lett. Review* **53** (1984), 1515-1518. MR **85k**:81055
- [Hu] M. Huxley, Introduction to Kloostermania, *Banach Center Publ.* **17**, PWN, Warsaw (1982), 217-306. MR **87j**:11046

- [Iw1] H. Iwaniec, The spectral growth of automorphic L -functions, *J. reine angew. Math.* **428** (1992), 139-159. MR **93g**:11049
- [Iw2] H. Iwaniec, "Introduction to the spectral theory of automorphic forms," *Revista Mat. Iber.* (1995). MR **96f**:11078
- [I-S1] H. Iwaniec and P. Sarnak, L^∞ norms of eigenfunctions of arithmetic surfaces, *Annals of Math. (2)* **141** (1995), 301-320. MR **96d**:11060
- [I-S2] H. Iwaniec and P. Sarnak, Perspectives on the analytic theory of L -functions, GAFA 2000 (Tel Aviv, 1999), *Geom. Funct. Anal.* 2000, Special Volume, Part II, 705-741. MR **2002b**:11117
- [Ja-Sh] H. Jacquet and J. Shalika, On Euler products and the classification of automorphic representations. I. *Amer. Jnl. Math.* **103** (1981), 499-558. MR **82m**:10050a
- [Ja] D. Jakobson, Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathrm{PSL}_2(\mathbb{R})$, *Ann de l'Inst. Fourier* **44** (1994), 1477-1504. MR **96b**:11068
- [Ju] C. Judge, On the existence of Maass cusp forms on hyperbolic surfaces with cone points, *JAMS* **8** (1995), 715-759. MR **96b**:11069
- [Ka-Sp] A. Katok and R. Spatzier, Invariant measures for higher-rank hyperbolic abelian actions, *Erg. and Dyn. Systems* **16** (1996), 751-778. MR **97d**:58116
- [Ka-Sa] N. Katz and P. Sarnak, Zeroes of zeta functions and symmetry, *BAMS* **36** (1999), 1-26. MR **2000f**:11114
- [Ke-Sn] J. Keating and N. Snaith, Random matrix theory and L -functions at $s = 1/2$, *CMP* **214**, No. 1 (2000), 91-111. MR **2002c**:11108
- [K] H. Kim, "Functoriality of the exterior square of GL_4 and the symmetric fourth power of GL_2 ", *JAMS*, Vol. 16, 1 (2003), 139-183.
- [Ki-Sa] H. Kim and P. Sarnak, Appendix to [K], *JAMS* **16** (2003), no. 1, 139-183 (electronic).
- [K-S1] H. Kim and F. Shahidi, Functorial products for $GL_2 \times GL_3$ and the symmetric cube for GL_2 , *Annals of Math.* **155** (2002), 837-893.
- [K-S2] H. Kim and F. Shahidi, Cuspidality of symmetric powers with applications, *Duke Math. J.* **112** (2002), 177-197. MR **2003a**:11057
- [Kl] F. Klein, "Lectures on the Icosahedron", Paul French Frübner Co. (1913).
- [Kne] M. Kneser, "Quadratische Formen", Lecture Notes, Göttingen (1974).
- [Ku] N. Kuznecov, The Petersson conjecture for cusp forms of weight zero and the Linnik conjecture. Sums of Kloosterman sums. (Russian), *Mat. Sb. (NS)* **111(153)** (1980), no. 3, 334-383, 479. MR **81m**:10053
- [Lan1] S. Lang, $SL_2(\mathbb{R})$, Addison-Wesley (1975). MR **55**:3170
- [Lan2] S. Lang, "Algebraic Number Theory", Addison Wesley (1970). MR **44**:181
- [La1] R. Langlands, *Base change for $GL(2)$* , Annals of Math. Studies **96**, Princeton University Press (1980). MR **82a**:10032
- [La2] R. Langlands, Problems in the theory of automorphic forms. Problems in the theory of automorphic forms, *Lectures in modern analysis and applications, III*, S. L. Notes, Vol. 170 (1970), 18-61. MR **46**:1758
- [La3] R. Langlands, On the functional equations satisfied by Eisenstein series, S.L. Notes, Vol. 544 (1976). MR **58**:28319
- [La4] R. Langlands, Eisenstein series, the trace formula, and the modern theory of automorphic forms, *Number Theory, Trace Formulas and Discrete Groups*, Academic Press (1988), 125-155. MR **90e**:11077
- [L-R] E. Lapid and S. Rallis, *Annals of Math.* **157** (2003), 1-26.
- [Lax] P. Lax, The Radon transform and translation representation, *J. Evol. Equ.* **1** (2001), 311-323. MR **2003a**:58047
- [L-P] P. Lax and R. Phillips, Scattering theory for automorphic functions, *Annals of Math. Studies*, Vol. 87 (1976). MR **58**:27768
- [Laz] V. F. Lazutkin, *KAM Theory and Semi Classical Approximation to Eigenfunctions*, *Ergeb. Math.*, Vol. 24, Springer (1993). MR **94m**:58069
- [Li1] E. Lindenstrauss, "Invariant measures and arithmetic quantum unique ergodicity", preprint (2003).

- [Li2] E. Lindenstrauss, On quantum unique ergodicity for $\Gamma \backslash \mathbb{H} \times \mathbb{H}$, *IMRN* **17** (2001), 913-933. MR **2002k**:11076
- [L-P-S] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, *Combinatorica* **8**, no. 3 (1988), 261-277. MR **89m**:05099
- [Lu] W. Luo, Nonvanishing of L -values and the Weyl law, *Annals of Math.* **154** (2001), 477-502. MR **2002i**:11084
- [L-R-S] W. Luo, Z. Rudnick and P. Sarnak, On Selberg's eigenvalue conjecture, *GAFSA* **5** (1995), 387-401. MR **96h**:11045
- [L-S1] W. Luo and P. Sarnak, Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}^2$, *Proc. Publ. IHES* **81** (1995), 207-237. MR **97f**:11037
- [L-S2] W. Luo and P. Sarnak, "Quantum ergodicity of eigenfunctions on $SL_2(\mathbb{Z}) \backslash \mathbb{H}^2$ " (in preparation).
- [Mas] H. Maass, *Math. Ann.* **121** (1949), 141-183. MR **11**:163c
- [M] G. Margulis, Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators, *J. Prob. Inf. Transmission* **24** (1988), 39-46. MR **89f**:68054
- [Mc] H. McKean, An upper bound to the spectrum of Δ on a manifold of negative curvature, *J.D.G.* **4** (1970), 359-366. MR **42**:1009
- [Mc-Si] H. McKean and I. Singer, Curvature and the eigenvalues of the Laplacian, *Jnl. of Diff. Geom.* **1** (1967), 43-69. MR **36**:828
- [Me] M. Mehta, *Random Matrices*, Academic Press (1980). MR **92f**:82002
- [Mi] S.D. Miller, On the existence and temperedness of cusp forms for $SL_3(\mathbb{Z})$, *J. reine angew. Math.* **533** (2001), 127-169. MR **2002b**:11070
- [M-S] S.D. Miller and W. Schmid, "Automorphic Distributions, L -functions and Voronoi Summation for $GL(3)$ " (2002), preprint.
- [Mu] W. Müller (2002) (in preparation).
- [Pe] Y. Petridis, Spectral data for finite volume hyperbolic surfaces at the bottom of the continuous spectrum, *J. Funct. Anal.* **124** (1994), no. 1, 61-94. MR **95g**:11045
- [P-S1] R. Phillips and P. Sarnak, On cusp forms for co-finite subgroups of $\mathrm{PSL}(2, R)$, *Invent. Math.* **80** (1985), 339-364. MR **86m**:11037
- [P-S2] R. Phillips and P. Sarnak, The Weyl theorem and the deformation of discrete groups, *Comm. Pure and Appl. Math* **38** (1985), 853-866. MR **87f**:11035
- [P-S3] R. Phillips and P. Sarnak, Perturbation theory for the Laplacian on automorphic functions, *JAMS* **5** (2), No. 1 (1992), 1-32. MR **92g**:11056
- [PS-R] I. Piatetsky-Shapiro and S. Rallis, Rankin triple L functions, *Compositio Math.* **64** (1987), 31-115. MR **89k**:11037
- [Ra] M. Ratner, The central limit theorem for geodesic flows on n -dimensional manifolds of negative curvature, *Israel Jnl. of Math.*, Vol. 16 (1973), 181-197. MR **48**:11446
- [R-S] Z. Rudnick and P. Sarnak, The behaviour of eigenstates of arithmetic hyperbolic manifolds, *CMP* **161** (1994), 195-213. MR **95m**:11052
- [Ru] D. Rudolph, $\times 2$ and $\times 3$ invariant measures and entropy, *Erg. Theory and Dyn. Systems* **10** (1990), 395-406. MR **91g**:28026
- [Sa1] P. Sarnak, "A panorama of number theory", ed. Wustholz, Cambridge (2002), 121-127.
- [Sa2] P. Sarnak, On cusp forms, *The Selberg trace formula and related topics* (Brunswick, Maine, 1984), AMS Contemp. Math. **53** (1986). MR **87j**:11047
- [Sa3] P. Sarnak, Letter to Z. Rudnick, August, 2002.
- [Sa4] P. Sarnak, "Arithmetic Quantum Chaos," *Israel Math. Conf. Proc.*, Vol. 8, Ramat Gan (1995), 183-236. MR **96d**:11059
- [Sa5] P. Sarnak, Estimates for Rankin-Selberg L -functions and quantum unique ergodicity, *J. Funct. Anal.* **184** (2001), 419-453. MR **2003c**:11050
- [Sa-Wa] P. Sarnak and T. Watson, " L^4 norms of eigenfunctions on the modular surface" (in preparation).
- [SP] R. Schulze-Pillot, Exceptional integers for genera of integral ternary positive definite quadratic forms, *D.M.J.* **102** (2000), 351-357. MR **2001a**:11068
- [Sel3] A. Selberg, *Collected Works*, Springer (1988), 341-355.

- [Sel1] A. Selberg, Collected Works, Vol. I, Springer (1989), 626-674. MR **92h**:01083
- [Sel2] A. Selberg, On the estimation of Fourier coefficients of modular forms, *Proc. Symp. Pure Math.*, VIII, Amer. Math. Soc. (1965), 1-15. MR **32**:93
- [Se] J-P. Serre, Modular forms of weight one and Galois representations, *Algebraic Number Fields*, ed. Fröhlich, Academic Press (1977), 193-268. MR **56**:8497
- [Sh] F. Shahidi, Automorphic L -functions: a survey, *Perspectives in Math.*, Vol. 10, Academic Press (1990), 415-437. MR **91i**:11170
- [Shi1] G. Shimura *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press (1971). MR **47**:3318
- [Shi2] G. Shimura, On the holomorphy of certain Dirichlet series, *Proc. London Math. Soc. (3)* **31** (1975), 79-98. MR **52**:3064
- [Shn] A. Shnirelman, *Uspekhi Mat. Nauk* **29/6** (1974), 181-182.
- [Si] B. Simon, Resonances in n -body quantum systems with dilatation analytic potentials and the foundations of time-dependent perturbation theory, *Ann. of Math. 2* **97** (1973), 247-274. MR **50**:6378
- [So] C. Sogge, Concerning the L^p norm of spectral clusters for second-order elliptic operators on compact manifolds, *J. Func. Anal.* **77** (1988), 123-138. MR **89d**:35131
- [Sp] F. Spinu, Thesis, Princeton University (2003).
- [Sta] H. Stark, *Modular Forms*, R. Rankin, ed., Ellis Howard (1984), 263-268. MR **86f**:11003
- [St] G. Steil, Diplom. Math. Univ. Hamburg (1992).
- [Ste] E. Stein, Oscillatory integrals in Fourier analysis, *Lectures in Harmonic Analysis*, Princeton University Press (1986), 307-356. MR **88g**:42022
- [Str] A. Strombergsson, private communication (2003).
- [Ta] K. Takeuchi, A characterization of arithmetic Fuchsian groups, *J. Math. Soc. Japan* **27**, 4 (1975), 600-612. MR **53**:2842
- [To] A. Toth, Roots of quadratic congruences, *IMRN* (2000), No. 14, 719-739. MR **2001m**:11159
- [Tu] J. Tunnell, Artin's conjecture for representations of octahedral type, *Bull. Amer. Math. Soc.* **5** (1981), 173-175. MR **82j**:12015
- [Ve] A. Venkov, Spectral theory of automorphic functions, A translation of Trudy Mat. Inst. Steklov. **153** (1981), *Proc. Steklov Inst. Math.* **1982**, no. 4 (153) (1983). MR **85j**:11060b
- [Wa] T. Watson, Thesis, Princeton University (2002).
- [Wei] A. Weil, *Automorphic forms and Dirichlet series*, Springer Lect. Notes, Vol. 189 (1971).
- [W] M. Wolf, Infinite energy harmonic maps and degeneration of hyperbolic surfaces in moduli space, *J.D.G.* **33** (1991), 487-539. MR **92b**:58055
- [Wol1] S. Wolpert, *Invent. Math.*, Vol. 108 (1992), 67-89, 91-129. MR **93b**:58160
- [Wol2] S. Wolpert, Disappearance of cusp forms in special families, *Annals of Math. (2)* **139** (1994), 239-291. MR **95e**:11062
- [Ze] S. Zelditch, Uniform distribution of eigenfunctions on compact hyperbolic surfaces, *Duke Math. Jnl.* **55** (1987), 919-941. MR **89d**:58129

COURANT INSTITUTE OF MATH. SCIENCES; AND DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NJ 05840

E-mail address: sarnak@Math.Princeton.EDU