JOURNAL OF FUNCTIONAL ANALYSIS **121**, 78–116 (1994)

The Circle Problem in the Hyperbolic Plane*

RALPH PHILLIPS

Department of Mathematics, Stanford University, Stanford, California 94305

AND

ZEÉV RUDNICK

Department of Mathematics, Princeton University, Princeton, New Jersey 08544

Communicated by the Editors

Received April 2, 1993; revised April 26, 1993

This paper deal with the analogue of the classical circle problem in the hyperbolic plane; that is, we count the number $N_\Gamma(s, z)$ of translates of a base point z by a Fuchsian group Γ, which lie in a geodesic ball of radius s about z. If $\Sigma(s, z)$ is the theoretical best approximation to $N(s, z)$ (which has $\pi e^s/\mathrm{vol}(\Gamma)$ as its leading term), we set $d(s, z) = N(s, z) - \Sigma(s, z)$, the best known upper bound for which is $O(e^{2s/3})$. We get omega results for $d(s, z)$, which in the co-compact case are $d(s, z) = \Omega(e^{s/2}\beta(s))$, where $\beta(s) \to \infty$ as $s \to \infty$. We also show that the normalized remainder term $e(s, z) = d(s, z)/e^{s/2}$ has finite mean, which is zero unless Γ is non-compact and has null forms. Further we carry out a numerical investigation of the Fermat groups and the results are consistent with an upper bound $e(s, z) = O(e^{\varepsilon s})$ for all $\varepsilon > 0$. The problem in hyperbolic n-space is also investigated. © 1994 Academic Press, Inc.

1. INTRODUCTION

The classical circle problem, as formulated by Gauss, is concerned with the discrepency between the number of lattice points contained in a circle and the area of the circle. Precisely, for

$$N(R) = |\{(m, n) \in \mathbf{Z}^2 : m^2 + n^2 \leqslant R^2\}| \tag{1.1}$$

and

$$d(R) = N(R) - \pi R^2. \tag{1.2}$$

* Supported by NSF Grants DMS-9303225 and DMS-9102082.

0022-1236/94 $6.00

Gauss posed the problem of determining the asymptotic behaviour of $d(R)$. By a simple packing argument, Gauss bounded $d(R)$ by the area of a boundary strip $\{R-\frac{1}{2}<|x|<R+\frac{1}{2}\}$ and obtained the estimate

$$d(R)=O(R). \tag{1.3}$$

Today this is considered a rather crude estimate. The first improvement was due to Sierpinski (1906), who used Fourier analysis (Poisson summation) to show that

$$d(R) \ll R^{2/3}. \tag{1.4}$$

The next major step was taken by Walfisz (1927), who proved that

$$d(R) \ll R^{2/3-\delta}, \tag{1.5}$$

for $\delta=5/741$. After succesive improvements by various authors, an important breakthrough was achieved by Iwaniec and Mozzochi [9], who obtained

$$d(R) \ll R^{7/11+\varepsilon} \tag{1.6}$$

for all $\varepsilon>0$. This has been refined by Huxley [7].

As for the truth, there are good reasons, as well as numerical results, for believing that for all $\varepsilon>0$,

$$d(R) \ll R^{1/2+\varepsilon}. \tag{1.7}$$

One cannot expect better, since Hardy and Landau showed that

$$d(R)=\Omega(R^{1/2}\log^{1/4} R). \tag{1.8}$$

We refer to Hafner's work [5] for the best results in this direction to date and to [8] for the historical references.

In this paper, we take up an analogous problem set in the hyperbolic plane $\mathbf{H}^2$. Instead of the lattice $\mathbf{Z}^2$, we treat the orbit of a point $z \in \mathbf{H}^2$ under the action of a Fuchsian group $\Gamma \subset PSL(2, \mathbf{R})$. The hyperbolic plane has a metric of constant negative curvature -1,

$$ds^2=\frac{dx^2+dy^2}{y^2}, \tag{1.9}$$

and we measure the radius s of the circle about z in this metric. We now set

$$N_\Gamma(s, z)=|\{\gamma \in \Gamma : \operatorname{dist}(z, \gamma z) \leqslant s\}|. \tag{1.10}$$

Remark. If z is an elliptic fixed-point for Γ, then our definition counts the number of actual lattice points Γz with multiplicity equal to the order of the stabilizer of z in Γ.

The problem of estimating $N(s, z)$ was considered by Delsarte [3], who proved that for Γ co-compact

$$N(s, z) \approx \frac{\pi}{\operatorname{vol}(\Gamma)} e^{s}, \qquad \text{as} \quad s \to \infty. \tag{1.11}$$

The case of more general Fuchsian groups was considered by Huber [6], Selberg [17–19], Margulis [11], Patterson [12], Günther [4], and Lax and Phillips [10]. As in the Euclidean case, one is interested in the remainder term, which in this setting requires more care in its definition since the main term may contain secondary terms in addition to that given by (1.11). Before establishing this it is well to point out that the packing argument used by Gauss to estimate the remainder does not work in the hyperbolic setting, because the length of the boundary of the ball is comparable to its area!

In order to define the main term, we have to make use of the Laplace–Beltrami operator

$$\Delta = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right), \tag{1.12}$$

which acts on Γ-automorphic functions in $L^2(\Gamma \backslash \mathbf{H}^2)$. If $F = \Gamma \backslash \mathbf{H}^2$ is compact the spectrum of Δ is discrete; when F is non-compact but of finite volume, there is also a continuous spectrum which fills out $[\frac{1}{4}, \infty)$ with multiplicity equal to the number of cusps of Γ. If κ is the number of cusps, then for each cusp one constructs an Eisenstein series $E_j(z, s)$, $1 \leqslant j \leqslant \kappa$, and these generate the continuous spectrum. In either case, let $\lambda_0 = 0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots$ be the eigenvalues, with a corresponding orthonormal basis of eigenfunctions $\{\phi_j\}$, $\phi_0 = 1/\sqrt{\operatorname{vol}(F)}$. Set $\lambda_j = 1/4 + r_j^2$. The eigenvalues $0 \leqslant \lambda_j < 1/4$ (r_j imaginary) are called *exceptional.*

We set the main term to be[1]

$$\Sigma_\Gamma(s, z) = \frac{\pi}{\operatorname{vol}(\Gamma)} e^{s} + \sqrt{\pi} \sum_{0 < \lambda_j < 1/4} \frac{\Gamma(|r_j|)}{\Gamma(|r_j| + 3/2)} e^{(1/2 + |r_j|)s} |\phi_j(z)|^2$$
$$+ 4(s + 2(\log 2 - 1)) e^{s/2} \sum_{\lambda_j = 1/4} |\phi_j(z)|^2. \tag{1.13}$$

[1] See the Remark following Theorem 1.1.

Next we define the remainder term for the hyperbolic circle problem to be

$$d_\Gamma(s, z) = N_\Gamma(s, z) - \Sigma_\Gamma(s, z). \tag{1.14}$$

When all non-constant eigenfunctions have eigenvalue $\lambda > 1/4$, $\Sigma(s, z)$ reduces to $\pi e^s/\mathrm{vol}(\Gamma)$ as in (1.11). Finally, we define the *normalized* remainder term to be

$$e_\Gamma(s, z) = \frac{N_\Gamma(s, z) - \Sigma_\Gamma(s, z)}{e^{s/2}}. \tag{1.15}$$

It was proved by Selberg [17–19] (see also Lax and Phillips [10]) that

$$d(s, z) \ll e^{2s/3}. \tag{1.16}$$

The question is, What is the true asymptotic behaviour of $d(s, z)$? Our interest in this problem began when Peter Sarnak asked us to check this out numerically. We have done so for the Fermat groups $\Phi(N)$. Before getting into the nature of the Fermat groups, let us state that our numerical investigations indicate that $d(s, z) \asymp e^{s/2}$. A summary of the numerical data along with the program are given in the appendix. In view of the Euclidean situation (1.7), this was not unexpected. What did surprise us as the fact that the data became strongly biased in the positive direction was the level of the group increased from 1 (i.e., $\Gamma(2)$) to 8. In order to quantify this, we computed the *mean normalized remainder*

$$\mathrm{MNR}(T) = \frac{1}{T}\int_0^T e(s, z)\, ds. \tag{1.17}$$

TABLE I. Mean Normalized Remainder

	Fermat level N							
T	1	2	3	4	5	6	7	8
10	−0.0139	0.0249	−0.0272	0.0273	0.03	0.0789	0.1424	0.1889
11	0.0067	0.0254	0.0065	−0.0045	0.0193	0.0728	0.1353	0.1844
12	−0.0004	0.0291	0.0057	−0.0278	0.024	0.0404	0.1262	0.1955
13	0.0097	0.0135	−0.018	−0.0126	0.0318	0.0386	0.1084	0.1917
14	−0.0037	0.0104	−0.0101	0.0041	0.0162	0.0491	0.0843	0.1708
15	−0.0006	0.0174	−0.0069	−0.0078	−0.0044	0.0536	0.0916	0.1560
16	−0.008	0.006	−0.0308	−0.0144	0.0189	0.0310	0.1165	0.1722
17	−0.0226	0.0201	−0.0074	−0.007	0.0172	0.0087	0.0999	0.1851
18	−0.0276	0.0	−0.0112	−0.0007	0.0159	0.0245	0.0731	0.1850
19	−0.0233	0.0031	−0.0128	−0.004	−0.0043	0.0398	0.0603	0.1896
20	−0.0173	0.0139	−0.0139	−0.0148	−0.0028	0.0353	0.0695	0.1687

The result for various level with $z=i$ and radii incremented by steps of 0.1 are listed in Table I.

In order to understand what is happening, it is necessary to go a bit further into the nature of the Fermat groups. These are all subgroups of the principal congruence subgroup $\Gamma(2)$:

$$\Gamma(2)=\left\{\begin{pmatrix} a & b \\ b & d \end{pmatrix} \in SL(2, \mathbf{Z}) : \begin{pmatrix} a & b \\ b & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod 2\right\}. \tag{1.18}$$

$\Gamma(2)$ is freely generated by the elements

$$A=\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B=\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}. \tag{1.19}$$

The Fermat group $\Phi(N)$ of level N is given by

$$\Phi(N)=\left\{\gamma=A^{p_1}B^{q_1}A^{p_2}B^{q_2}\cdots A^{p_k}B^{q_k} : \sum p_i \equiv 0 \equiv \sum q_j \bmod N\right\}. \tag{1.20}$$

These groups were studied in [13, 14], where it was found that for $N<8$, all of the Eisenstein series $E_j(z, s)$ vanished at $s=\frac{1}{2}$, while for $N=8$, several of the Eisenstein series were non-zero at $s=\frac{1}{2}$. What is happening is that a pole of the scattering operator approaches $s=\frac{1}{2}$ from the left as N increases to 8, where it becomes a "null vector" (a not quite square-integrable eigenfunction), and then at $N=9$ it becomes an exceptional eigenfunction. Consequently as N increases to 8, $E_j(z, s)$ tends to a non-zero value for some j in a neighborhood of $s=\frac{1}{2}$. This together with the fact that the kernel for MNR(T) in the spectral representation approaches a delta function at $s=\frac{1}{2}$ as $T\to\infty$ explains Table I. More precisely,

THEOREM 1.1. (1) *If Γ is co-compact, then* $\lim_T MNR(T)=0$;

(2) *If Γ is non-compact, then*

$$\lim_T \mathrm{MNR}(T, z)=\sum_j |E_j(z, \tfrac{1}{2})|^2, \tag{1.21}$$

where $E_j(z, s)$ is the Eisenstein series associated with the jth cusp.

Remark. Since we include terms of size $e^{s/2}$ in the main term (1.13) (coming from the eigenvalue $\lambda=\frac{1}{4}$), we could add the term $e^{s/2}\sum_j |E_j(z, \frac{1}{2})|^2$ to (1.13) as well, and then the remainder term will always have mean zero in Theorem 1.1. We have, however, decided to refrain from doing so.

The second part of our paper deals with Ω-results, that is, lower bounds for lim sup $e(s, z)$.

THEOREM 1.2. (1) *If Γ is co-finite, then for all $\delta > 0$,*

$$e(s, z) = \Omega(e^{-\delta s}). \tag{1.22}$$

(2) *If Γ is co-finite but not co-compact, and either has some eigenvalues $\lambda > \frac{1}{4}$ or has a null-vector, then $e(s, z) = \Omega(1)$.*

(3) *If Γ is co-compact or a subgroup of finite index in $\Gamma(1) = SL(2, \mathbf{Z})$, then for all $\delta > 0$*

$$e(s, z) = \Omega((\log s)^{1/4 - \delta}). \tag{1.23}$$

In view of the results of Phillips and Sarnak [15] and Wolpert [20] one might expect that the "generic" group in Teichmüller space has no cusp forms with eigenvalue $\lambda > 1/4$. However, at present no group is known to have this property.

In Section 5, we sketch the extension of these results to hyperbolic n-space. Most of them carry over, but there are differences. Although the main term in the asymptotic distribution of the lattice points is, as expected,

$$N(s, z) = \frac{\omega_n}{2^{n-1}(n-1)\,\mathrm{vol}(\Gamma)}\, e^{(n-1)s}, \tag{1.24}$$

the error term can grow like $e^{(n-2)s}$. This is to be compared to the known upper bound of $O(s^{3/(n+1)} e^{(n-2+2/(n+1))s})$ in [10]. For $n \geqslant 4$ this is significantly larger then $e^{(n-1)s/2}$, which one may expect from our data when $n = 2$.

2. The Mean Normalized Remainder

The mean normalized remainder is defined by

$$\mathrm{MNR} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{N(s, z) - \Sigma(s, z)}{e^{s/2}}\, ds, \tag{2.1}$$

where $N(s, z)$ and $\Sigma(s, z)$ are defined in (1.10) and (1.13).

THEOREM 2.1. *Denoting the Eisenstein series for the jth cusp by $E_j(z, s)$, we have*

$$\mathrm{MNR} = \sum_j |E_j(z, \tfrac{1}{2})|^2. \tag{2.2}$$

Our proof of this follows a by now classical path. We first define a "point-pair invariant" k_s:

$$k_s(z, z') = k_s(t(z, z')) = \begin{cases} 1, & \operatorname{dist}(z, z') \leqslant s \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Here

$$t(z, z') = \frac{|z - z'|^2}{yy'} = 2\cosh \operatorname{dist}(z, z') - 2. \tag{2.4}$$

Setting

$$K_s(z, z') = \sum_{\gamma \in \Gamma} k_s(\gamma z, z'), \tag{2.5}$$

we see that

$$N(s, z) = K_s(z, z). \tag{2.6}$$

We now express $K_s(z, z')$ in terms of the spectral function $h_s(r)$, and the eigenfunctions of the Laplacian (this is sometimes called the "pre-trace formula"):

$$K_s(z, z') = \sum h_s(r_j)\, \phi_j(z)\, \overline{\phi_j(z')}$$
$$+ \sum_j \frac{1}{4\pi} \int_{-\infty}^{\infty} h_s(r)\, E_j(z, 1/2 + ir)\, \overline{E_j(z', 1/2 + ir)}\, dr.$$

This expression converges in the strong operator topology. The mapping $k_s \to h_s$, called the Selberg transform, is realized as follows [17]:

$$\begin{aligned} Q_s(w) &= \int_w^{\infty} \frac{k_s(t)}{\sqrt{t - w}}\, dt \\ &= \begin{cases} 2(t - w)^{1/2}, & \text{if } \ 0 \leqslant \operatorname{dist}(z, z') \leqslant s \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.8}$$

Replacing t by $2\cosh s - 2$ and w by $2\cosh u - 2$, this becomes

$$g_s(u) = \begin{cases} 2^{3/2}(\cosh s - \cosh u)^{1/2}, & |u| \leqslant s \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

The spectral function $h_s(r)$ is the Fourier transform of $g_s(u)$:

$$h_s(r)=\int_{-s}^{s} g_s(u)e^{iru}\,du. \tag{2.10}$$

We note that $g_s(u)$ is even in u, and $h_s(r)$ is even in both s and r.

We proceed now to estimate the terms in (2.7). The main term, $\Sigma(s, z)$, is obtained by approximating $h_s(r)$ at imaginary values of r.

LEMMA 2.1. *For* $\operatorname{Im} r>0$,

$$h_s(r)=\sqrt{\pi}\,\frac{\Gamma(|r|)}{\Gamma(|r|+3/2)}\,e^{(1/2+|r|)s}+O(e^{(|r|-1/2)s}). \tag{2.11}$$

Proof. Let $p=|r|$. Then $h_s(ip)$ can be rewritten as

$$h_s(ip)=2^{5/2}\int_0^s(\cosh s-\cosh u)^{1/2}\cosh(pu)\,du. \tag{2.12}$$

Next make the substitution $X=e^s$, $Xy=e^u$ and get

$$h_s(ip)=2X^{1/2}\int_{1/X}^{1}(1-y)^{1/2}\left(1-\frac{1}{X^2y}\right)^{1/2}((Xy)^p+(Xy)^{-p})\,\frac{dy}{y}. \tag{2.13}$$

This can be evaluated from the following two estimates, which we state with the n-dimensional case in mind: For $p>0$ and $n>1$,

$$\begin{aligned}&\int_{1/X}^{1}(1-y)^{(n-1)/2}\,\frac{1}{y^p}\,\frac{dy}{y}=\frac{X^p}{p}+O(X^{p-1}),\\&\int_{1/X}^{1}(1-y)^{(n-1)/2}\,y^p\,\frac{dy}{y}=\frac{\Gamma((n+1)/2)\,\Gamma(p)}{\Gamma((n+1)/2+p)}-\frac{1}{pX^p}+O\left(\frac{1}{X^{p+1}}\right).\end{aligned} \tag{2.14}$$

Substituting these relations (with $n=2$) into (2.13) we obtain (2.11). For $p=\frac{1}{2}$ $(n=2)$, (2.12) can be integrated directly and we get

$$h_s\left(\frac{i}{2}\right)=\pi(e^s-2+e^{-s}). \quad ∎ \tag{2.15}$$

The expression (2.13) with $p=0$ can be evaluated by first integrating by parts to get

$$h_s(0)=c_n X^{(n-1)/2}\int_{1/X}^{1}(1-y)^{(n-1)/2}\,d\log y+O(X^{(n-3)/2})$$

$$=c_n X^{(n-1)/2}\left[\log y\,(1-y)^{(n-1)/2}\big|_{1/X}^{1}+\frac{n-1}{2}\int_{1/X}^{1}\log y\,(1-y)^{(n-3)/2}\,dy\right]$$

$$+O(X^{(n-3)/2})$$

$$=c_n X^{(n-1)/2}s+\frac{n-1}{2}\int_0^1\log y\,(1-y)^{(n-3)/2}\,dy+O(X^{(n-3)/2}\log X),$$

and then evaluating the remaining integral by differentiating

$$\int_0^1 (1-y)^{(n-3)/2}\,y^p\,dy=\frac{\Gamma((n-1)/2)\,\Gamma(p)}{\Gamma((n-1)/2+p)}$$

with respect to p at $p=0$. We end up with

LEMMA 2.2.

$$h_s(0)=4(s+2(\log 2-1))e^{s/2}+O(e^{-s/2}). \tag{2.16}$$

Note that the remainder terms in Lemmas 2.1, 2.2 do not affect the MNR and hence are omitted from the main term $\Sigma(s,z)$. It remains to estimate the terms in (2.7) for real r. We begin with:

LEMMA 2.3. *For each* z,

$$\sum_{|r_j|<R}|\phi_j(z)|^2+\sum_j\frac{1}{4\pi}\int_{-R}^{R}\left|E_j\left(z,\frac{1}{2}+ir\right)\right|^2dr\sim cR^2 \quad \textit{as} \quad R\to\infty, \tag{2.17}$$

where $c=c(z)$ *depends only on the number of elements of* Γ *fixing* z.

Proof. The kernel for the automorphic heat equation can be written as

$$G(z,z,t)=\sum e^{-\lambda_j t}\,|\phi_j(z)|^2+\sum_j\frac{1}{4\pi}\int_{-\infty}^{\infty}e^{-\lambda t}\left|E_j\left(z,\frac{1}{2}+ir\right)\right|^2dr, \tag{2.18}$$

where as usual, $\lambda=\frac{1}{4}+r^2$. Since

$$G(z,z,t)\sim\frac{c}{t}, \qquad \text{as} \quad t\to 0, \tag{2.19}$$

where $c = c(z)$ depends only on the number of elements of Γ fixing z, it follows from Karamata's Tauberian theorem that

$$\sum_{|r_j| < R} |\phi_j(z)|^2 + \sum_j \frac{1}{4\pi} \int_{-R}^{R} \left| E_j\left(z, \frac{1}{2} + ir\right) \right|^2 dr \sim cR^2 \qquad \text{as} \quad R \to \infty, \tag{2.20}$$

as desired. ∎

The spectral function of the mean normalized K_s is

$$H_T(r) = \frac{1}{T} \int_0^T e^{-s/2} h_s(r)\, ds. \tag{2.21}$$

LEMMA 2.4.

$$\lim_{T \to \infty} \int_{-\infty}^{\infty} H_T(r)\, dr = 4\pi. \tag{2.22}$$

Proof. Interchanging the order of integration, we can rewrite (2.21) as a Fourier transform,

$$H_T(r) = \int_{-\infty}^{\infty} e^{iru} \Phi_T(u)\, du, \tag{2.23}$$

where

$$\Phi_T(u) = \begin{cases} 2^{3/2}/T \int_{|u|}^{T} e^{-s/2} (\cosh s - \cosh u)^{1/2}\, ds, & |u| < T \\ 0, & \text{otherwise.} \end{cases} \tag{2.24}$$

Thus $\Phi_T(u)$ is continuous and compactly supported so that

$$\begin{aligned} \int_{-\infty}^{\infty} H_T(r)\, dr &= 2\pi \Phi_T(0) \\ &= \frac{4\pi}{T} \int_0^T e^{-s/2} (e^{s/2} - e^{-s/2})\, ds \\ &= \frac{4\pi}{T} (T - 1 + e^{-T}) \to 4\pi \qquad \text{as} \quad T \to \infty. \ \blacksquare \end{aligned} \tag{2.25}$$

Next we give estimates on the spectral functions $h_s(r)$ and $H_T(r)$.

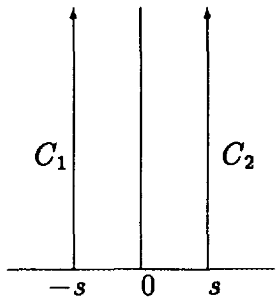

FIGURE 1

LEMMA 2.5. *If* $h_s'(r) = 2e^{-s/2} \int_{-s}^{s} (e^s + e^{-s} - e^u - e^{-u})^{1/2} e^{iru}\, du$, *then*

$$h_s'(r) = \begin{cases} \alpha(r) e^{irs} + \overline{\alpha(r)} e^{-irs} + O(e^{-2s}/r^{3/2}), & r > 0 \\ \overline{\alpha(-r)} e^{irs} + \alpha(-r) e^{-irs} + O(e^{-2s}/|r|^{3/2}), & r < 0, \end{cases} \tag{2.26}$$

where

$$\alpha(r) = \frac{-2i}{1 - e^{-2\pi r}} \int_0^{2\pi} (1 - e^{iv})^{1/2}\, e^{-rv}\, dv. \tag{2.27}$$

Proof. We first evaluate $h_s'(r)$ by deforming the path of integration from $-s$ to s into two vertical half-lines C_1 and C_2, one from $-s$ to $-s + i\infty$ and the other from s to $s + i\infty$, as in Fig. 1.

For $r > 0$ we then obtain

$$h_s'(r) = 2 \int_{-s}^{s} (1 + e^{-2s} - e^{-s+u} - e^{-s-u})^{1/2}\, e^{iru}\, du$$

$$= \int_{C_1} - \int_{C_2}$$

$$= 2 \int_0^{\infty} (1 + e^{-2s} - e^{-s+(-s+iv)} - e^{-s-(-s+iv)})^{1/2}\, e^{ir(-s+iv)} i\, dv$$

$$- 2 \int_0^{\infty} (1 + e^{-2s} - e^{-s+(s+iv)} - e^{-s-(s+iv)})^{1/2}\, e^{ir(s+iv)} i\, dv$$

$$= 2i \int_0^{\infty} (1 - e^{-iv} + e^{-2s}(1 - e^{iv}))^{1/2}\, e^{-rv}\, dv \cdot e^{-irs}$$

$$- 2i \int_0^{\infty} (1 - e^{iv} + e^{-2s}(1 - e^{-iv}))^{1/2}\, e^{-rv}\, dv \cdot e^{irs}$$

and using $1-e^{iv}=-e^{iv}(1-e^{-iv})$, this becomes

$$=2i\int_0^\infty (1-e^{-iv})^{1/2}(1-e^{-2s}e^{iv})^{1/2}e^{-rv}\,dv\cdot e^{-irs}$$

$$-2i\int_0^\infty (1-e^{iv})^{1/2}(1-e^{-2s}e^{-iv})^{1/2}e^{-rv}\,dv\cdot e^{irs}.$$

Thus we see that for $r>0$,

$$h_s'(r)=-2i\int_0^\infty (1-e^{iv})^{1/2}e^{-rv}\,dv\cdot e^{irs}$$

$$+2i\int_0^\infty (1-e^{-iv})^{1/2}e^{-rv}\,dv\cdot e^{-irs}+O\left(\frac{e^{-2s}}{r^{3/2}}\right),$$

which is the required answer upon noting that

$$-2i\int_0^\infty (1-e^{iv})^{1/2}e^{-rv}\,dv=\frac{-2i}{1-e^{-2\pi r}}\int_0^{2\pi}(1-e^{iv})^{1/2}e^{-rv}\,dv=\alpha(r). \quad (2.30)$$

For $r<0$, the computation is similar, except that we deform the original contour of integration into the lower half-plane as in Fig 2. ▌

LEMMA 2.6. (1) *$\alpha(r)$ is meromorphic, with its only poles being simple poles at $r\in i\mathbf{Z}$.*

(2) $\alpha(r)=c/r^{3/2}+O(1/r^{5/2})$, $c=\sqrt{\pi}\,e^{-i3\pi/4}$.

(3) $\overline{\alpha(-\bar{r})}=\alpha(r)$ *for all* $r\in\mathbf{C}$.

(4) *For r real,* $\mathrm{Im}(e^{irs}\alpha(r))$ *is an odd function of r.*

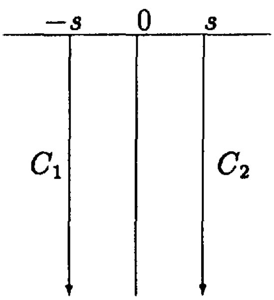

FIGURE 2

Proof. (1) This is clear from the integral representation (2.27).

(2) Since for $0 \leqslant v < 2\pi$,

$$(1 - e^{iv})^{1/2} = \left(2 \sin \frac{v}{2}\right)^{1/2} e^{i((v-\pi)/4)} \tag{2.31}$$

we see that as $r \to \infty$,

$$\alpha(r) = \frac{c}{r^{3/2}} + O\left(\frac{1}{r^{5/2}}\right), \qquad c = \sqrt{\pi}\, e^{-i3\pi/4}. \tag{2.32}$$

(3) Since both $\alpha(r)$ and $\overline{\alpha(-\bar{r})}$ are meromorphic, it suffices to check this for r pure imaginary. Thus we need to show that $\alpha(it) = \overline{\alpha(it)}$ for t real, $t \notin \mathbf{Z}$. We have

$$\alpha(it) = \frac{-2i}{1 - e^{-2\pi it}} \int_0^{2\pi} (1 - e^{iv})^{1/2}\, e^{-itv}\, dv \tag{2.33}$$

changing variables to $u = 2\pi - v$,

$$= \frac{-2ie^{-2\pi it}}{1 - e^{-2\pi it}} \int_0^{2\pi} (1 - e^{-iu})^{1/2}\, e^{itu}\, du$$

$$= \overline{\alpha(it)}.$$

(4) Clear from (3). ∎

Lemma 2.7. (1) $H_T(r) = O(1/r)$;

(2) $H_T(r) = O(r^{-5/2}/T)$.

Proof. (1) It is clear from (2.28) that

$$|h_s'(r)| \ll \int_0^\infty e^{-rv}\, dv = \frac{1}{r} \tag{2.34}$$

and so

$$|H_T(r)| = \left| \frac{1}{T} \int_0^T e^{-s/2} h_s(r)\, ds \right| \ll \max_s |h_s'(r)| \ll \frac{1}{r}. \tag{2.35}$$

(2) We apply the contour-shifting idea of Lemma 2.5 to the 0 to T integration, substituting $s = \pm i\tau$ on one leg and $s = T \pm i\tau$ on the other. We end up with

$$TH_T(r) = \left(\int_0^\infty \int_0^\infty ((1 - e^{-iv}) + e^{2i\tau}(1 - e^{iv}))^{1/2} e^{-r\tau} e^{-rv} \, dv \, d\tau \right.$$

$$\left. - \int_0^\infty \int_0^\infty ((1 - e^{-iv}) + e^{-2T} e^{2i\tau}(1 - e^{iv}))^{1/2} e^{-irT} e^{-r\tau} e^{-rv} \, dv \, d\tau \right)$$

$$+ \left(\int_0^\infty \int_0^\infty ((1 - e^{iv}) + e^{-2i\tau}(1 - e^{-iv}))^{1/2} e^{-r\tau} e^{-rv} \, dv \, d\tau \right.$$

$$\left. - \int_0^\infty \int_0^\infty ((1 - e^{iv}) + e^{-2T} e^{-2i\tau}(1 - e^{-iv}))^{1/2} e^{irT} e^{-r\tau} e^{-rv} \, dv \, d\tau \right). \tag{2.36}$$

If we bound the square root expression in all the integrals by $v^{1/2}$, we see that

$$|TH_T(r)| \ll \int_0^\infty \int_0^\infty v^{1/2} e^{-r\tau} e^{-rv} \, dv \, d\tau \ll r^{-5/2}. \tag{2.37}$$

This completes the proof of the lemma. ∎

We are now ready to complete the proof of Theorem 1.1. It follows from Lemmas 2.1 and 2.2 that

$$\frac{1}{T} \int_0^T (N(s, z) - \Sigma(s, z)) \, e^{-s/2} \, ds \tag{2.38}$$

$$= \sum_j H_T(r_j) \, |\phi_j(z)|^2 + \sum_j \frac{1}{4\pi} \int_{-\infty}^\infty H_T(r) \left| E_j \left(z, \frac{1}{2} + ir \right) \right|^2 dr,$$

where the first sum on the right is over the eigenfunctions with real non-zero r. Let

$$M_1(R) = \sum_{|r_j| \leq R} |\phi_j(z)|^2 \tag{2.39}$$

$$M_2(R) = \sum_j \frac{1}{4\pi} \int_{-R}^R \left| E_j \left(z, \frac{1}{2} + ir \right) \right|^2 dr. \tag{2.40}$$

According to Lemma 2.3, both these quantities are $O(R^2)$. We can therefore bound the sum on the right in (2.38) by

$$\frac{c}{T}\int_\delta^\infty r^{-5/2}\, dM_1(r) \ll \frac{c}{T}\int_\delta^\infty r^{-7/2} M_1(r)\, dr = O\left(\frac{1}{T}\right). \tag{2.41}$$

The second term in (2.38) is more interesting. If all the Eisenstein series vanish at $r=0$, then $\sum_j |E(z, \frac{1}{2}+ir)|^2 \ll r$ for small r and by Lemma 2.7(1),

$$\int_{-\delta}^{\delta} |H_T(r)| \sum_j \left|E_j\left(z, \frac{1}{2}+ir\right)\right|^2 dr \ll \int_0^\delta \frac{c}{r}\, r\, dr = O(\delta). \tag{2.42}$$

Arguing as in (2.41) we see that

$$\left|\int_\delta^\infty H_T(r)\, dM_2(r)\right| = O\left(\frac{1}{T}\right). \tag{2.43}$$

Since $\delta > 0$ is arbitrary, it follows that MNR $= 0$ in this case.

Finally, if $E_j(z, \frac{1}{2}) \neq 0$ for some j (i.e., there is a null vector), we write the integral as

$$\begin{aligned} \int_{-\infty}^{\infty} & H_T(r)\, |E_j(z, \tfrac{1}{2}+ir)|^2\, dr \\ &= \int_{-\infty}^{\infty} H_T(r)\, |E_j(z, \tfrac{1}{2})|^2\, dr \\ &\quad + \int_{-\infty}^{\infty} H_T(r)\, (|E_j(z, \tfrac{1}{2}+ir)|^2 - |E_j(z, \tfrac{1}{2})|^2)\, dr. \end{aligned} \tag{2.44}$$

Since $|E_j(z, \frac{1}{2}+ir)|^2 - |E_j(z, \frac{1}{2})|^2 = O(r)$ for small r, it follows as above that the second expression on the right in (2.44) tends to zero as $T \to \infty$. Making use of Lemma 2.4, we obtain the stated result (2.2).

3. Lower Bounds

3.1. *Preliminaries*

Let $\psi \in C_c^\infty(\mathbf{R})$ with $\operatorname{Supp} \psi \subseteq [-1, 1]$, ψ even, $\psi \geqslant 0$, $\hat{\psi} \geqslant 0$, and $\int \psi(x)\, dx = 1$. For $\varepsilon > 0$, set $\psi_\varepsilon(x) = (1/\varepsilon)\, \psi(x/\varepsilon)$. Then $0 \leqslant \hat{\psi}_\varepsilon \leqslant 1$, $\hat{\psi}_\varepsilon(0) = 1$. We consider a normalized smoothed remainder term

$$e_\varepsilon(s) = \int_{-\infty}^{\infty} \psi_\varepsilon(s-x)\, e(x)\, dx. \tag{3.1}$$

$e_\varepsilon(s)$ is a smooth average of $e(s)$ over an interval of radius ε around s. We first prove part (iii) of Theorem 1.2.

To show that $\limsup |e(s)| = \infty$, it clearly suffices to do the same for the smooth version, since if $|e(s)| < K$ in an ε-interval about s, then $|e_\varepsilon(s)| \leqslant \int \psi_\varepsilon(s-x)\, K\, dx = K$. The advantage of using a smoothed remainder term is that we then have (in the co-compact case) a pointwise equality

$$e_\varepsilon(s) = \sum_{0 \neq r_j \in \mathbf{R}} h'_{s,\varepsilon}(r_j)\, B(r_j), \tag{3.2}$$

where $B(r) = B(r, z)$ is

$$B(r) = \sum_{r_j = r} |\phi_j(z)|^2 \tag{3.3}$$

the sum running over an orthonormal basis of eigenfunctions with eigenvalue $\lambda = 1/4 + r^2$, and

$$h'_{s,\varepsilon}(r) = \int \psi_\varepsilon(s-x)\, h'_x(r)\, dx \tag{3.4}$$

$$h'_s(r) = 2^{3/2} e^{-s/2} \int_{-s}^{s} (\cosh(s) - \cosh(u))^{1/2}\, e^{iru}\, du. \tag{3.5}$$

3.2. *Asymptotics*

We begin with a lemma on the asymptotics of the spectral functions. As in Section 2, we set for $r > 0$

$$\alpha(r) = \frac{-2i}{1 - e^{-2\pi r}} \int_0^{2\pi} (1 - e^{iv})^{1/2}\, e^{-rv}\, dv. \tag{3.6}$$

LEMMA 3.1. *For $r > 0$, we have*

$$h'_{s,\varepsilon}(r) = 2 \operatorname{Re} \alpha(r)\, \hat{\psi}_\varepsilon(r)\, e^{irs} + O\left(\frac{e^{-2s}}{r^N}\right), \tag{3.7}$$

for all $N \gg 1$.

Proof. We have from Lemma 2.5 that

$$h'_{s,\varepsilon}(r) = 4 \operatorname{Re} i \int \psi_\varepsilon(s-x)\, e^{-irx} \int_0^\infty (1-e^{-iv})^{1/2} (1-e^{iv}e^{-2x})^{1/2} e^{-rv}\, dv\, dx$$

$$= 2 \operatorname{Re} \int \psi_\varepsilon(s-x)\, e^{-irx} \bar{\alpha}(r)\, dx$$

$$+ 4 \operatorname{Re} i \int \psi_\varepsilon(s-x)\, e^{-irx} \int_0^\infty (1-e^{-iv})^{1/2}$$

$$\times [(1-e^{iv}e^{-2x})^{1/2} - 1]\, e^{-rv}\, dv\, dx$$

$$= 2 \operatorname{Re} \alpha(r)\, \hat{\psi}_\varepsilon(r)\, e^{irs}$$

$$+ 4 \operatorname{Re} i \int_0^\infty e^{-rv} (1-e^{-iv})^{1/2} \int \psi_\varepsilon(s-x)$$

$$\times [(1-e^{iv}e^{-2x})^{1/2} - 1]\, e^{-irx}\, dx\, dv.$$

We are done if we show that for all $N \gg 1$,

$$\int \psi_\varepsilon(s-x) [(1-e^{iv}e^{-2x})^{1/2} - 1]\, e^{-irx}\, dx \ll \frac{e^{-2s}}{r^N}. \tag{3.8}$$

Indeed, setting $u(x) = (1-e^{iv}e^{-2x})^{1/2} - 1$, we see that $u(x)$ and all its derivatives are rapidly decaying in x in the support of $\psi_\varepsilon(s-x)$:

$$u^{(k)}(x) \ll_k e^{-2x}. \tag{3.9}$$

Upon integrating by parts N times, and noting that the integration is over an interval of radius ε around s, we see that

$$\int \psi_\varepsilon(s-x)\, u(x)\, e^{-irx}\, dx \ll \frac{e^{-2s}}{r^N}. \quad \blacksquare \tag{3.10}$$

3.3. *The Co-compact Case*

In this section we give lower bounds for the remainder term when the group Γ is co-compact.

LEMMA 3.2. *Let* $M_1(R) = \sum_{|r_j| < R} |\phi_j(z)|^2$.

(1) *If* Γ *is co-compact, then as* $R \to \infty$,

$$\sum_{|r_j| < R} \frac{B(r_j)}{r_j^\sigma} = \begin{cases} \dfrac{2c}{2-\sigma} R^{2-\sigma} + O(1), & \sigma \neq 2 \\ 2c \log R + O(1), & \sigma = 2. \end{cases} \tag{3.11}$$

(2) *In the general finite volume case, the sum* $\sum_{|r_j|<R}(B(r_j)/r_j^\sigma)$ *converges for* $\sigma > 2$.

Proof. This is an immediate application of partial summation to Lemma 2.3. Indeed, for Γ co-compact, we know that $M_1(R) \approx cR^2$ and so

$$\begin{aligned}\sum_{|r_j|<R} \frac{B(r_j)}{r_j^\sigma} &= \int_0^R \frac{dM_1(r)}{r^\sigma} \\ &= \frac{M_1(R)}{R^\sigma} + \sigma \int_0^R \frac{M_1(r)}{r^{\sigma+1}}\, dr \\ &\approx cR^{2-\sigma} + c\sigma \int_1^R r^{1-\sigma}\, dr + O(1) \\ &= \begin{cases} \dfrac{2c}{2-\sigma} R^{2-\sigma} + O(1), & \sigma \neq 2 \\ 2c \log R + O(1), & \sigma = 2. \end{cases}\end{aligned} \tag{3.12}$$

The claim for Γ non-compact follows in the same manner, except that then we only have an upper bound $M_1(R) \ll R^2$. ∎

THEOREM 3.1. *If the group* Γ *is co-compact, then for all* $\delta > 0$,

$$e(s) = \Omega((\log s)^{1/4-\delta}). \tag{3.13}$$

Proof. We have

$$e_\varepsilon(s) = \sum_{0 \neq r_j \in \mathbf{R}} h'_{s,\varepsilon}(r_j)\, B(r_j).$$

Using Lemmas 2.6 and 3.1, we get

$$e_\varepsilon(s) = 4\,\mathrm{Re} \sum_{r_j>0} \alpha(r_j)\, \hat{\psi}_\varepsilon(r_j)\, e^{ir_js} B(r_j) + O(e^{-2s}).$$

Divide the above sum into two parts, one for $r_j < R$ and the other for $r_j > R$. The latter sum can be estimated using $\hat{\psi}_\varepsilon(r) \ll_k (\varepsilon r)^{-k}$ as

$$\sum_{r_j>R} \alpha(r_j)\, \hat{\psi}_\varepsilon(r_j)\, e^{ir_js}\, B(r_j) \ll_k \sum_{r_j>R} \frac{1}{r_j^{3/2}} \frac{1}{(\varepsilon r_j)^k} B(r_j) \ll \frac{1}{\varepsilon^k} \frac{1}{R^{k-1/2}}. \tag{3.14}$$

For the sum over $r_j < R$, using Lemma 3.1 to write

$$\alpha(r) = \frac{c}{r^{3/2}} + O\left(\frac{1}{r^{5/2}}\right),$$

we have

$$\operatorname{Re} \sum_{r_j < R} \alpha(r_j)\, \hat{\psi}_\varepsilon(r_j)\, e^{ir_j s}\, B(r_j) = \operatorname{Re} c \sum_{r_j < R} e^{ir_j s} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) + O\left(\sum_{r_j < R} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{5/2}} B(r_j) \right) \tag{3.15}$$

and using $\hat{\psi}_\varepsilon(r) < 1$ we find as above that the O term in (3.15) is $O(1)$. Putting (3.14) and (3.15) together, we find that

$$e_\varepsilon(s) = 4 \operatorname{Re} c \sum_{r_j < R} e^{ir_j s} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) + O_k\left(\frac{1}{\varepsilon^k R^{k-1/2}} \right). \tag{3.16}$$

We need a lemma in order to "line up" the phases in (3.16).

LEMMA 3.3 (Dirichlet Box Principle). *Given n distinct real numbers $r_1, ..., r_n$ and $M > 0$, $T > 1$, there is an s, $M \leqslant s \leqslant MT^n$ such that*

$$|e^{ir_j s} - 1| < 1/T, \qquad j = 1, ..., n.$$

Thus denoting by $n(\mathrm{R})$ the number of *distinct* eigenvalues $r_j < R$, we can find for any $T > 1$ and $M > 0$ some s, $M \leqslant s \leqslant MT^{n(R)}$ such that

$$\begin{aligned} \sum_{r_j < R} e^{ir_j s} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) &= \sum_{r_j < R} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) + O\left(\frac{1}{T} \sum_{r_j < R} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) \right) \\ &= \sum_{r_j < R} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) + O\left(\frac{R^{1/2}}{T} \right). \end{aligned} \tag{3.17}$$

To conclude, we have shown that given $T > 1$, $M > 0$, there is an s, $M \leqslant s \leqslant MT^{n(R)}$, such that

$$e_\varepsilon(s) = \operatorname{Re} c \sum_{r_j < R} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) + O_k\left(\frac{R^{1/2}}{T} + \frac{1}{\varepsilon^k R^{k-1/2}} \right). \tag{3.18}$$

We can now choose $T = R^{1/2}$, $1/\varepsilon = R^{1-1/(2k)}$. Furthermore, choose τ, $0 < \tau < 1$ such that $\hat{\psi}(x) \geqslant \frac{1}{2}$ if $|x| < \tau$. Then since $\hat{\psi}_\varepsilon(r) \geqslant 0$ we have

$$\begin{aligned} \sum_{r_j < R} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) &\geqslant \sum_{r_j < \tau/\varepsilon} \frac{\hat{\psi}_\varepsilon(r_j)}{r_j^{3/2}} B(r_j) \\ &\geqslant \frac{1}{2} \sum_{r_j < \tau/\varepsilon} \frac{B(r_j)}{r_j^{3/2}} \approx \kappa \varepsilon^{-1/2}. \end{aligned} \tag{3.19}$$

In particular, we see that there are arbitrarily large values of s for which

$$|e_\varepsilon(s)| \gg \varepsilon^{-1/2}. \tag{3.20}$$

From (3.20) and the choice of s in Lemma 3.3, namely $M \leqslant s \leqslant MT^{n(R)} \ll R^{R^2/2}$, we see that $R^2 \gg \log s/\log\log s$, and so from $1/\varepsilon = R^{1-1/(2k)}$ we see

$$|e_\varepsilon(s)| \gg_\delta (\log s)^{1/4-\delta} \tag{3.21}$$

for all $\delta > 0$, and the same must then be true of the unsmoothed remainder $e(s)$. ∎

3.4. *The Non-compact Case*

In the finite volume, non-compact case we have to handle, in addition to the discrete spectrum, the contribution of the Eisenstein series. We have

$$e_\varepsilon(s) = \sum_{0 \neq r_j \in \mathbf{R}} h'_{s,\varepsilon}(r_j)\, B(r_j) + \sum_{j=1}^{\kappa} \frac{1}{4\pi} \int_{-\infty}^{\infty} h'_{s,\varepsilon}(r) \left| E_j\left(z, \frac{1}{2} + ir\right)\right|^2 dr, \tag{3.22}$$

where $E_j(z, s)$ is the Eisenstein series associated to the jth cusp. We show that if $E_j(z, \frac{1}{2}) = 0$ the contribution of the jth cusp to the sum (3.22) is $o(1)$.

We have seen (Lemma 3.1) that for $r > 0$

$$h'_{s,\varepsilon}(r) = \operatorname{Re} e^{irs}\, \alpha(r)\, \hat{\psi}_\varepsilon(r) + O\left(\frac{e^{-2s}}{r^N}\right) \tag{3.23}$$

and likewise for $r < 0$

$$h'_{s,\varepsilon}(r) = \operatorname{Re} e^{irs}\, \overline{\alpha(-r)}\, \hat{\psi}_\varepsilon(r) + O\left(\frac{e^{-2s}}{r^N}\right). \tag{3.24}$$

Set

$$\theta(r) = \alpha(r)\, \hat{\psi}_\varepsilon(r)(|E(z, \tfrac{1}{2} + ir)|^2 - |E(z, \tfrac{1}{2})|^2) \tag{3.25}$$

Since both $\hat{\psi}_\varepsilon(r)$ and $|E(z, \frac{1}{2} + ir)|^2$ are even, Lemma 2.6 shows that $\theta(-r) = \overline{\theta(r)}$.

LEMMA 3.4. *$\theta(r)$ has a meromorphic extension for all r complex, its poles in the strip $|\mathrm{Im}(r)| < 1$ are those of $|E(z, \frac{1}{2} + ir)|^2$, and for r real, $\theta(r)$ is regular and rapidly decreasing.*

Proof. For r real we have

$$|E(z, \tfrac{1}{2}+ir)|^2 = E(z, \tfrac{1}{2}-ir)\, E(z, \tfrac{1}{2}+ir) \tag{3.26}$$

and the right-hand side of (3.26) is meromorphic in r. $\hat{\psi}_\varepsilon(r)$ is entire, and $\alpha(r)$ is meromorphic in r, with simple poles at the points $r \in i\mathbf{Z}$. At $r=0$ the simple pole of $\alpha(r)$ is cancelled by the zero of $|E(z, \frac{1}{2}+ir)|^2 - |E(z, \frac{1}{2})|^2$. To see that $\theta(r)$ is rapidly decreasing, use the fact that $\alpha(r)\,\hat{\psi}_\varepsilon(r)$ is rapidly decreasing, while $E(z, \frac{1}{2}+ir)$ grows polynomially in r [1]. ∎

3.5. *Decay of the Continuous Contribution*

We now show that the continuous spectrum contribution in (3.22) is decaying in s if there are no null vectors. Set

$$c_j(s) = \int_{-\infty}^{\infty} h'_{s,\varepsilon}(r)\, |E_j(z, \tfrac{1}{2}+ir)|^2\, dr. \tag{3.27}$$

LEMMA 3.5. $c_j(s) = 4\pi |E_j(z, \frac{1}{2})|^2 + \hat{\theta}_j(s) + O(e^{-s})$.

Proof. We have by Lemma 3.1

$$c(s) = \int_{-\infty}^{\infty} \operatorname{Re}(e^{irs}\alpha(r))\, \hat{\psi}_\varepsilon(r)\, |E(z, \tfrac{1}{2}+ir)|^2\, dr + O(e^{-2s}). \tag{3.28}$$

Since $\operatorname{Im}(e^{irs}\alpha(r))$ is an *odd* function while $\hat{\psi}_\varepsilon(r)$ and $|E(z, \frac{1}{2}+ir)|^2$ are *even*, we see that $\operatorname{Im}(\theta(r)e^{irs})$ is also odd, and so has mean zero. Now $h'_{s,\varepsilon} = \hat{g}_{s,\varepsilon}$, and hence (3.28) shows that

$$\begin{aligned} c(s) &= \int_{-\infty}^{\infty} \operatorname{Re}(\theta(r)\, e^{irs})\, dr + |E(z, \tfrac{1}{2})|^2 \int_{-\infty}^{\infty} h'_{s,\varepsilon}(r)\, dr + O(e^{-2s}) \\ &= \int_{-\infty}^{\infty} \theta(r)\, e^{irs}\, dr + 2\pi g_{s,\varepsilon}(0)\, |E(z, \tfrac{1}{2})|^2 + O(e^{-2s}) \\ &= \hat{\theta}(s) + 4\pi\, |E(z, \tfrac{1}{2})|^2 + O(e^{-s}), \end{aligned} \tag{3.29}$$

where we have used the Fourier inversion

$$\begin{aligned} \int_{-\infty}^{\infty} h'_{s,\varepsilon}(r)\, dr &= 2\pi g_{s,\varepsilon}(0) \\ &= 2\pi \int \psi_\varepsilon(s-x)\, e^{-x/2} g_x(0)\, dx \\ &= 4\pi(1 + O(e^{-s})). \quad ∎ \end{aligned} \tag{3.30}$$

Now using $\theta \in L^1(R)$, we immediately see:

COROLLARY 3.1 (Riemann–Lebesgue). *As* $s \to \infty$,

$$\int_{-\infty}^{\infty} h'_{s,\varepsilon}(r)\, |E_j(z, \tfrac{1}{2}+ir)|^2\, dr \to 4\pi\, |E_j(z, \tfrac{1}{2})|^2. \tag{3.31}$$

3.6. *Arithmetic Groups*

We can now use Corollary 3.1 and Theorem 3.1 to deduce a lower bound for $e(s)$ whenever the Fuchsian group Γ has as many Maass cusp forms, as in the co-compact case, in particular when Γ is a subgroup of $SL(2, \mathbf{Z})$.

THEOREM 3.2. *If the group* Γ *has sufficiently many cusp forms in the sense that* $|\{r_j \leqslant R\}| \gg R^2$, *then* (3.13) *holds.*

Proof. We have seen from Theorem 3.1 that the contribution of the cuspidal spectrum is unbounded if there are sufficiently many cusp forms, in fact if

$$\sum \frac{B(r_j)}{r_j^{3/2}} = \infty. \tag{3.32}$$

Corollary 3.1 shows that the continuous spectrum contributes a bounded amount. ∎

Remark. Subgroups of $SL(2, \mathbf{Z})$ satisfy our assumption—in fact, if Γ is a *congruence* subgroup, then it is known [18] that

$$|\{r_j \leqslant R\}| \approx \frac{\mathrm{vol}(\Gamma)}{4\pi} R^2, \qquad R \to \infty.$$

3.7. *Loose Ends*

If Γ has a null form ($E_j(z, \frac{1}{2}) \neq 0$ for some j) then the results of Section 2 show that $e(s) = \Omega(1)$. To finish the proof of Theorem 1.2, we still need to consider the (hypothetical) case when Γ has no null forms, and either has no cusp forms or does have discrete spectrum above $\lambda > \frac{1}{4}$, but not too much of it in the sense that

$$\sum_j \frac{B(r_j)}{r_j^{3/2}} < \infty \tag{3.33}$$

(no such example is known). In the latter case, in (3.22) the contribution of the discrete spectrum to $e_\varepsilon(s)$ is an almost periodic function if we assume (3.33), and so is $\Omega(1)$, while the continuous spectrum contributes a term which decays as $s \to \infty$. Thus we see that for such Γ, we get $e(s) = \Omega(1)$.

We now assume that Γ has no cusp forms with eigenvalue $\lambda > 1/4$. From the work of Phillips and Sarnak [15] and the recent work of Wolpert [20], it appears that there are good reasons for believing that this is the case for *generic* Fuchsian groups Γ which are not co-compact. We show that in this case, assuming all Eisenstein series $E(\cdot, \frac{1}{2})$ vanish at z, the normalized remainder term $e(s)$ cannot be exponentially small. The mechanism responsible for this is a theorem of Selberg [19], which asserts that if there are no cusp forms then the Eisenstein series $E(z, s)$ have poles arbitrarily close to the "critical line" $\operatorname{Re}(s) = \frac{1}{2}$.

THEOREM 3.3. *Assume that Γ has no cusp forms with $r_j \neq 0$ and that all Eisenstein series vanish at $s = \frac{1}{2}$: $E_j(z, \frac{1}{2}) = 0$ for all $1 \leq j \leq \kappa$. Then for all $\delta > 0$ we have*

$$e(s) = \Omega(e^{-\delta s}). \tag{3.34}$$

Proof. As this is a lower bound, it suffices to prove it for the smoothed remainder term $e_\varepsilon(s)$. If there is no discrete spectrum in (3.22), then by Lemma 3.5,

$$e_\varepsilon(s) = \sum \hat{\theta}_j(s) + O(e^{-2s}), \tag{3.35}$$

with $\theta_j(r)$ given in (3.25). Suppose by contradiction that for some $K > 0$,

$$|e_\varepsilon(s)| \leq K e^{-\delta s}. \tag{3.36}$$

$e_\varepsilon(s)$, defined for $s < 0$ by $e_\varepsilon(-s) = e_\varepsilon(s)$, is smooth and exponentially decaying by assumption. We may use Fourier inversion on (3.35) to find

$$\sum_{j=1}^{\kappa} \theta_j(r) = \int_{-\infty}^{\infty} e_\varepsilon(s)\, e^{irs}\, ds + f(r), \tag{3.37}$$

where $f(r)$ comes from the error term (e^{-2s}) in (3.35) and so is holomorphic for $|\operatorname{Im}(r)| < 2$. Likewise, assuming (3.36) shows that the Fourier transform of $e_\varepsilon(s)$ is holomorphic in the strip $|\operatorname{Im}(r)| < \delta$. Thus $\sum_j \theta_j(r)$ is holomorphic in the strip $|\operatorname{Im}(r)| < \delta$.

By Lemma 3.4, the poles of $\theta_j(r)$ in the strip $|\operatorname{Im}(r)| < 1$ are those of the Eisenstein series $E_j(z, \frac{1}{2} + ir)$. In turn, the poles of the Eisenstein series are the same as those of the determinant of the scattering matrix $\phi(\frac{1}{2} + ir)$ [18]. Recall that for any group Γ, $\phi(s)$ is holomorphic in $\operatorname{Re}(s) \geq \frac{1}{2}$ except for poles in the real line. Let $N(\frac{1}{2}, T)$ be the number of poles $\rho = \beta + i\gamma$ of $\phi(s)$

with $\beta < \frac{1}{2}$, $|\gamma| < T$, and let $N(T)$ denote the number of Maass cusp forms with eigenvalue $\frac{1}{4}+r_j^2$, $0 < r_j < T$. Then Selberg [19] shows that for all Γ

$$N(T)+N\left(\frac{1}{2}, T\right)=\frac{\operatorname{vol}(\Gamma)}{4\pi} T^2+B_1 T \log T+O(T) \tag{3.38}$$

$$\sum_{\substack{\rho=\beta+i\gamma \\ 0 \leqslant \gamma \leqslant T}}\left(\frac{1}{2}-\beta\right)=\frac{\kappa}{4\pi} T \log T+B_2 T+O(\log T). \tag{3.39}$$

The sum in (3.39) over the poles is $\phi(s)$, and as usual κ is the number of cusps of Γ.

In our case, $N(T)=0$, and so (3.38) shows

$$N\left(\frac{1}{2}, T\right) \approx \frac{\operatorname{vol}(\Gamma)}{4\pi} T^2. \tag{3.40}$$

Our assumption (3.36) implied that there are no poles of $\theta_j(r)$ in the strip $|\operatorname{Im}(r)| < \delta$, and so $\beta \leqslant \frac{1}{2}-\delta$ for all summands in (3.39), giving

$$\sum_{\substack{\rho=\beta+i\gamma \\ 0 \leqslant \gamma \leqslant T}}\left(\frac{1}{2}-\beta\right) \geqslant \delta N\left(\frac{1}{2}, T\right) \approx \delta \frac{\operatorname{vol}(\Gamma)}{4\pi} T^2, \tag{3.41}$$

which contradicts (3.39). ∎

3.8. *Mean Square*

It is natural to try to check moments of the remainder term $e(s)$. For example, the numerical data (see Section 5) indicate that $e(s)$ has finite non-zero second moment. We have not been able to prove finiteness; however, from the discussion of this section one sees that the second moment is non-zero. Indeed, we saw that the smoothed remainder term $e_\varepsilon(s)$ is almost periodic, and so

$$V(\varepsilon)=\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |e_\varepsilon(s)|^2 \, ds$$

is finite and non-zero. We then see that for all $\varepsilon > 0$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |e(s)|^2 \, ds \geqslant V(\varepsilon). \tag{3.42}$$

Indeed, we have on using Cauchy–Schwartz (and $\int \psi_\varepsilon(x)\,dx = 1$)

$$|e_\varepsilon(s)|^2 = \left| \int_{|s-x|<\varepsilon} e(x)\,\psi_\varepsilon(s-x)\,dx \right|^2$$

$$\leqslant \int_{|s-x|<\varepsilon} |e(x)|^2\,\psi_\varepsilon(s-x)\,dx \tag{3.43}$$

and on integrating with respect to s we find

$$\int_0^T |e_\varepsilon(s)|^2\,ds \leqslant \int_0^T \int_{|s-x|<\varepsilon} |e(x)|^2\,\psi_\varepsilon(s-x)\,dx\,ds$$

$$\leqslant \int_0^{T+\varepsilon} |e(x)|^2 \int_{x-\varepsilon}^{x+\varepsilon} \psi_\varepsilon(s-x)\,ds\,dx$$

$$= \int_0^{T+\varepsilon} |e(x)|^2\,dx. \tag{3.44}$$

After dividing by T and letting $T \to \infty$, we get (3.42).

4. Hyperbolic n-Space

We again consider the lattice problem for an orbit of a discrete cofinite group Γ of motions of hyperbolic n-space $\mathbf{H}^n$. We take k_s as in (2.3), as our point-pair invariant. The Selberg transform in n dimensions is

$$Q_s(w) = \frac{\omega_{n-1}}{2} \int_w^\infty k_s(t)(t-w)^{(n-3)/2}\,dt$$

$$= \begin{cases} \dfrac{\omega_{n-1}}{n-1}(t-w)^{(n-1)/2} & 0 \leqslant \operatorname{dist}(z, z') \leqslant s \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The spectral function h_s is now given by

$$h_s(r) = c_n 2^{(n-1)/2} \int_{-s}^{s} (\cosh s - \cosh u)^{(n-1)/2}\,e^{iru}\,du, \tag{4.2}$$

where $c_n = \omega_{n-1}/(n-1)$.

As before, the complete asymptotic distribution $\Sigma_\Gamma(s, z)$ is computed from the exceptional eigenvalues. However, we may now have to include some of the lower order terms in $h_s(ip)$. For $p>0$ and $X=e^s$,

$$\begin{aligned} h_s(ip) &= c_n X^{p+(n-1)/2} \frac{\Gamma((n+1)/2)\,\Gamma(p)}{\Gamma((n+1)/2+p)} \\ &+ c_n \sum_{j=1}^{[p/2]} \frac{(-1)^j}{j!} \frac{\Gamma((n+1)/2)\,\Gamma(p-j)}{\Gamma((n+1)/2+p-j)} \frac{n-1}{2} \frac{n-3}{2} \cdots \frac{n+1-2j}{2} \\ &\times X^{p+(n+1)/2-2j} + o(X^{(n-1)/2}). \end{aligned} \tag{4.3}$$

To prove this, we again make the substitution $X=e^s$, $Xy=e^u$ after which $h_s(ip)$ can be written as

$$\begin{aligned} h_s(ip) &= c_n X^{(n-1)/2} \int_{1/X}^{1} (1-y)^{(n-1)/2} \left(1-\frac{1}{X^2 y}\right)^{(n-1)/2} \\ &\times ((Xy)^p + (Xy)^{-p}) \frac{dy}{y}. \end{aligned} \tag{4.4}$$

We then expand the expression $(1-X^{-2}y^{-1})^{(n-1)/2}$ in powers of X. The resulting terms have been evaluated in (2.14) and using this we obtain (4.3).

The analogue of Lemma 2.2, which we derive in the same way, is

$$h_s(0) = c_n \left(s + 2 \left(\log 2 - \left(1 + \frac{1}{3} + \frac{1}{5} \cdots \frac{1}{n-1} \right) \right) \right) \tag{4.5}$$

for n even, and

$$h_s(0) = c_n \left(s - \left(1 + \frac{1}{2} + \cdots \frac{2}{n-1} \right) \right) \tag{4.6}$$

for n odd. Further, instead of (2.17) we now have

$$\sum_{|r_j| \leqslant R} |\phi_j(z)|^2 + \sum_j \frac{1}{4\pi} \int_{-R}^{R} |E_j(z, (n-1)/2 + ir)|^2 \, dr \sim cR^n \qquad \text{as} \quad R \to \infty. \tag{4.7}$$

In defining the mean normalized spectral function we found that we had to take a smoother mean then that used for $n=2$. To this end, we introduce a function $\theta \in C_c^\infty(\mathbf{R})$ such that $\theta(s) \geqslant 0$, $\int_{-\infty}^{\infty} \theta(s)\, ds = 1$, $\operatorname{supp} \theta \subset [0, 1]$.

We now set

$$H_T(r)=\frac{1}{T}\int_0^T e^{(n-1)s/2}h_s(r)\,\theta\left(\frac{s}{T}\right)ds. \tag{4.8}$$

As before,

$$H_T(r)=\int_{-\infty}^{\infty} e^{iru}\Phi_T(u)\,du, \tag{4.9}$$

where

$$\Phi_T(u)=\begin{cases}(c_n 2^{(n-1)/2}/T) \\ \times\int_{|u|}^T e^{-(n-1)s/2}(\cosh s-\cosh u)^{(n-1)/2}\,\theta(s/T)\,ds, & |u|\leqslant T \\ 0, & \text{otherwise.}\end{cases} \tag{4.10}$$

It follows that

$$\int_{-\infty}^{\infty} H_T(r)\,dr=2\pi\Phi_T(0)\to 2\pi c_n, \qquad \text{as} \quad T\to\infty. \tag{4.11}$$

We now define the normalized spectral function as

$$h'_s(r)=e^{-(n-1)s/2}h_s(r)=c_n\int_{-s}^{s}\left(\frac{e^s+e^{-s}-e^u-e^{-u}}{e^s}\right)^{(n-1)/2} e^{iru}\,du, \tag{4.12}$$

in terms of which the analogue of Lemma 2.5 is

$$h'_s(r)=2\operatorname{Re}\alpha(|r|)\,e^{i|r|s}+O\left(\frac{e^{-2s}}{|r|^{(n+1)/2}}\right), \qquad r\neq 0, \tag{4.13}$$

where now

$$\alpha(r)=\frac{-ic_n}{1-e^{-2\pi r}}\int_0^{2\pi}(1-e^{it})^{(n-1)/2}\,e^{-rt}\,dt. \tag{4.14}$$

Likewise, Lemma 2.6 holds with part (2) replaced by

$$\alpha(r)=\frac{c}{r^{(n+1)/2}}+O\left(\frac{1}{r^{(n+3)/2}}\right). \tag{4.15}$$

In order to prove the extended Theorem 1.1 we still need the estimates

$$H_T(r)=O\left(\frac{1}{r}\right) \tag{4.16}$$

and

$$|H_T(r)| \leqslant \frac{c(N)}{Tr^N}, \qquad \text{for all} \quad N \geqslant 1. \tag{4.17}$$

The first of these is clear from the n-dimensional analogue of (2.28), namely

$$h_s'(r) = 2c_n \operatorname{Im} e^{irs} \int_0^\infty \psi(s, v, n, n)\, e^{-rv}\, dv, \tag{4.18}$$

where

$$\psi(s, v, n, m) = (1 - e^{iv})^{(n-1)/2} (1 - e^{-2s} e^{-iv})^{(m-1)/2}. \tag{4.19}$$

Using this we can write

$$H_T(r) = 2c_n \operatorname{Im} \int_0^1 \theta(s)\, e^{irsT} \int_0^\infty \psi(sT, v, n, n)\, e^{-rv}\, dv\, ds. \tag{4.20}$$

We now integrate by parts with respect to s to obtain

$$H_T(r) = 2c_n \operatorname{Im} \int_0^1 -\frac{e^{irsT}}{irT} \left\{ \theta'(s) \int_0^\infty \psi(sT, v, n, n)\, e^{-rv}\, dv \right.$$

$$\left. + (n-1)\, Te^{-2sT} \theta(s) \int_0^\infty \psi(sT, v, n, n-2)\, e^{-iv} e^{-rv}\, dv \right\} ds. \tag{4.21}$$

Taking absolute values, it is easy to see that the v-integration is bounded by $c \sup_v |1 - e^{-2sT} e^{-iv}|^{\mu} / r^{(n+1)/2}$, where $\mu = (n-1)/2$ for the first integral and $(n-3)/2$ for the second. The subsequent s-integration adds nothing to the bound for the first term. However, in the second term the exponential e^{-2sT} allows us to take advantage of the fact that θ vanishes at $s = 0$ to infinite order. Omitting the factor of $1/Tr^{(n+3)/2}$, we can express the remaining integral as

$$T \int_0^1 e^{-2sT} \theta(s) \sup_v |1 - e^{-2sT} e^{-iv}|^{\mu}\, ds$$

$$\ll \begin{cases} T \int_0^1 e^{-2sT} s^N\, ds \ll T^{-N}, & \mu \geqslant 0 \\ T \int_0^1 e^{-2sT} s^N (sT)^{\mu}\, ds \ll T^{-N}, & \mu < 0. \end{cases} \tag{4.22}$$

Thus after one integration by parts one finds $|H_T(r)| \ll 1/Tr^{(n+3)/2}$. We can continue to integrate by parts indefinitely, each time gaining a factor of $1/r$. This proves the estimate in (4.17).

The relations (4.11) and (4.17) show that $H_T(r)$ approximates a delta function as $T \to \infty$. Arguing as in the proof of Theorem 1.1 we get

THEOREM 4.1. (1) *If Γ is co-compact, then* $\lim_T MNR(T) = 0$.

(2) *If Γ is co-finite but not co-compact, then*

$$\lim_T MNR(T) = \frac{c_n}{2} \sum_j \left| E_j\left(z, \frac{n-1}{2}\right) \right|^2. \tag{4.23}$$

Lower bounds. By repeating the arguments of Section 3, one can get lower bounds analogous to Theorem 3.1, in the case Γ is co-compact or a congruence subgroup. In the latter case, one needs to know Weyl's law holds for cusp forms. This has recently been done by A. Reznikov [16]. For such Γ, we have for all $\delta > 0$,

$$e(s) = \Omega((\log s)^{(n-1)/2n - \delta}). \tag{4.24}$$

However, this is far from the truth in some cases, where the remainder term may oscillate much more.

EXAMPLE. For $n \geqslant 4$, we consider the quadratic form

$$S(x_1, ..., x_n) = x_1^2 + \cdots + x_n^2 - x_{n+1}^2. \tag{4.25}$$

The two-sheeted hyperboloid $C = \{x : S(x) = -1\}$ gives a model of hyperbolic n-space: If $C_1^+ = \{x : S(x) = -1, x_{n+1} > 0\}$, then the connected component of the identity $G = SO_e(n, 1)$ of the orthogonal group of S acts transitively on C_1^+, and the stabiliser of the point $P_0 = (0, ..., 0, 1) \in C_1^+$ is $K = SO(n)$. Thus $C_1^+ = G/K$. The metric is induced from the indefinite metric $dx_1^2 + \cdots dx_n^2 - dx_{n+1}^2$ on $\mathbf{R}^{n+1}$ and restricted to C has constant curvature -1. The Euclidean balls $B_T = \{x : x_1^2 + \cdots + x_{n+1}^2 \leqslant T^2\}$ intersect C_1^+ in geodesic balls centered at P_0, of geodesic radius $s \approx \log T$.

Let $\Gamma = G \cap SL_{n+1}(\mathbf{Z}) \subset O(S, \mathbf{Z})$; this is a lattice in G. Γ acts on the integer points $C_1^+ \cap \mathbf{Z}^{n+1}$ with finitely many orbits. For our purposes, there is no harm in pretending that there is only one orbit, so that counting Γ-translates of P_0 is equivalent to counting the number $N(T)$ of integer points on C_1^+:

$$N(T) = \{x \in \mathbf{Z}^{n+1} : S(x) = -1, x_1^2 + \cdots + x_{n+1}^2 \leqslant T^2\}. \tag{4.26}$$

To understand the oscillation of $N(T)$, look at the integer points lying on the boundary $\partial B_T \cap C_1^+$, for T an odd integer. Their number is

$$N(\partial B_T \cap C_1^+) = |\{x \in \mathbf{Z}^{n+1} : x_{n+1} > 0, S(x) = -1, x_1^2 + \cdots + x_{n+1}^2 = T^2\}|$$
$$= \left| \left\{ x \in \mathbf{Z}^{n+1} : x_{n+1} = \sqrt{(T^2+1)/2}, x_1^2 + \cdots + x_n^2 = \frac{T^2-1}{2} \right\} \right|. \tag{4.27}$$

Thus the remainder term $e(\log T)$ is at least as large as $N(\partial B_T \cap C_1^+)$, which is the number of integer solutions of the equation

$$x_1^2 + \cdots + x_n^2 = \frac{T^2-1}{2}. \tag{4.28}$$

It is known that the number of ways to represent an integer N as a sum of n squares for $n \geqslant 4$ is [2]

$$r_n(N) \sim c_n \mathscr{S}(N) N^{n/2-1}, \qquad \text{as} \quad N \to \infty, \tag{4.29}$$

where $c_n > 0$, the "singular series" $\mathscr{S}(N)$ depends on the divisibility properties of N, and $\mathscr{S}(N) \geqslant b_n > 0$. Thus (4.28) has $\gg T^{n-2} \gg e^{(n-2)s}$ solutions, and so the remainder term $e(s)$ can get as large as $e^{(n-2)s}$.

Note that this kind of "jump" in the size of the remainder term is not due to the existence of exceptional eigenvalues, as these contribute a continuous function of s to the lattice count. It is this same feature that causes the remainder term for the "circle problem" in $\mathbf{R}^n$ to be large if $n \geqslant 4$. We point out that for the hyperbolic plane ($n=2$), the above boundary behaviour is not noticeable, since as is well known, the number of ways of writing $N = x_1^2 + x_2^2$ is $O(N^\varepsilon)$, for all $\varepsilon > 0$.

5. APPENDIX: NUMERICAL RESULTS

We computed the number of lattice points $N_\Gamma(s, z)$ in circles about z for radii s starting at 6 and increasing in steps of 0.1 up to 20, for a variety of groups Γ and base points z. Tables III–X in this Appendix give the results at integer values of the radius for $N_\Gamma(s, z)$, $d(s, z) = N(s, z) - \Sigma(s, z)$ and $\alpha(s, z) = \log |d(s, z)|/s$ when the base point is $z = i$ and Γ is the Fermat group $\Phi(N)$ of level $N = 1$ through 8. Tables I and II list the mean normalized remainder term and the mean square of the normalized remainder term at integer radii (abstract from the data taken at steps of 0.1). We also gathered data for these Fermat groups and for semi-Fermat groups $\Psi(N)$ (i.e., $\sum p_i \equiv 0 \bmod N$ in (5.2) below), both at $z = i$ and at $z = 0.5i$.

TABLE II. Mean Square Normalized Remainder

	Fermat level N							
T	1	2	3	4	5	6	7	8
10	0.3109	0.1301	0.0705	0.0455	0.0390	0.0321	0.0337	0.0423
11	0.2974	0.1108	0.0666	0.0414	0.0398	0.0292	0.0304	0.0418
12	0.2786	0.1016	0.0632	0.0401	0.0402	0.0279	0.0277	0.0464
13	0.2832	0.0922	0.0688	0.0390	0.0360	0.0262	0.0249	0.0447
14	0.2886	0.0951	0.0652	0.0387	0.0336	0.0258	0.0233	0.0407
15	0.2898	0.1170	0.0612	0.0432	0.0398	0.0256	0.0243	0.0368
16	0.3073	0.1109	0.0686	0.0421	0.0420	0.0316	0.0339	0.0464
17	0.2932	0.1114	0.0753	0.0429	0.0401	0.0334	0.0331	0.0522
18	0.2783	0.1104	0.0727	0.0414	0.0398	0.0355	0.0356	0.0518
19	0.2790	0.1094	0.0711	0.0415	0.0427	0.0371	0.0352	0.0546
20	0.2713	0.1124	0.0695	0.0444	0.0410	0.0350	0.0363	0.0521

TABLE III. $\Gamma(2)$

s	$N(s)$	Remainder	α
10	11069	55.8	0.4021
11	30121	183.9	0.4741
12	81361	−16.4	0.2331
13	221525	318.3	0.4433
14	601909	606.9	0.4577
15	1634433	−75.7	0.2884
16	4441989	−1066.3	0.4357
17	12078397	920.6	0.4015
18	32831265	1280.4	0.3975
19	89236321	−4829.5	0.4464
20	242582909	311.3	0.287

TABLE IV. $\Phi(2)$

s	$N(s)$	Remainder	α
10	2761	7.7	0.204
11	7557	72.7	0.3897
12	20353	8.7	0.1798
13	55137	−164.7	0.3926
14	151137	811.5	0.4785
15	407805	−822.2	0.4475
16	1111017	253.2	0.3459
17	3016837	−2532.1	0.461
18	8207337	−159.1	0.2817
19	22307437	−2850.6	0.4187
20	60637297	−8352.4	0.4515

TABLE V. $\Phi(3)$

s	$N(s)$	Remainder	α
10	1241	17.3	0.2851
11	3353	26.7	0.2985
12	9013	−28.9	0.2804
13	24621	42.5	0.2884
14	67117	305.7	0.4087
15	181745	132.9	0.326
16	493065	−607.8	0.4006
17	1343445	1503.2	0.4303
18	3648253	476.9	0.3426
19	9912105	−3578.4	0.4307
20	26956829	3207.0	0.4037

TABLE VI. $\Phi(4)$

s	$N(s)$	Remainder	α
10	701	12.7	0.2539
11	1845	−26.1	0.2964
12	5125	38.9	0.3051
13	13873	47.6	0.2971
14	37777	195.6	0.3769
15	101829	−327.8	0.3862
16	277685	−6.0	0.1115
17	754093	−749.3	0.3894
18	2051549	−325.0	0.3213
19	5572497	−5074.9	0.4491
20	15156977	−4435.4	0.4199

TABLE VII. $\Phi(5)$

s	$N(s)$	Remainder	α
10	457	16.5	0.2802
11	1113	−84.5	0.4033
12	3329	73.9	0.3586
13	8845	−3.3	0.0911
14	24029	−23.1	0.2242
15	65621	240.7	0.3656
16	178313	590.8	0.3988
17	483945	845.9	0.3965
18	1313597	397.6	0.3325
19	3565629	−4017.0	0.4368
20	9709569	6265.1	0.4371

TABLE VIII. $\Phi(6)$

s	$N(s)$	Remainder	α
10	333.0	27.1	0.329868
11	805.0	−26.6	0.298214
12	2201.0	−59.5	0.340475
13	6125.0	−19.6	0.229006
14	16969.0	266.2	0.398865
15	45465.0	62.0	0.275122
16	122357.0	−1061.2	0.435447
17	334713.0	−772.5	0.391151
18	914421.0	2477.0	0.434155
19	2479637.0	716.2	0.345994
20	6737861.0	−544.5	0.314993

TABLE IX. $\Phi(7)$

s	$N(s)$	Remainder	α
10	265.0	40.2	0.369487
11	633.0	22.0	0.281166
12	1665.0	4.2	0.120318
13	4549.0	34.6	0.272554
14	12373.0	101.5	0.330024
15	33937.0	579.7	0.424165
16	91441.0	766.4	0.415107
17	245217.0	−1262.1	0.420032
18	669989.0	−10.7	0.131603
19	1822749.0	1501.0	0.384942
20	4958733.0	8067.7	0.449781

TABLE X. $\Phi(8)$

s	$N(s)$	Remainder	α
10	201.0	28.9	0.336447
11	489.0	21.2	0.277779
12	1421.0	149.5	0.417263
13	3577.0	120.6	0.368681
14	9413.0	17.7	0.205069
15	25521.0	−18.2	0.193422
16	70793.0	1370.3	0.451422
17	190161.0	1450.4	0.428213
18	513789.0	820.5	0.372772
19	1392593.0	−1800.0	0.394502
20	3790089.0	−264.1	0.278814

The α results were essentially the same in all cases: The exponent of growth of the remainder never exceeded 0.5 by much, indicating that the true upper bound for $d(s, z)$ was probably of the order $s^A e^{s/2}$. The bias in the reminder for $\Phi(8)$ is explained in Section 2 by the presence of a null vector at level 8. As for the mean square of the normalized remainder, Table II indicates that it has a limiting value, but we have not been able to prove this.

5.1. *The Program*

A copy of this program has also been included in this appendix. It is written in Pascal in order to take advantage of the stack properties of this language, which are especially suitable for tree structures. It will be recalled (see Section 1) that $\Gamma(2)$ is freely generated by

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \tag{5.1}$$

and that the Fermat group of level N consists of words of type

$$A^{p_1}B^{q_1}A^{p_2}B^{q_2} \cdots A^{p_k}B^{q_k}, \qquad \sum p_i \equiv 0 \equiv \sum q_j \bmod N \tag{5.2}$$

none of the p_i or q_j is zero excepts perhaps p_1 or q_k.

The lattice points of $\Gamma(2)$ (i.e., (5.2) acting on the base point z) are ordered according to word length and a record is kept of their coordinates, their distance from z, $\sum p_i \bmod N$, and $\sum q_j \bmod N$. Each additional letter to the word on the left is either a translation by $2p$, $p \neq 0$ (procedure `next1` $\sim A$) or such a translation conjugated by an inversion (procedure `next2` $\sim B$). Each translation takes a point in the strip $-1 < x < 1$ out of this strip and therefore increases the distance from z, which lies on the imaginary axis. B then inverts the point back into this strip, translates out of this strip, and ends up inverting back into the strip. Thus the distance from the base point is increased at each step in the program.

In procedure `next1` all the translates (power of A) except for the identity translate which stay within the circle C of maximum radius from this base point are found. All of these are stored except one, which is recorded and then used as the beginning point for `next2`. The procedure `next2` then inverts, obtains all the translates whose inversions lie within C, inverts again, and stores all but one of these points, which is then recorded and used as the beginning point for `next1`, and so on.

In this way the program proceeds down a branch of the tree until it reaches the point where the brach first extends outside of C. It then backs up one or more steps until it gets back in C and can proceed along another direction in a part of the tree which is still inside of C. This process continues until all of the tree in C has been exhausted.

The purpose of calling next1 followed by next2 in the main part of the program is first to treat all words beginning with A and then to treat all words beginning with B. If the base point is i then, since i is fixed under the inversion, it does not matter whether we start with A or B. We can then eliminate one of the calls in the main program and record each ensuing word as two; this halves the running time.

```
program Fermat (input, output);
       {lattice count for Fermat group F(n) of level n}
type dataset = record
      x, y, g: real;
      p, q: integer
      end;
var a: array [0..500] of real;
    root: dataset;
    current: dataset;
    d, c, j, n: integer;
    f, m, r, s, t: real;
procedure info;
       {determines input parameters
         and initializes variables}
    var index: integer;
    begin
      writeln('give the base point/i');
      readln(t);
      writeln('give fermat level, fraction of step
        and max dist.');
      readln(n, d, c);    {c=radius of circle C
        about base point = t*i}
      for index:= 6*d to c*d do a[index]:= 1;
        {the 1 adjusts for root}
      m:= 1; r:= 0; s:= 0;
    end;
procedure next2 (current: dataset); forward;
procedure next1 (current: dataset);
  {performs (1 2, 0 1)}
    var u, v, g: real;
      b, i, j: integer;
      temp: dataset;
    begin
      v:= current.y;
      b:= round(sqrt(v*t*exp(c))/2);
```

```
    {x limits in C at height y=v}
  for i:= -b to b do
    if i <> 0 then {i=0 is already included
      in previous procedure}
      begin
        u:= current.x + 2*i;
        g:= ln((sqr(u) + sqr(v) + sqr(t))/(v*t));
          {distance from (u, v) to t*i}
        if g < c then
          begin
            m:= m + 1;    {counts number of points in C}
            temp.x:= u;
            temp.y:= v;
            temp.g:= g;
            temp.p:= (current.p + i) mod n;
            temp.q:= current.q;
            if (temp.p = 0) and (temp.q = 0) then
                                  {selects points in F(n)}
              for j:= 6*d to c*d do    {counts points
                of dist.<j/d}
                if g*d < j then a[j]:= a[j] + 1;
            next2(temp);
          end;
      end;
end;

procedure next2;
      {performs (1 0, -2 0)
        = (0 -1, 1 0)(1 2, 0 1)(0 1, -1 0)}
  var e, o, u, v, x, y, w, g: real;
    b, i, j: integer;
    temp: dataset;

  begin
    x:= current.x;
    y:= current.y;
    e:= sqr(x) + sqr(y);    {inversion}
    u:= -x/e;
    v:= y/e;
    b:= round(sqrt(v*exp(c)/t)/2);
              {max. x dist. from t*i after inversion}
    for i:= -b to b do
      if i <> 0 then    {i=0 already included
        in previous procedure}
```

```
        begin
          w:= u + 2*i;
          o:= sqr(w) + sqr(v);
          x:= -w/o;
          y:= v/o;
          g:= ln((sqr(x) + sqr(y) + sqr(t))/(t*y));
          if g < c then
            begin
              m:= m + 1;    {counts number of points in C}
              temp.x:= x;
              temp.y:= y;
              temp.g:= g;
              temp.p:= current.p;
              temp.q:= (current.q + i) mod n;
              if (temp.p = 0) and (temp.q = 0) then
                                {selects points in F(n)}
                for j:= 6*d to c*d do
                  {counts points of dist.<j/d}
                  if g*d < j then a[j]:= a[j] + 1;
              nextl(temp);
            end;
        end;
end;

    begin    {beginning of program}
      info;
      root.x:= 0; root.y:= t; root.g:= 0;
      root.p:= 0; root.q:= 0;
      nextl(root);    {words beginning with (1 2, 0 1)}
      next2(root);    {words beginning with (1 0, -2 1)}
      write(chr(7), chr(7), chr(7));
      writeln('fermat level = ', n:2, ' c = ', c:3);
      writeln(' m = ', m:10:0, ' base point = ',
        t:4:3, '*i');
      for j:= 6*d to c*d do
        begin
          write(j/d:4:2, ' ', a[j]:10:0,' ');
          f:= a[j] - exp(j/d)/(2*sqr(n));
          r:= r + f/exp(j/(2*d));    {r = mean
            normalized remainder}
          s:= s + sqr(f/exp(j/(2*d)));
            {s = mean square normalized remainder}
          write((d*ln(abs(f))/j):5:4, ' ',
```

```
      r/((j - 6*d) + 1):5:4, ' ');
    write(s/((j - 6*d) + 1):5:4);
    writeln;
  end;
end.
```

5.2. *The Mean Square*

We used the data generated by the program to check the mean square normalized remainder term

$$\frac{1}{T}\int_0^T |e(s, z)|^2\, ds. \tag{5.3}$$

The mean square normalized remainder term, shown in Table II has values around 0.27 for $\Phi(1)$, 0.11 for $\Phi(2)$, 0.07 for $\Phi(3)$, and 0.04 for $\Phi(4)$ through $\Phi(8)$. As one would expect from this, the frequency distribution for the normalized remainder (not shown here) was noticeably flatter for $\Phi(1)$ then for the Fermat groups of higher levels. We were not able to explain this phenomenon, nor were we able to establish the existence of the limit

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T |e(s, z)|^2\, ds < \infty, \tag{5.4}$$

which is indicated in Table II.

Acknowledgment

We thank Peter Sarnak for his interest and helpful comments.

References

1. P. Cohen and P. Sarnak, Discontinuous groups and harmonic analysis, mimeographed notes.
2. H. Davenport, "Analytic Methods for Diophantine Equations and Diophantine Inequalities," Ann Arbor Science Pub., Ann Arbor, MI, 1962.
3. J. Delsarte, Sur le Gitter Fuchsien, *C.R. Acad. Sci. Paris* **214** (1942), 147–149.
4. P. Günther, Gitterpunktprobleme in symmetrischen Riemannschen Räumen vom Rang 1, *Math. Nachr.* **94** (1980), 5–27.
5. J. Hafner, New omega theorems for two classical lattice point problems, *Invent. Math.* **63** (1981), 181–186.
6. H. Huber, Über eine neue Klass automorpher Functionen und eine Gitterpunktproblem in der hyperbolische Ebene, *Comment. Math. Helv.* **30** (1956), 20–62.
7. M. N. Huxley, Exponential sums and lattice points, *Proc. London Math. Soc.* **60** (1990), 471–502.

8. A. Ivič, The Riemann Zeta Function," Wiley, New York, 1985.
9. H. Iwaniec and C. J. Mozzochi, On the divisor and circle problems, *J. Number Theory* **29** (1988), 60–93.
10. P. Lax and R. Phillips, The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces, *J. Funct. Anal.* **46** (1982), 280–350.
11. G. Margulis, Applications of ergodic theory to the investigation of manifolds of negative curvature, *Funct. Anal. Appl.* **4** (1969), 335.
12. S. J. Patterson, A lattice point problem in hyperbolic space, *Mathematika* **22** (1975), 81–88.
13. R. Phillips and P. Sarnak, The spectrum of Fermat curves, *Geom. Funct. Anal.* **1** (1991), 80–146.
14. R. Phillips and P. Sarnak, Perturbation theory for the Laplacian on automorphic functions, *J. Amer. Math. Soc.* **5** (1992), 1–32.
15. R. Phillips and P. Sarnak, Cusp forms for character varieties, *Geom. Funct. Anal.*, to appear.
16. A. Reznikov, Eisenstein matrix and existence of cusp forms in rank one symmetric spaces, *Geom. Funct. Anal.*, to appear.
17. A. Selberg, Bombay lectures, *in* "Collected Works I," Springer-Verlag, New York/Berlin, 1988.
18. A. Selberg, Gottingen lecture, *in* "Collected Works I," Springer-Verlag, New York/Berlin, 1988.
19. A. Selberg, Remarks on the distribution of poles of Eisenstein series, *in* "Collected Works II," Springer-Verlag, New York/Berlin, 1992.
20. S. Wolpert, Disappearance of cursp forms in famlies, preprint, 1992.