Russian Journal of Mathematical Physics, Vol. 12, No. 4, 2005, pp. 472–482.

Dedicated to the memory of B. M. Levitan

The Variance of the Hyperbolic Lattice Point Counting Function

R. Hill and L. Parnovski
Department of Mathematics, University College London
Gower Street, London, WC1E 6BT, England

Received August, 2005

Abstract. Let Γ be a discrete group of isometries of the d-dimensional hyperbolic space $\mathcal{H}$, such that $\Gamma\backslash\mathcal{H}$ has finite volume. For points $\mathbf{q}, \mathbf{p} \in \mathcal{H}$, we study the number $N_\rho(\mathbf{q}, \mathbf{p})$ of points of the lattice $\Gamma\mathbf{p}$ in a ball of radius ρ, centered at $\mathbf{q}$. This counting function depends only on the images of $\mathbf{q}$ and $\mathbf{p}$ in $\Gamma\backslash\mathcal{H}$. We shall regard $N_\rho(\mathbf{q}, \mathbf{p})$ as a function of $\mathbf{q} \in \Gamma\backslash\mathcal{H}$ and estimate its variance.

1. INTRODUCTION

The problem of estimating the number of points of a lattice that lie in a ball, is often called the *circle problem*. In the case of lattices in Euclidean space, this question goes back at least as far as Gauss. If we call N_ρ the number of points of $\mathbb{Z}^2$ inside the ball $B(0,\rho)$, then one easily sees that the leading term of N_ρ is the area, $\pi\rho^2$, of $B(0,\rho)$. It is not difficult to show that the error term in this estimate is bounded by circumference of $B(0,\rho)$ and is therefore $O(\rho)$. The first improvement was due to Sierpiński (1906) who used the Poisson summation formula to show that the error term is $O(\rho^{2/3})$. It is conjectured that this error term is in fact $O(\rho^{1/2+\epsilon})$. This conjecture has been extensively studied (see, for example, [18, 25, 5, 12, 13]) but is not proved. The best estimate of which we are aware is $N_\rho = \pi\rho^2 + O(\rho^{46/73})$ (see [12]).

More generally, given an arbitrary lattice L is $\mathbb{R}^d$ and an arbitrary ball $B(v,\rho)$ in $\mathbb{R}^d$, we let $N_\rho(v)$ denote the number of points of L in the ball. Again if we fix v and L, then for large ρ, we have

$$N_\rho(v) \sim \mathrm{vol}(B(0,\rho))/\mathrm{vol}(\mathbb{R}^d/L). \tag{1}$$

The easy bound on the error term here is $O(\rho^{d-1})$ and this has been improved by a number of authors (see [15, 3], and [7]).

It is obvious that the left-hand side of (1) depends on the center v of the ball, whereas the right-hand side depends only on its volume. The extent of this dependence is measured by the variance:

$$\mathrm{Var}(N_\rho) = (1/\mathrm{vol}(\mathbb{R}^d/L)) \int_{\mathbb{R}^d/L} \left(N_\rho(v) - \frac{\mathrm{vol}(B(0,\rho))}{\mathrm{vol}(\mathbb{R}^d/L)} \right)^2 dv.$$

In [20], this variance was estimated in connection with the Bethe–Sommerfeld Conjecture. In particular, it was shown that $\mathrm{Var}(N_\rho) \asymp \rho^{d-1}$, $d \not\equiv 1 \bmod 4$. However if d is congruent to 1 modulo 4, it is shown that $\mathrm{Var}(N_\rho)$ is not comparable to ρ^{d-1}, and instead one has for positive ϵ

$$\rho^{d-1-\epsilon} \ll \mathrm{Var}(N_\rho) \ll \rho^{d-1}, \quad d \equiv 1 \bmod 4. \tag{2}$$

These results are saying that, after averaging over v, the size of the error term in the circle problem is $O(\rho^{(d-1)/2})$. A similar result was obtained for the L^1-norm of $N_\rho - \mathrm{vol}(B(0,\rho))/\mathrm{vol}(\mathbb{R}^d/L)$. The first estimate in (2) was later improved in [14].

Now let Γ be a discrete group of isometries of the d-dimensional real hyperbolic space $\mathcal{H}$. For a point $\mathbf{p} \in \mathcal{H}$, we shall write $\Gamma\mathbf{p}$ for the Γ-orbit of $\mathbf{p}$. Again one may estimate the number of points of $\Gamma\mathbf{p}$ in a ball $B(\mathbf{q},\rho)$ in $\mathcal{H}$: $N_\rho(\mathbf{q},\mathbf{p}) = \#(\Gamma\mathbf{p} \cap B(\mathbf{q},\rho))$. One easily sees that $N_\rho(\mathbf{q},\mathbf{p})$ depends only on the images of $\mathbf{q}$ and $\mathbf{p}$ in $\Gamma\backslash\mathcal{H}$, and we also have $N_\rho(\mathbf{q},\mathbf{p}) = N_\rho(\mathbf{p},\mathbf{q})$.

For $d = 2$, the numbers $N_\rho(\mathbf{q},\mathbf{p})$ were estimated by Huber [10, 11] for the case of co-compact Γ. Huber's results were extended by Patterson [21] to the case in which $\Gamma\backslash\mathcal{H}$ has finite hyperbolic area, but may have cusps. Patterson's result was generalized to arbitrary dimensions by Lax and Phillips [16] and Levitan [17]. In fact, Lax and Phillips obtained results which continue to hold when

Γ is geometrically finite, but does not necessarily have finite co-volume. In this context, we remark that the circle problem has more recently been studied on other symmetric spaces, including those of higher rank (see [8, 2] and [4]).

To establish some notation, we describe the result of [16] for the case in which $\Gamma\backslash\mathcal{H}$ has finite volume. By an *eigenfunction* we shall always mean a nonzero, smooth, square-integrable function $\phi : \Gamma\backslash\mathcal{H} \to \mathbb{C}$ such that $-\Delta\phi = \lambda\phi$. Here Δ is the Laplace–Beltrami operator on $\mathcal{H}$ (see (12) below). There are at most countable many linearly independent eigenfunctions; we shall call them $\phi_0, \phi_1, \ldots$ We shall assume that these are chosen to be orthonormal, i.e.,

$$\int_{\Gamma\backslash\mathcal{H}} \phi_j(\mathbf{p})\bar{\phi}_k(\mathbf{p})d\mu(\mathbf{p}) = \delta_{jk},$$

where μ is the invariant measure on $\mathcal{H}$, see (11). The corresponding eigenvalues λ_j are all nonnegative real numbers, and we shall order them so that $0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots$ The constant function has eigenvalue 0; if there are infinitely many ϕ_j, then $\lambda_j \to \infty$. For convenience, we shall often use the notation $\mathcal{D} = \frac{d-1}{2}$, where d is the dimension of $\mathcal{H}$. We shall also choose complex numbers s_j such that $\lambda_j = s_j(d-1-s_j)$. If $\lambda_j \geqslant \mathcal{D}^2$, then s_j is on the line $\operatorname{Re} s = \mathcal{D}$; we shall refer to this line as the *critical line*. However if $\lambda_j < \mathcal{D}^2$, we have $s_i \in [0, 2\mathcal{D}]$, and we shall choose $s_j \in (\mathcal{D}, 2\mathcal{D}]$. We shall call λ an *exceptional eigenvalue* if $\lambda = s(d-1-s)$ with real $\mathcal{D} < s < 2\mathcal{D}$ (i.e., $0 < \lambda < \mathcal{D}^2$). There are only finitely many exceptional eigenvalues $0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_N < \mathcal{D}^2$.

With this notation, the estimate of [16] is as follows:

$$N_\rho(\mathbf{q},\mathbf{p}) = \operatorname{vol}(B(\mathbf{p},\rho))/\operatorname{vol}(\Gamma\backslash\mathcal{H}) + \sum_{j\,:\,d-1-\frac{d-1}{d+1}<s_j<d-1} w(s_j)\phi_j(\mathbf{q})\bar{\phi}_j(\mathbf{p})e^{s_j\rho} + O\left(\rho^{\frac{3}{d+1}}e^{\left(d-1-\frac{d-1}{d+1}\right)\rho}\right), \quad (3)$$

where the coefficients $w(s)$ are given by

$$w(s) = \pi^{\mathcal{D}}\Gamma(s-\mathcal{D})/\Gamma(s+1). \quad (4)$$

(We hope it will not cause too much confusion to use the symbol Γ to refer to both a discrete group and to the Euler gamma function.)

If one averages the numbers $N_\rho(\mathbf{q},\mathbf{p})$ in some way, then one can hope to obtain a better bound on the error term in (3). For example, Phillips and Rudnick [22] have studied the average of the numbers $N_\rho(\mathbf{q},\mathbf{p})$ as ρ ranges over an interval. In another direction, for co-compact Γ in the 2-dimensional case, Wolfe estimated the variance of $N_\rho(\mathbf{q},\mathbf{p})$ as a function of $\mathbf{q}$ and $\mathbf{p}$:

$$\int_{\Gamma\backslash\mathcal{H}}\int_{\Gamma\backslash\mathcal{H}} \left(N_\rho(\mathbf{q},\mathbf{p}) - \frac{\operatorname{vol}(B(\mathbf{p},\rho))}{\operatorname{vol}(\Gamma\backslash\mathcal{H})}\right)^2 d\mu(\mathbf{q})d\mu(\mathbf{p}).$$

A similar result is outlined in the 3-dimensional case in [23]. As Wolfe points out, this question only makes sense for co-compact Γ, since otherwise N_ρ fails to be square-integrable on $\Gamma\backslash\mathcal{H}\times\Gamma\backslash\mathcal{H}$. One can however fix $\mathbf{p}$ and study the variance of $N_\rho(\mathbf{q},\mathbf{p})$ as a function of $\mathbf{q}$; this is the subject of the present paper. Thus we shall study the quantity

$$\operatorname{Var}(N_\rho(\cdot,\mathbf{p})) = \frac{1}{\operatorname{vol}(\Gamma\backslash\mathcal{H})}\int_{\Gamma\backslash\mathcal{H}} \left(N_\rho(\mathbf{q},\mathbf{p}) - \frac{\operatorname{vol}(B(\mathbf{p},\rho))}{\operatorname{vol}(\Gamma\backslash\mathcal{H})}\right)^2 d\mu(\mathbf{q}).$$

This number describes to what extent the number of lattice points in a ball depends on the center of the ball.

1.1. *Statements*

To make matters simpler, we assume throughout that Γ is torsion-free. We first describe our result for the case in which Γ is co-compact. Again we shall write $\lambda_1, \ldots, \lambda_N$ for the exceptional eigenvalues; $\phi_1, \ldots \phi_N$ will denote the corresponding normalized eigenfunctions, and $\lambda_j = s_j(d-1-s_j)$ with $\mathcal{D} < s_j < d-1$. For an exceptional eigenvalue $\lambda = s(d-1-s)$, we define a polynomial

$$f_s(x) = w(s)(1-x)^d \sum_{0\leqslant n<s-\mathcal{D}} \frac{(d-s)_n(\mathcal{D}+1)_n}{(\mathcal{D}-s+1)_n}\frac{x^n}{n!}. \quad (5)$$

Here, $w(s)$ is the same constant (4) as in the Lax-Phillips theorem. We are using the notation $(x)_n = x(x+1)\cdots(x+n-1)$. We also define a constant $c_{\log}(s)$ by

$$c_{\log}(s) = \begin{cases} \frac{2(-1)^{s-\mathcal{D}}\pi^{d-1}}{(s-\mathcal{D})\Gamma(s+1)\Gamma(d-s)} & \text{if } \ s-\mathcal{D}\in\mathbb{Z}, \\ 0 & \text{otherwise.}\end{cases} \quad (6)$$

Finally, we define a constant h_d, depending only on the dimension d, by

$$h_d = \begin{cases} -\sum_{k=1}^{(d-1)/2} 1/k & \text{if } d \text{ is odd,} \\ \log 4 - \sum_{k=0}^{(d-2)/2} 2/(2k+1) & \text{if } d \text{ is even.} \end{cases} \tag{7}$$

Theorem 1. *If Γ is co-compact, then for large ρ, we have*

$$\operatorname{vol}(\Gamma\backslash\mathcal{H}) \operatorname{Var}(N_\rho(\cdot,\mathbf{p})) = \sum_{j:0<\lambda_j<\mathcal{D}^2} f_{s_j}\left(e^{-2\rho}\right)^2 |\phi_j(\mathbf{p})|^2 e^{2s_j\rho} + \sum_{j:0<\lambda_j<\mathcal{D}^2} c_{\log}(s_j)|\phi_j(\mathbf{p})|^2 \rho e^{(d-1)\rho}$$
$$+ \frac{4\pi^{d-1}}{\Gamma(\mathcal{D}+1)^2} \sum_{j:\lambda_j=\mathcal{D}^2} |\phi_j(\mathbf{p})|^2 (h_d+\rho)^2 e^{(d-1)\rho} + O(e^{(d-1)\rho}).$$

Here f_s, $c_{\log}$ and h_d are as defined in (5), (6), and (7). *The implied constant in the error term depends on Γ.*

One could obtain a result with a much weaker error term by integrating the Lax-Phillips estimate (3) with respect to $\mathbf{q}$. Such a result would however only involve a subset of the exceptional eigenvalues; it would only involve the constant term of each polynomial f_s, and it would not contain the $c_{\log}$ terms. As far as we are aware, these extra terms have not been observed before.

The reader wishing for slightly less information might be happier with the constant term of f_i: $f_s(e^{-2\rho}) = w(s)+O(e^{-2\rho})$, where $w(s)$ is as in the Lax–Phillips theorem (4). Indeed, in dimensions 2 and 3, the other terms of the polynomial f_s may be absorbed into the error term, and we have the following corollary.

Corollary 1. *Let Γ be as in Theorem* 1 *and assume in addition that $d = 2$ or $d = 3$. Then*

$$\operatorname{vol}(\Gamma\backslash\mathcal{H}) \operatorname{Var}(N_\rho(\cdot,\mathbf{p})) = \sum_{j\ :\ 0<\lambda_j<\mathcal{D}^2} w(s_j)^2 |\phi_j(\mathbf{p})|^2 e^{2s_j\rho}$$
$$+ \frac{4\pi^{d-1}}{\Gamma(\mathcal{D}+1)^2} \sum_{j\ :\ \lambda_j=\mathcal{D}^2} |\phi_j(\mathbf{p})|^2 (h_d+\rho)^2 e^{(d-1)\rho} + O(e^{(d-1)\rho}). \tag{8}$$

The implied constant in the error term depends on Γ.

Remark 1. Integrating (8) with respect to $\mathbf{p}$, we improve Wolfe's estimate from [26].

Now suppose that $\Gamma\backslash\mathcal{H}$ has finite volume, but has a finite number of cusps $\kappa_1,\ldots,\kappa_M$. For each cusp κ, we let $E_\kappa(\mathbf{q},s)$ denote the corresponding Eisenstein series; to make matters more precise, we now describe the normalization of $E_\kappa(\mathbf{q},s)$. We represent $\mathcal{H}$ in the usual way as the upper half-space: $\mathcal{H} = \{(\mathbf{x},y) : \mathbf{x}\in\mathbb{R}^{d-1},\ y>0\}$. For $\mathbf{p}\in\mathcal{H}$, we shall write $y(\mathbf{p})$ for the y-coordinate of $\mathbf{p}$. We let N denote the group of translations of $\mathcal{H}$ in the $\mathbf{x}$-coordinate. We identify N with $\mathbb{R}^{d-1}$ and equip it with the Lebesgue measure from $\mathbb{R}^{d-1}$. Each cusp κ_i is a point on the boundary of $\mathcal{H}$, i.e., $\mathbb{R}^{d-1}\cup\{\infty\}$. We shall write Γ_κ for the stabilizer of κ in Γ. We choose an isometry g_κ of $\mathcal{H}$ so that

(1) $g_\kappa(\kappa) = \infty$;
(2) $\operatorname{vol}(N/(g_\kappa\Gamma_\kappa g_\kappa^{-1})) = 1$.

(Since we are assuming that Γ is torsion-free, it follows that $g_\kappa\Gamma_\kappa g_\kappa^{-1}$ is a lattice in N.) Then for $\operatorname{Re} s > d-1$, we define the *Eisenstein series* by

$$E_\kappa(\mathbf{q},s) = \sum_{\gamma\in\Gamma_\kappa\backslash\Gamma} y(g_\kappa\gamma\mathbf{q})^s. \tag{9}$$

Each $E_\kappa(\mathbf{q},s)$ converges absolutely and uniformly on compact subsets of the half plane $\operatorname{Re} s > d-1$. It is therefore analytic in s in this region. Furthermore, E_κ has a meromorphic continuation to all $s\in\mathbb{C}$. Each function $E_\kappa(\cdot,s)$ is automorphic with respect to Γ, so we may regard it as a function on $\Gamma\backslash\mathcal{H}$. For convenience, we shall write $\mathcal{E}(\mathbf{q},s)$ for the column vector of Eisenstein series: $\mathcal{E}_j(\mathbf{q},s) = E_{\kappa_j}(\mathbf{q},s)$, $j=1,\ldots,M$. This vector of Eisenstein series satisfies a functional equation of the form

$$\mathcal{E}(\mathbf{q},d-1-s) = \Phi(s)\mathcal{E}(\mathbf{q},s), \tag{10}$$

where $\Phi(s)$ is an $M\times M$ matrix of meromorphic functions, called the *scattering matrix*. The scattering matrix $\Phi(s)$ is unitary for s on the critical line $\operatorname{Re} s = \mathcal{D}$. Thus $s = \mathcal{D}$ is the center of symmetry of the functional equation.

Theorem 2. *If* Γ *is co-finite, then for large* ρ *we have*

$$\operatorname{vol}(\Gamma\backslash\mathcal{H})\operatorname{Var}(N_\rho(\cdot,\mathbf{p})) = \sum_{j:0<\lambda_j<\mathcal{D}^2} f_{s_j}\left(e^{-2\rho}\right)^2 |\phi_j(\mathbf{p})|^2 e^{2s_j\rho} + \Big(\sum_{j:0<\lambda_j<\mathcal{D}^2} c_{\log}(s_j)|\phi_j(\mathbf{p})|^2 + \frac{\pi^{d-1}}{\Gamma(\mathcal{D}+1)^2}\left|\mathcal{E}(\mathbf{p},\mathcal{D})\right|^2\Big)\rho e^{(d-1)\rho} + \frac{4\pi^{d-1}}{\Gamma(\mathcal{D}+1)^2}\sum_{j:\lambda_j=\mathcal{D}^2}|\phi_j(\mathbf{p})|^2(h_d+\rho)^2 e^{(d-1)\rho} + O(e^{(d-1)\rho}).$$

Here, $|\mathcal{E}(\mathbf{p},\mathcal{D})|$ *is the usual Euclidean norm of the vector of Eisenstein series. The implied constant in the error term depends on both* Γ *and* $\mathbf{p}$.

Suppose that ϕ is one of the exceptional eigenfunctions. After multiplying ϕ by a constant of absolute value 1 if necessary, we may assume that ϕ is real valued. Since ϕ is orthogonal to the constant function, we know that its integral is zero. Hence ϕ takes both positive and negative values, so it follows that ϕ vanishes on a subset of $\mathcal{H}$ of codimension 1. Similarly, one can show that $E_\kappa(\mathcal{D},\mathbf{p})$ is zero on such a subset. Thus there are points $\mathbf{p}\in\mathcal{H}$ for which some of the terms of the estimates of Theorems 1 and 2 vanish. For these points $\mathbf{p}$, the variance of $N_\rho(\cdot,\mathbf{p})$ is significantly smaller than for other points, so the corresponding lattices $\Gamma\mathbf{p}$ are more evenly distributed in balls.

2. BACKGROUND MATERIAL

2.1. *The Plancherel Formula for* Γ

We recall that the standard metric on the upper half space model of $\mathcal{H}$ is given by

$$y^{-2}\left(dx_1^2+\cdots+dx_{d-1}^2+dy^2\right).$$

The measure invariant under isometries is

$$d\mu(\mathbf{x},y) = y^{-d}dx_1\ldots dx_{d-1}dy, \tag{11}$$

and the positive Laplace–Beltrami operator is

$$-\Delta = -y^2\Big(\frac{\partial^2}{\partial x_1^2}+\cdots+\frac{\partial^2}{\partial x_{d-1}^2}+\frac{\partial^2}{\partial y^2}\Big)+(d-2)y\frac{\partial}{\partial y}. \tag{12}$$

Our method for calculating the variance of $N_\rho(\cdot,\mathbf{p})$ is to find its Fourier transform and then apply the Plancherel formula, which we now recall.

The space $L^2(\Gamma\backslash\mathcal{H})$ is a Hilbert space with the following inner product:

$$\langle f,g\rangle = \int_{\Gamma\backslash\mathcal{H}} f(\mathbf{p})\bar{g}(\mathbf{p})d\mu(\mathbf{p}).$$

Suppose first that $\Gamma\backslash\mathcal{H}$ is compact. In this case, the orthonormal eigenfunctions $\phi_0,\phi_1,\phi_2,\ldots$ span a dense subspace of $L^2(\Gamma\backslash\mathcal{H})$. The discrete Fourier transform of a function $f\in L^2(\Gamma\backslash\mathcal{H})$ is defined by $\hat{f}_j^{(\mathrm{disc})} := \langle f,\phi_j\rangle$. Thus we have (with convergence in $L^2(\Gamma\backslash\mathcal{H})$)

$$f = \sum_{j=0}^{\infty}\hat{f}_j^{(\mathrm{disc})}\cdot\phi_j.$$

From this, we deduce the Plancherel formula in this case,

$$\|f\|^2 = \sum_{j=0}^{\infty}\left|\hat{f}_j^{(\mathrm{disc})}\right|^2.$$

In particular, the variance of f is simply this sum without the coefficient of ϕ_0,

$$\operatorname{Var}(f) = \frac{1}{\operatorname{vol}(\Gamma\backslash\mathcal{H})}\sum_{j=1}^{\infty}\left|\hat{f}_j^{(\mathrm{disc})}\right|^2. \tag{13}$$

Now suppose that Γ has cusps $\kappa_1,\ldots,\kappa_M$. In this case, the closed span of the eigenfunctions ϕ_j is a proper subspace of $L^2(\Gamma\backslash\mathcal{H})$. The orthogonal complement of this subspace consists of integrals of Eisenstein series over the critical line $\operatorname{Re}s=\mathcal{D}$. More precisely, recall that, for each cusp κ, we have an Eisenstein series $E_\kappa(\mathbf{p},s)$. For each s which is not one of the poles of E_κ, the function $E_\kappa(\cdot,s)$ is the solution of the eigenfunction equation for $-\Delta$,

$$-\Delta E_\kappa(\cdot,s) = s(d-1-s)E_\kappa(\cdot,s). \tag{14}$$

Since $E_\kappa(\mathbf{p},\bar{s}) = \overline{E_\kappa(\mathbf{p},s)}$, it follows that $E_\kappa(\cdot,s)$ is real valued for real s. Furthermore, by the Fourier expansion of E_κ around κ, it follows that $E_\kappa(\mathbf{p},\mathcal{D})$ is bounded below and tends to infinity as $\mathbf{p}$ moves into the κ. For fixed s on the critical line (i.e., $\operatorname{Re} s = \mathcal{D}$), the function $E_\kappa(\mathbf{p},s)$ is not square-integrable; it is in fact in $L^p(\Gamma\backslash\mathcal{H})$ if and only if $p < 2$. However, for $a \in C_0^\infty(\mathbb{R})$, the function

$$\hat{a}_\kappa(\mathbf{p}) = \int_{\mathbb{R}} a(t) E_\kappa\left(\mathcal{D}+it,\mathbf{p}\right) dt$$

is square-integrable on $\Gamma\backslash\mathcal{H}$. The orthogonal complement of the eigenfunctions is the closed span of the functions $\hat{a}_\kappa$ for the various cusps κ and $a \in C_0^\infty(\mathbb{R})$.

We therefore define the continuous part of the Fourier transform of f by

$$\hat{f}_\kappa^{(\mathrm{cts})}(s) = \int_{\Gamma\backslash\mathcal{H}} f(\mathbf{p})\bar{E}_\kappa(\mathbf{p},s)d\mu(\mathbf{p}), \quad \operatorname{Re} s = \mathcal{D}.$$

In this case, the Plancherel formula states that

$$\|f\|^2 = \sum_{j=0}^{\infty} \left|\hat{f}_j^{(\mathrm{disc})}\right|^2 + \frac{1}{4\pi}\sum_\kappa \int_{\mathbb{R}} \left|\hat{f}_\kappa^{(\mathrm{cts})}\left(\mathcal{D}+it\right)\right|^2 dt.$$

The constant $1/4\pi$ depends on the normalization of the Eisenstein series chosen in the introduction. Again we obtain the variance by simply missing out the constant term:

$$\operatorname{Var}(f) = \frac{1}{\operatorname{vol}(\Gamma\backslash\mathcal{H})}\left(\sum_{j=1}^{\infty} \left|\hat{f}_j^{(\mathrm{disc})}\right|^2 + \frac{1}{4\pi}\sum_\kappa \int_{\mathbb{R}} \left|\hat{f}_\kappa^{(\mathrm{cts})}\left(\mathcal{D}+it\right)\right|^2 dt\right). \tag{15}$$

2.2. *Hypergeometric Functions*

We collect here the facts which we shall need about the hypergeometric function ${}_2F_1$. First, for $|x| < 1$ and for $c \notin \mathbb{Z}_{\leqslant 0}$, the function is defined by

$${}_2F_1(a,b;c;x) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n}\frac{x^n}{n!}.$$

Here we are using the standard notation $(a)_n := \Gamma(a+n)/\Gamma(a) = a(a+1)\cdots(a+n-1)$, $(a)_0 = 1$. For the case in which $c \in \mathbb{Z}_{\leqslant 0}$, we define

$${}_2\mathbf{F}_1(a,b;c;x) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{\Gamma(n+c)}\frac{x^n}{n!}.$$

Here we are using the convention $1/\Gamma(n+c) = 0$ for $n+c = -1,-2,-3,\ldots$ Note that generically we have ${}_2F_1(a,b;c;z) = \Gamma(c){}_2\mathbf{F}_1(a,b;c;z)$. If $\operatorname{Re} c > \operatorname{Re} b > 0$, then we have an integral representation (Theorem 2.2.1 of [1])

$${}_2F_1(a,b;c;x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-xt)^{-a}dt. \tag{16}$$

Using the analytic continuation of $(1-x)^{-a}$ to the cut plane $\mathbb{C}\setminus[1,\infty)$, the identity (16) gives an analytic continuation of ${}_2\mathbf{F}_1$ to the same cut plane, as long as $\operatorname{Re} c > \operatorname{Re} b > 0$, and in fact ${}_2\mathbf{F}_1$ has an analytic continuation to this cut plane for all $a,b,c \in \mathbb{C}$. When we refer to ${}_2F_1(a,b;c;x)$ with $|x| \geqslant 1$, we shall always mean this continuation.

From (16), one easily obtains Pfaff's identity (Theorem 2.2.5 of [1]):

$${}_2\mathbf{F}_1(a,b;c;x) = (1-x)^{-a}{}_2\mathbf{F}_1\big(a,c-b;c;x/(x-1)\big). \tag{17}$$

The hypergeometric function ${}_2\mathbf{F}_1(a,b;c;\cdot)$ is a solution to the following differential equation (see section 2.3 of [1]):

$$(1-x)d^2f/dx^2 + (c-(a+b+1)x)df/dx - abf = 0. \tag{18}$$

This equation has two linearly independent solutions on the interval $(0,1)$; the form of the second solution depends on which of the numbers $c, a-b, c-a-b$ are integers. If c is not an integer, then a second solution is given by $x^{1-c}{}_2F_1(a+1-c,b+1-c;2-c;x)$. Now consider the function ${}_2F_1(a,b;a+b+1-c;1-x)$ for $x \in (0,1)$. It follows that this is also a solution to (18), so we may write ${}_2F_1(a,b;a+b+1-c;1-x)$ as a linear combination of the two linearly independent solutions. If c is not an integer, then this expansion is:

$${}_2F_1(a,b;a+b+1-c;1-x) = A\,{}_2F_1(a,b,c;x) + Bx^{1-c}\,{}_2F_1(a+1-c,b+1-c;2-c;x), \tag{19}$$

where
$A = \Gamma(1-c)\Gamma(a+b+1-c)/\Gamma(a+1-c)\Gamma(b+1-c), \quad B = \Gamma(a+b-c)\Gamma(a+b+1-c)/\Gamma(a)\Gamma(b).$
Formula (19) gives us a good understanding of ${}_2F_1(a,b;a+b+1-c;1-x)$ when x is small and $c \notin \mathbb{Z}$. We shall need a formula of this form valid for $c \in \mathbb{Z}$: if $c = 1-m$ with $m = 1, 2, \dots$, then

$$ {}_2F_1(a,b;a+b+m;1-x) = \frac{\Gamma(m)\Gamma(a+b+m)}{\Gamma(a+m)\Gamma(b+m)} \sum_{n=0}^{m-1} \frac{(a)_n(b)_n}{(1-m)_n} \frac{x^n}{n!} + \frac{\Gamma(a+b+m)}{\Gamma(a)\Gamma(b)}(-x)^m \sum_{n=0}^{\infty} \frac{(a+m)_n(b+m)_n}{(n+m)!} (\log(x)+r_n) \frac{x^n}{n!}, \quad (20) $$

where $r_n = \psi(a+m+n) + \psi(b+m+n) - \psi(1+m+n) - \psi(n+1)$, $\psi(x) = \Gamma'(x)/\Gamma(x)$. This may be obtained by taking the limit of (19) as $c \to 1-m$; it can be found in [19, Sec. 2.4.1].

3. PROOFS OF THE MAIN RESULTS

Theorems 1 and 2 are proved by calculating the Fourier transform of the counting function $N_\rho(\cdot, \mathbf{p})$ and applying the Plancherel formula (13) or (15). We shall make brief use of the following function on $\mathcal{H} \times \mathcal{H}$: $\chi_\rho(\mathbf{p}, \mathbf{q}) = 1$ if $\operatorname{dist}(\mathbf{p}, \mathbf{q}) < \rho$, $\chi_\rho(\mathbf{p}, \mathbf{q}) = 0$ if $\operatorname{dist}(\mathbf{p}, \mathbf{q}) \geqslant \rho$. Obviously, we have $\chi_\rho(g\mathbf{p}, g\mathbf{q}) = \chi_\rho(\mathbf{p}, \mathbf{q})$ for any isometry g of $\mathcal{H}$, so χ_ρ is a point-pair invariant in the sense of [24] (the fact that χ_ρ is not continuous will play no role here). To calculate the Fourier transform $\hat{N}_\rho$, we need to integrate N_ρ against eigenfunctions and Eisenstein series. As a first step in this, we note the following lemma.

Lemma 1. *Let $\rho > 0$. For any automorphic function g on $\mathcal{H}$, we have*

$$ \int_{\Gamma \backslash \mathcal{H}} N_\rho(\mathbf{q}, \mathbf{p}) g(\mathbf{q}) d\mu(\mathbf{q}) = \int_{\mathcal{H}} \chi_\rho(\mathbf{p}, \mathbf{q}) g(\mathbf{q}) d\mu(\mathbf{q}). $$

Proof. As Γ is torsion-free, we may express N_ρ as $N_\rho(\mathbf{q}, \mathbf{p}) = \sum_{\gamma \in \Gamma} \chi_\rho(\gamma \mathbf{q}, \mathbf{p})$. We may unfold the integral as follows:

$$ \int_{\Gamma \backslash \mathcal{H}} N_\rho(\mathbf{q}, \mathbf{p}) g(\mathbf{q}) d\mu(\mathbf{q}) = \int_{\Gamma \backslash \mathcal{H}} \sum_\gamma \chi_\rho(\gamma\mathbf{q}, \mathbf{p}) g(\mathbf{q}) d\mu(\mathbf{q}) = \int_{\mathcal{H}} \chi_\rho(\mathbf{q}, \mathbf{p}) g(\mathbf{q}) d\mu(\mathbf{q}). \quad \square $$

Thus to calculate $\widehat{N_\rho(\cdot, \mathbf{p})}$, we will need to integrate eigenfunctions against the point-pair invariant χ_ρ. To do this, we recall the following proposition.

Proposition 1. *Fix $\rho > 0$ and $s \in \mathbb{C}$. There is a number $I(\rho, s)$ with the following property: for any smooth function $\phi\colon \mathcal{H} \to \mathbb{C}$ satisfying*

$$ -\Delta\phi = s(d-1-s)\phi, \quad (21) $$

we always have

$$ \int_{\mathcal{H}} \chi_\rho(\mathbf{p}, \mathbf{q}) \phi(\mathbf{q}) d\mu(\mathbf{q}) = I(\rho, s)\phi(\mathbf{p}). \quad (22) $$

The function I does not depend on ϕ or $\mathbf{p}$.

Proof. This is true for any point-pair invariant (see, for example, Theorem 1 of [23]). $\square$

Putting the last two results together, we have

$$ \widehat{N_\rho(\cdot, \mathbf{p})}_j^{(\mathrm{disc})} = I(\rho, s_j)\bar{\phi}_j(\mathbf{p}), \qquad \widehat{N_\rho(\cdot, \mathbf{p})}^{(\mathrm{cts})}(s) = I(\rho, s)\bar{E}(s, \mathbf{p}). \quad (23) $$

It is therefore important for us to understand $I(\rho, s)$. Before going on, we note the following obvious properties of I:

(1) for all $\rho > 0$ and all $s \in \mathbb{C}$,

$$ I(\rho, s) = \int_{B(\mathcal{O}, \rho)} y(\mathbf{q})^s d\mu(\mathbf{q}), \quad (24) $$

where $\mathcal{O}$ is the point $(0, \dots, 0, 1)$ of the upper half-space $\mathcal{H}$;

(2) for fixed $\rho > 0$, the function $I(\rho, \cdot)$ is entire, and satisfies the functional equation $I(\rho, s) = I(\rho, d-1-s)$. In particular, $I(\rho, s)$ is real for s real or s on the critical line $\operatorname{Re} s = \mathcal{D}$;

(3) For fixed $\rho > 0$, the function $I(\rho, s)$ is bounded in vertical strips.

The identity (24) follows by substituting $\phi(\mathbf{q}) = y(\mathbf{q})^s$ and $\mathbf{p} = \mathcal{O}$ into (22). The functional equation is immediate from the defining property (22) of I. The analytic properties of I follow from (24).

We need to investigate the behavior of $I(\rho, s)$ as $\rho \to \infty$. We begin with the following result.

Proposition 2. *For all $\rho > 0$ and all $s \in \mathbb{C}$, we have*

$$I(\rho, s) = \frac{\pi^{\mathcal{D}}\Gamma(\mathcal{D}+1)}{\Gamma(d+1)}(2\sinh(\rho))^d e^{(s-d)\rho} {}_2F_1\left(d-s, \mathcal{D}+1; d+1; 1-e^{-2\rho}\right).$$

Proof. After substituting the definition (11) of the invariant measure μ into the expression (24), we have

$$I(\rho, s) = \int_{B(\mathcal{O},\rho)} y^{s-d} dx_1 \dots dx_n dy.$$

To calculate the integral, we shall express $B(\mathcal{O}, \rho)$ as a Euclidean ball. The lowest point of this ball has y-coordinate $e^{-\rho}$ and the highest point e^{ρ}. Thus the Euclidean radius is $\sinh(\rho)$ and the Euclidean center is $(0, \cosh(\rho))$. In the integral, the y coordinate runs from $e^{-\rho}$ to e^{ρ}. For a fixed value of y, the $\mathbf{x}$-coordinate runs over a $(d-1)$-dimensional Euclidean ball of radius $\sqrt{(e^\rho - y)(y - e^{-\rho})}$. We therefore have

$$I(\rho, s) = \int_{e^{-\rho}}^{e^{\rho}} \int_{B\left(0, \sqrt{(e^\rho - y)(y-e^{-\rho})}\right)} y^{s-d} d\mathbf{x} dy.$$

Since the Euclidean ball $B(0, r)$ in $\mathbb{R}^{d-1}$ has volume $\frac{\pi^{\mathcal{D}}}{\Gamma(\mathcal{D}+1)} r^{d-1}$, we have

$$I(\rho, s) = \frac{\pi^{\mathcal{D}}}{\Gamma(\mathcal{D}+1)} \int_{e^{-\rho}}^{e^{\rho}} (e^\rho - y)^{\mathcal{D}} (y - e^{-\rho})^{\mathcal{D}} y^{s-d} dy.$$

Now, making the change of variable $2\sinh(\rho)\eta = y - e^{-\rho}$, we have

$$I(\rho, s) = \frac{\pi^{\mathcal{D}}}{\Gamma(\mathcal{D}+1)} (2\sinh(\rho))^d e^{(d-s)\rho} \int_0^1 (1-\eta)^{\mathcal{D}} \eta^{\mathcal{D}} (1 - (1-e^{2\rho})\eta)^{s-d} d\eta.$$

By Euler's integral formula (16), this is equal to

$$I(\rho, s) = \frac{\pi^{\mathcal{D}}\Gamma(\mathcal{D}+1)}{\Gamma(d+1)} (2\sinh(\rho))^d e^{(d-s)\rho} {}_2F_1(d-s, \mathcal{D}+1; d+1; 1-e^{2\rho}).$$

Finally, by Pfaff's identity (17), we have

$$I(\rho, s) = \frac{\pi^{\mathcal{D}}\Gamma(\mathcal{D}+1)}{\Gamma(d+1)} (2\sinh(\rho))^d e^{(s-d)\rho} {}_2F_1(d-s, \mathcal{D}+1; d+1; 1-e^{-2\rho}). \quad \square$$

In passing, we note the following corollary.

Corollary 2. *The volume of a ball in $\mathcal{H}$ of radius ρ is given by:*

$$\mathrm{vol}(B(\mathbf{p}, \rho)) = (\pi^{\mathcal{D}}\Gamma(\mathcal{D}+1)/\Gamma(d+1))(2\sinh(\rho))^d e^{-\rho} {}_2F_1\left(1, \mathcal{D}+1; d+1; 1-e^{-2\rho}\right).$$

Proof. We simply note that $\mathrm{vol}(B(\mathbf{p}, \rho)) = I(\rho, d-1)$. $\square$

3.1. *Leading Terms from the Discrete Spectrum*

Recall that, for a real $s > \mathcal{D}$, we defined: $w(s) = \pi^{\mathcal{D}}\Gamma(s-\mathcal{D})/\Gamma(s+1)$,

$$f_s(x) = w(s)(1-x)^d \sum_{0 \leqslant n < s-\mathcal{D}} \frac{(d-s)_n(\mathcal{D}+1)_n}{(\mathcal{D}-s+1)_n} \frac{x^n}{n!},$$

$$h_d = \begin{cases} -\sum_{k=1}^{(d-1)/2} 1/k & \text{if } d \text{ is odd}, \\ \log(4) - \sum_{k=0}^{(d-2)/2} 2/(2k+1) & \text{if } d \text{ is even}. \end{cases}$$

Proposition 3. *Fix a real* $s \in [\mathcal{D}, d-1)$.

(i) *If* $s-\mathcal{D} \notin \mathbb{Z}$, *then* $I(\rho, s) = f_s(e^{-2\rho})e^{s\rho} + O(e^{(d-1-s)\rho})$.

(ii) *If* $s-\mathcal{D}$ *is a positive integer, then*

$$I(\rho, s) = f_s(e^{-2\rho})e^{s\rho} + \frac{2(-1)^{s-\mathcal{D}}\pi^{\mathcal{D}}}{(s-\mathcal{D})!\Gamma(d-s)} e^{(d-1-s)\rho}\rho + O(e^{(d-1-s)\rho}).$$

(iii) *In the case* $s = \mathcal{D}$, *we have* $I(\rho, \mathcal{D}) = (2\pi^{\mathcal{D}}/\Gamma(\mathcal{D}+1))\,(h_d + \rho)\, e^{\mathcal{D}\rho} + O(\rho e^{\rho(\mathcal{D}-2)})$.

The implied constant in case (i) *depends on* s.

Proof. In case (i), we use the expansion (19) to show that for $x \in (0,1)$:

$${}_2F_1(d-s, \mathcal{D}+1; d+1; 1-x) = A\, {}_2F_1(d-s, \mathcal{D}+1; \mathcal{D}+1-s; x) + Bx^{s-\mathcal{D}}\, {}_2F_1(d-\mathcal{D}, s+1; 1; x),$$

where $A = \Gamma(d+1)\Gamma(s-\mathcal{D})/\Gamma(\mathcal{D}+1)\Gamma(s+1)$. Bounding the second term above by $x^{s-\mathcal{D}}$, we have

$${}_2F_1(d-s, \mathcal{D}+1; d+1; 1-x) = \frac{\Gamma(d+1)\Gamma(s-\mathcal{D})}{\Gamma(\mathcal{D}+1)\Gamma(s+1)}\, {}_2F_1(d-s, \mathcal{D}+1; \mathcal{D}+1-s; x) + O(x^{s-\mathcal{D}}).$$

Truncating the hypergeometric series at the error term, we obtain

$${}_2F_1(d-s, \mathcal{D}+1; d+1; 1-x) = \frac{\Gamma(d+1)\Gamma(s-\mathcal{D})}{\Gamma(\mathcal{D}+1)\Gamma(s+1)} \sum_{0 \leqslant n < s-\mathcal{D}} \frac{(d-s)_n(\mathcal{D}+1)_n}{(\mathcal{D}+1-s)_n}\frac{x^n}{n!} + O(x^{s-\mathcal{D}}).$$

Substituting this into Proposition 2, we can write

$$I(\rho, s) = w(s)(2\sinh(\rho))^d e^{(s-d)\rho} \sum_{0 \leqslant n < s-\mathcal{D}} \frac{(d-s)_n(\mathcal{D}+1)_n}{(\mathcal{D}+1-s)_n}\frac{e^{-2n\rho}}{n!} + O(e^{(d-1-s)\rho}).$$

Replacing $2\sinh(\rho)$ by $e^{\rho}(1-e^{-2\rho})$, we obtain $I(\rho, s) = f_s(e^{-2\rho})e^{s\rho} + O(e^{(d-1-s)\rho})$. This gives (i). For (ii), we use the expansion (20) instead of (19). To prove (iii), we first use (20) as well, holding on to one extra term. This gives $I(\rho, \mathcal{D}) = (2\pi^{\mathcal{D}}/\Gamma(\mathcal{D}+1))e^{\mathcal{D}\rho}\,(\psi(1) - \psi(\mathcal{D}+1) + \rho) + O(\rho e^{\rho(\mathcal{D}-2)})$, where $\psi = \Gamma'/\Gamma$. By Theorem 1.2.7 of [1], we have $\psi(1) - \psi(\mathcal{D}+1) = h_d$. □

Theorem 3.

- *For an exceptional eigenvalue* $0 < \lambda_j < \mathcal{D}^2$,

$$\left|\widehat{N_\rho(\cdot, \mathbf{p})}_j^{(\mathrm{disc})}\right|^2 = f_{s_j}(e^{-2\rho})^2|\phi_j(\mathbf{p})|^2 e^{2s_j\rho} + c_{\log}(s_j)\rho e^{(d-1)\rho} + O(e^{(d-1)\rho}),$$

where $c_{\log}$ *is as defined in* (6).

- *For* $\lambda_j = \mathcal{D}^2$,

$$\left|\widehat{N_\rho(\cdot, \mathbf{p})}_j^{(\mathrm{disc})}\right|^2 = (4\pi^{d-1}/\Gamma(\mathcal{D}+1)^2)(h_d+\rho)^2|\phi_j(\mathbf{p})|^2 e^{2s_j\rho} + O(e^{(d-1)\rho}).$$

The implied constants depend on Γ.

Proof. This follows from (23) and Proposition 3. □

3.2. *The Leading Term from the Continuous Spectrum*

Theorem 4. *For sufficiently large* ρ,

$$\frac{1}{4\pi}\sum_{\kappa}\int_{-1}^{1}\left|\widehat{N_\rho(\cdot, \mathbf{p})}_\kappa^{(\mathrm{cts})}(\mathcal{D}+it)\right|^2 dt = \frac{\pi^{d-1}}{\Gamma(\mathcal{D}+1)^2}\left|\mathcal{E}\left(\mathcal{D}, \mathbf{p}\right)\right|^2 \rho e^{(d-1)\rho} + O(e^{(d-1)\rho}).$$

The implied constant depends on $\mathbf{p}$ *and* Γ.

Proof. Let $s = \mathcal{D} + it$ be on the critical line, with $t \in [-1, 1]$ and $t \neq 0$. By (19), we have

$$\begin{aligned} {}_2F_1(d-s, \mathcal{D}+1; d+1; 1-e^{-2\rho}) &= \frac{\Gamma(d+1)\Gamma(it)}{\Gamma(\mathcal{D}+1)\Gamma(\mathcal{D}+1+it)}\, {}_2F_1(\mathcal{D}+1-it, \mathcal{D}+1; 1-it; e^{-2\rho}) \\ &\quad + \frac{\Gamma(d+1)\Gamma(-it)e^{-2it\rho}}{\Gamma(\mathcal{D}+1)\Gamma(\mathcal{D}+1-it)}\, {}_2F_1(\mathcal{D}+1+it, \mathcal{D}+1; 1+it; e^{-2\rho}). \end{aligned}$$

Substituting this into Proposition 2, we obtain

$$I(\rho,s)=\pi^{\mathcal{D}}(2\sinh(\rho))^d e^{-(\mathcal{D}+1)\rho}\Big(\frac{\Gamma(it)}{\Gamma(\mathcal{D}+1+it)}e^{it\rho}{}_2F_1(\mathcal{D}+1-it,\mathcal{D}+1;1-it;e^{-2\rho}) + \frac{\Gamma(-it)}{\Gamma(\mathcal{D}+1-it)}e^{-it\rho}{}_2F_1(\mathcal{D}+1+it,\mathcal{D}+1;1+it;e^{-2\rho})\Big).$$

Replacing $\Gamma(it)$ by $\frac{\Gamma(1+it)}{it}$, we can write this as

$$I(\rho,s)=2\pi^{\mathcal{D}}(2\sinh(\rho))^d e^{-(\mathcal{D}+1)\rho}\,\mathrm{Im}\left(\frac{e^{it\rho}}{t}\frac{-\Gamma(1+it)}{\Gamma(\mathcal{D}+1+it)}{}_2F_1(\mathcal{D}+1-it,\mathcal{D}+1;1-it;e^{-2\rho})\right).$$

Hence by (23), we have

$$\frac{1}{4\pi}\sum_\kappa\int_{-1}^1\left|\widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})}(\mathcal{D}+it)\right|^2dt=\pi^{d-2}(2\sinh(\rho))^{2d}e^{-(d+1)\rho}\int_{-1}^1 h(t,\rho)^2\left|\mathcal{E}(\mathbf{p},\mathcal{D}+it)\right|^2dt, \quad (25)$$

where $h(t,\rho)=\mathrm{Im}\left((e^{it\rho}/t)(\Gamma(1+it)/\Gamma(\mathcal{D}+1+it)){}_2F_1(\mathcal{D}+1-it,\mathcal{D}+1;1-it;e^{-2\rho})\right)$. We shall estimate the integral on the right-hand side of (25). To this end, we first break h up into more manageable pieces as follows: $h(t,\rho)=\mathrm{Im}\left(e^{it\rho}/t\right)\mathrm{Re}\left(g(e^{-2\rho},t)\right)+\mathrm{Re}\left(e^{it\rho}/t\right)\mathrm{Im}\left(g(e^{-2\rho},t)\right)$, where $g(x,t)=(\Gamma(1+it)/\Gamma(\mathcal{D}+1+it)){}_2F_1(\mathcal{D}+1-it,\mathcal{D}+1;1-it;x)$. We note that $g(x,t)$ is real analytic for $|x|<\frac{1}{2}$ and $t\in[-1,1]$. We also have $g(x,-t)=\overline{g(x,t)}$. The latter observations show that $\mathrm{Re}\,g(x,t)$ is an even function of t, whereas $\mathrm{Im}\,g(x,t)=tf(x,t)$ for some analytic function f. Hence

$$h(t,\rho)=\mathrm{Im}\left(\frac{e^{it\rho}}{t}\right)\left(\frac{{}_2F_1(\mathcal{D}+1,\mathcal{D}+1;1;e^{-2\rho})}{\Gamma(\mathcal{D}+1)}+O(t^2)\right)+\mathrm{Re}\left(e^{it\rho}\right)f(e^{-2\rho},t),$$

or, more simply, $h(t,\rho)=({}_2F_1(\mathcal{D}+1,\mathcal{D}+1;1;e^{-2\rho})/\Gamma(\mathcal{D}+1))(\sin(t\rho)/t)+\cos(t\rho)f(e^{-2\rho},0)+O(|t|)$. Squaring this, we obtain

$$h(t,\rho)^2=\frac{{}_2F_1(\mathcal{D}+1,\mathcal{D}+1;1;e^{-2\rho})^2}{\Gamma(\mathcal{D}+1)^2}\left(\frac{\sin(t\rho)}{t}\right)^2 + \frac{{}_2F_1(\mathcal{D}+1,\mathcal{D}+1;1;e^{-2\rho})f(e^{-2\rho},0)}{\Gamma(\mathcal{D}+1)}\frac{\sin(2t\rho)}{t}+O(1).$$

Estimating the hypergeometric function by $1+O(e^{-2\rho})$, we have

$$h(t,\rho)^2=\frac{1}{\Gamma(\mathcal{D}+1)^2}\left(\frac{\sin(t\rho)}{t}\right)^2+O(\rho^2e^{-2\rho})+\frac{f(e^{-2\rho},0)}{\Gamma(\mathcal{D}+1)}\frac{\sin(2t\rho)}{t}+O(\rho e^{-2\rho})+O(1).$$

Since $\rho e^{-2\rho}$ and $\rho^2e^{-2\rho}$ are bounded, we obtain

$$h(t,\rho)^2=\frac{1}{\Gamma(\mathcal{D}+1)^2}\left(\frac{\sin(t\rho)}{t}\right)^2+\frac{f(e^{-2\rho},0)}{\Gamma(\mathcal{D}+1)}\frac{\sin(2t\rho)}{t}+O(1). \quad (26)$$

On the other hand, by the functional equation (10) for $\mathcal{E}(\mathbf{p},s)$, we have

$$|\mathcal{E}(\mathbf{p},\mathcal{D}+it)|^2=|\mathcal{E}(\mathbf{p},\mathcal{D})|^2+O(t^2). \quad (27)$$

Putting the estimates (26) and (27) together, we find that

$$h(t,\rho)^2|\mathcal{E}(\mathbf{p},s)|^2=\frac{|\mathcal{E}(\mathbf{p},\mathcal{D})|^2}{\Gamma(\mathcal{D}+1)^2}\left(\frac{\sin(t\rho)}{t}\right)^2+\frac{|\mathcal{E}(\mathbf{p},\mathcal{D})|^2f(e^{-2\rho},0)}{\Gamma(\mathcal{D}+1)}\frac{\sin(2t\rho)}{t}+O(1).$$

Integrating this, we obtain

$$\int_{-1}^1 h(t,\rho)^2|\mathcal{E}(\mathbf{p},s)|^2dt=\frac{|\mathcal{E}(\mathbf{p},\mathcal{D})|^2}{\Gamma(\mathcal{D}+1)^2}\int_{-1}^1\frac{\sin^2(t\rho)}{t^2}dt+\frac{|\mathcal{E}(\mathbf{p},\mathcal{D})|^2f(e^{-2\rho},0)}{\Gamma(\mathcal{D}+1)}\int_{-1}^1\frac{\sin(2t\rho)}{t}dt+O(1).$$

The change of variable $u=\rho t$ gives

$$\int_{-1}^1 h(t,\rho)^2|\mathcal{E}(\mathbf{p},\mathcal{D})|^2dt=\frac{|\mathcal{E}(\mathbf{p},\mathcal{D})|^2}{\Gamma(\mathcal{D}+1)^2}\rho\int_{-\rho}^{\rho}\frac{\sin^2(u)}{u^2}du+\frac{|\mathcal{E}(\mathbf{p},\mathcal{D})|^2f(e^{-2\rho},0)}{\Gamma(\mathcal{D}+1)}\int_{-\rho}^{\rho}\frac{\sin(2u)}{u}dt+O(1).$$

Integrating by parts, we note that the second term on the right-hand side above is bounded; it may therefore be absorbed into the error term. The other integral may be replaced by $\int_{\mathbb{R}} \frac{\sin^2(u)}{u^2} du + O(\rho^{-1}) = \pi + O(\rho^{-1})$. This gives us

$$\int_{-1}^{1} h(t,\rho)^2 |\mathcal{E}(\mathbf{p},\mathcal{D})|^2 dt = \frac{\pi |\mathcal{E}(\mathbf{p},\mathcal{D})|^2}{\Gamma(\mathcal{D}+1)^2} \rho + O(1).$$

Substituting this into (25), we have

$$\frac{1}{4\pi} \sum_{\kappa} \int_{-1}^{1} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) \right|^2 dt = \frac{\pi^{d-1} |\mathcal{E}(\mathbf{p},\mathcal{D})|^2}{\Gamma(\mathcal{D}+1)^2} \rho e^{(d-1)\rho} + O(e^{(d-1)\rho}). \quad \square$$

3.3. *The End of the Critical Line*

Theorem 5.

$$\sum_{\kappa} \int_{1}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) \right|^2 dt + \sum_{j\,:\,\lambda_j > \mathcal{D}^2}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_j^{(\mathrm{disc})} \right|^2 \ll e^{(d-1)\rho}.$$

The implied constant depends on Γ *and* $\mathbf{p}$.

To prove this, we require the following lemma.

Lemma 2. *Let* $\delta > 0$. *For* $s = \mathcal{D} + it$ *with* $|t| > \delta$ *and* $\rho > 1$, *we have* $|I(\rho,s)| \ll e^{\mathcal{D}\rho} |t|^{-\mathcal{D}-1}$. *Moreover, there exists a constant* $C_1 > 0$ *such that for* $|t| > C_1$ *and* $\rho > 1$, *we have* $|I(\rho,s)| \asymp e^{\mathcal{D}\rho} |t|^{-\mathcal{D}-1}$.

Proof. Let $s = \mathcal{D} + it$. By formula (c2) of section II.2.6 of [19], we have for large t

$$\begin{aligned}{}_2F_1(d-s,\mathcal{D}+1;d+1;1-e^{-2\rho}) &= \frac{\Gamma(d+1)}{\Gamma(\mathcal{D}+1)} \left((d-s)(1-e^{-2\rho})\right)^{-\mathcal{D}-1} \\ &\quad \times \left(e^{i\pi(\mathcal{D}+1)} + e^{(d-s)(1-e^{-2\rho})}\right) \left(1 + O(|(d-s)(1-e^{-2\rho})|^{-1})\right).\end{aligned}$$

For large t, the term $(1 + O(|(d-s)(1-e^{-2\rho})|^{-1}))$ is in fact $1 + O(|t|^{-1})$ and, therefore, is bounded away from zero and infinity. Hence, for large $|t|$, we have $\left|{}_2F_1(d-s,\mathcal{D}+1;d+1;1-e^{-2\rho})\right| \asymp t^{-\mathcal{D}-1}$. The result follows from this estimate together with Proposition 2. $\square$

Proof of the theorem. We have by (23)

$$\widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) = I(\rho,\mathcal{D}+it) E_\kappa(\mathcal{D}+it,\mathbf{p}), \qquad \widehat{N_\rho(\cdot,\mathbf{p})}_j^{(\mathrm{disc})} = I(\rho,s_j)\phi_j(\mathbf{p}).$$

Let $|t| > C_1$. Then by the previous lemma, we have

$$\left| \widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) \right| \asymp e^{\mathcal{D}\rho} |t|^{-\frac{d+1}{2}} |E_\kappa(\mathcal{D}+it,\mathbf{p})|, \qquad \left| \widehat{N_\rho(\cdot,\mathbf{p})}_j^{(\mathrm{disc})} \right| \asymp e^{\mathcal{D}\rho} |t|^{-\frac{d+1}{2}} |\phi_j(\mathbf{p})|$$

by the previous lemma. Hence

$$\begin{aligned}&\sum_{\kappa} \int_{C_1}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) \right|^2 dt + \sum_{j\,:\,\mathrm{Im}\, s_j > C_1}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_j^{(\mathrm{disc})} \right|^2 \\ &\qquad \asymp e^{(d-1)\rho} \left(\int_{C_1}^{\infty} |\mathcal{E}(\mathcal{D}+it,\mathbf{p})|^2 t^{-d-1} dt + \sum_{j\,:\,\mathrm{Im}\, s_j > C_1}^{\infty} |\phi_j(\mathbf{p})|^2 t^{-d-1} \right). \quad (28)\end{aligned}$$

The left-hand side of the above formula is finite by the Plancherel formula. Therefore the right-hand side is finite, so the sum and the integral on the right-hand side both converge; this convergence also follows from Chapter 29 of [9]. As the expression in the brackets in the right-hand side of (28) does not depend on ρ, we have

$$\sum_{\kappa} \int_{C_1}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) \right|^2 dt + \sum_{j\,:\,\mathrm{Im}\, s_j > C_1}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_j^{(\mathrm{disc})} \right|^2 \asymp e^{(d-1)\rho}.$$

A similar argument, using the first part of Lemma 2, shows that we can extend this bound to

$$\sum_{\kappa} \int_{1}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_\kappa^{(\mathrm{cts})} (\mathcal{D}+it) \right|^2 dt + \sum_{j\,:\,\lambda_j > \mathcal{D}^2}^{\infty} \left| \widehat{N_\rho(\cdot,\mathbf{p})}_j^{(\mathrm{disc})} \right|^2 \asymp e^{(d-1)\rho}. \quad \square$$

3.4. *The Main Theorems*

To prove Theorem 1, we simply add up the estimates given in Theorems 3 and 5 and apply the Plancherel formula (13). Similarly, to obtain Theorem 2, we simply add up the estimates given in Theorems 3, 4, and 5 and apply (15).

REFERENCES

1. G. E. Andrews, R. Askey, and R. Roy, *Special Functions,* Encyclopedia of Mathematics and Its Applications **71** (Cambridge University Press, 1999).
2. H.-J. Bartels, "Nichteuklidische Gitterpunktprobleme und Gleichverteilung in linearen algebraischen Gruppen," Comment. Math. Helvetici **57**, 158–172 (1982).
3. V. Bentkus and F. Götze, "On the Lattice Point Problem for Ellipsoids," Acta Arith. **80** (2), 101–125 (1997).
4. R. W. Bruggeman, R. J. Miatello, and N. R. Wallach, "Resolvent and Lattice Points on Symmetric Spaces of Strictly Negative Curvature," Math. Ann. **315**, 617–639 (1999).
5. J. R. Cheng, "The Lattice Points in a Circle," Sci. Sinica **12**, 633–649 (1963).
6. J. Delsarte, "Sur le Gitter Fuchsien," C. R. Acad. Sci. Paris **214**, 147–149 (1942).
7. F. Götze, "Lattice Point Problems and Values of Quadratic Forms," Invent. Math. **157**, 195–226 (2004).
8. P. Günther, "Gitterpunktprobleme in symmetrischen Riemannschen Räumen vom Rang 1," Math. Nachr. **94**, 5–27 (1980).
9. L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Vol. 4 (Springer, 1985).
10. H. Huber, "Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. I," Comment. Math. Helv. **30**, 20–62 (1956).
11. H. Huber, "Zur analytischen Theorie Hyperbolischer Raumformen und Bewegungsgruppen," Math. Ann. **138**, 1–26 (1959); **142**, 385–398 (1961); **143**, 463–464 (1961).
12. M. N. Huxley, "Exponential Sums and Lattice Points," Proc. London Math. Soc. **60**, 471–502 (1990).
13. M. N. Huxley, "Corrigenda: 'Exponential Sums and Lattice Points'," Proc. London Math. Soc. **66**, 70 (1993).
14. S. V. Konyagin, M. M. Skriganov, A. V. Sobolev, "On a Lattice Point Problem Arising in the Spectral Analysis of Periodic Operators," Mathematika **50**, 87–98 (2003).
15. E. Landau, "Über Gitterpunkte in mehrdimensionalen Ellipsoiden," Math. Z. **21**, 126–132 (1924).
16. P. D. Lax and R. S. Phillips, "The Asymptotic Distribution of Lattice Points in Euclidean and Non-Euclidean Spaces," Journal of Functional Analysis **46**, 280–350 (1982).
17. B. M. Levitan, "Asymptotic Formulas for the Number of Lattice Points in Euclidean and Lobachevskii Spaces," Uspekhi Mat. Nauk **42** (3), 13–38 (1987).
18. J. E. Littlewood and A. Walfisz, "The Lattice Points of a Circle. (With a Note by Prof. E. Landau.)," Proc. Roy. Soc. London (A) **106**, 478–488 (1925).
19. W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, 3rd ed. (Springer, 1966).
20. L. Parnovski and A. V. Sobolev, "On the Bethe–Sommerfeld Conjecture for the Polyharmonic Operator," Duke Math. J. **107** (2), 209–238 (2001).
21. S. J. Patterson, "A Lattice-Point Problem in Hyperbolic Space," Mathematika **22**, 81–88 (1975); **23**, 227 (1976).
22. R. Phillips and Z. Rudnick, "The Circle Problem in the Hyperbolic Plane," Journal of Functional Analysis **121**, 78–116 (1994).
23. B. Randol, "The Selberg Trace Formula," chapter in *Eigenvalues in Riemannian Geometry* by I. Chavel (Academic Press, 1984).
24. A. Selberg, "Harmonic Analysis and Discontinuous Groups in Weakly Symmetric Riemannian Spaces with Applications to Dirichlet Series," J. Indian Math. Soc. B. **20**, 47–87 (1956).
25. I. M. Vinogradov, "The Number of Lattice Points in a Sphere," Trav. Inst. Math. Stekloff (Leningrade) **9**, 17–38 (1935).
26. W. Wolfe, "The Asymptotic Distribution of Lattice Points in Hyperbolic Space," Journal of Functional Analysis **31**, 333–340 (1979).