JEMS

Daniel El-Baz · Bingrong Huang · Min Lee

Effective joint equidistribution of primitive rational points on expanding horospheres

Abstract. We prove an effective version of a result due to Einsiedler, Mozes, Shah and Shapira who established the equidistribution of primitive rational points on expanding horospheres in the space of unimodular lattices in at least three dimensions. Their proof uses techniques from homogeneous dynamics and relies in particular on measure-classification theorems – an approach which does not lend itself to effective bounds. We implement a strategy based on spectral theory, Fourier analysis and Weil's bound for Kloosterman sums in order to quantify the rate of equidistribution for a specific horospherical subgroup in any dimension. We apply our result to provide a rate of convergence to the limiting distribution for the appropriately rescaled diameters of random circulant graphs.

Keywords. Effective equidistribution, homogeneous spaces, spectral theory, Kloosterman sums

1. Introduction

In recent years, there has been an increased focus on obtaining effective versions of equidistribution theorems in homogeneous dynamics. For the method it introduced, we single out Strömbergsson's breakthrough paper [17] and mention the related work by Browning and Vinogradov [2]. Particularly interesting targets, of which these two papers are instances, consist of results whose proof relies on rigidity theorems such as Ratner's, which are by nature not effective. The primary purpose of this paper is to accomplish this to get an effective version of a result due to Einsiedler, Mozes, Shah and Shapira [5]. Their theorem confirmed a conjecture due to Marklof, who had been able to prove an averaged version thereof and made great use of it [11]. His proof relied on the mixing property of a certain diagonal flow on the space of unimodular lattices and was made effective,

Daniel El-Baz: Institute of Analysis and Number Theory, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria; danielelbaz88@gmail.com

Bingrong Huang: Data Science Institute and School of Mathematics, Shandong University, Jinan, Shandong 250100, China; brhuang@sdu.edu.cn

Min Lee: School of Mathematics, University of Bristol, Woodland Road, Bristol BS8 1UG, United Kingdom; min.lee@bristol.ac.uk

using estimates on the decay of matrix coefficients, by Li [10] who applied it to obtain a quantitative version of Marklof's result concerning the distribution of Frobenius numbers. An article by Marklof and the third author [9] provided a rate of convergence for the Einsiedler–Mozes–Shah–Shapira result for a certain horospherical subgroup in the two-dimensional setting according to the set-up below. We now state our main result, which yields such a rate in any dimension for certain horospherical subgroups.

For $d \geq 1$, let $\Gamma = \mathrm{SL}_{d+1}(\mathbb{Z})$ and define

$$\mathrm{H} = \left\{ \begin{pmatrix} A & \boldsymbol{v} \\ {}^t\boldsymbol{0} & 1 \end{pmatrix} : A \in \mathrm{SL}_d(\mathbb{R}), \, \boldsymbol{v} \in \mathbb{R}^d \right\} \subset \mathrm{SL}_{d+1}(\mathbb{R}). \tag{1.1}$$

Denote by μ_{H} the H-invariant Haar probability measure on $\Gamma\backslash\Gamma\mathrm{H}$. Finally, for $\boldsymbol{x} \in \mathbb{R}^d$, define

$$n_+(\boldsymbol{x}) = \begin{pmatrix} I_d & \boldsymbol{0} \\ {}^t\boldsymbol{x} & 1 \end{pmatrix} \in \mathrm{SL}_{d+1}(\mathbb{R}). \tag{1.2}$$

We note that the group of all matrices of this form is the expanding horospherical subgroup corresponding to the semigroup of matrices of the form $\mathrm{diag}(e^t, \ldots, e^t, e^{-dt}) \in \mathrm{SL}_{d+1}(\mathbb{R})$ with $t > 0$.

Define, for every positive integer q,

$$D(q) = \mathrm{diag}(q^{1/d}, \ldots, q^{1/d}, q^{-1}) \in \mathrm{SL}_{d+1}(\mathbb{R}). \tag{1.3}$$

Note that for every $\boldsymbol{r} \in \mathbb{Z}^d$ satisfying $\gcd(q, \boldsymbol{r}) = 1$, by [5, Lemma 2.1] (see also [11, Remark 3.3, (3.53)] and [10, Lemma 4.1]) we have

$$\Gamma n_+(q^{-1}\boldsymbol{r})D(q) \in \Gamma\backslash\Gamma\mathrm{H}. \tag{1.4}$$

We also provide an explicit proof in Lemma 2.3. We define

$$\mathcal{R}_q = \{\boldsymbol{r} \in (\mathbb{Z} \cap [1, q])^d : \gcd(\boldsymbol{r}, q) = 1\}. \tag{1.5}$$

Let $\mathbb{T} \cong \mathbb{Z}\backslash\mathbb{R}$ be the unit circle and $\mathrm{C}_b^k(\Gamma\backslash\Gamma\mathrm{H} \times \mathbb{T}^d)$ be the space of k times continuously differentiable functions with all derivatives bounded and denote by $\|\cdot\|_{\mathrm{C}_b^k}$ the Sobolev norm (see (4.1)). Our main result is the following theorem.

Theorem 1.1. *For every $d \geq 3$, every $\varepsilon > 0$ and every integer $k \geq 2d^2 - d + 1$, there exists a constant $c > 0$ such that for every function $f \in \mathrm{C}_b^k(\Gamma\backslash\Gamma\mathrm{H} \times \mathbb{T}^d)$ and every $q \in \mathbb{Z}_{\geq 1}$,*

$$\left| \frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{r} \in \mathcal{R}_q} f\left(\Gamma n_+\left(\frac{1}{q}\boldsymbol{r}\right)D(q), \frac{1}{q}\boldsymbol{r} \right) - \int_{\Gamma\backslash\Gamma\mathrm{H}\times\mathbb{T}^d} f \, d\mu_{\mathrm{H}} \, d\boldsymbol{x} \right|$$

$$\leq c\|f\|_{\mathrm{C}_b^k} q^{-1/2 + \frac{d^2(2k-2d+1)}{2k^2} + \varepsilon}. \tag{1.6}$$

Remark 1.1. For $d = 1$, this result was already known to Marklof in an effective form, with rate $O_f(q^{-1/2+\varepsilon})$ [12]. See also [5, Section 2.1] for a more detailed presentation of the argument. We merely mention that it relies on Weil's bound for Kloosterman sums as well, but is otherwise much simpler.

Remark 1.2. We note that our proof also works when $d = 2$ and hence recovers the previous result by Marklof and the third author [9]. In this case, the error term becomes $c\|f\|_{C_b^k} q^{-1/2+\varepsilon}(q^\theta + q^{2(2k-3)/k^2})$ for some constant $c > 0$, where $\theta \geq 0$ is a Ramanujan bound for GL_2 over $\mathbb{Q}$. The Ramanujan conjecture is the assertion that $\theta = 0$ and the current record towards it is a result due to Kim and Sarnak which states that $\theta \leq \frac{7}{64}$, proved in [8, Appendix 2]. The reason for this discrepancy is that for $d \geq 3$, the use of bounds towards the Ramanujan conjecture for GL_d over $\mathbb{Q}$ can be bypassed. Instead, Clozel, Oh and Ullmo [3] exploit the uniform version of Kazhdan's property (T) for $SL_d(\mathbb{Q}_p)$ for all primes p, when $d \geq 3$, as was obtained by Oh [14].

As already hinted at, this result has several applications, for instance to the distribution of Frobenius numbers as in [10, 11] or to results about the shape of lattices as in [5]. We highlight one in particular, which concerns the limiting distribution of the diameters of random Cayley graphs of $\mathbb{Z}/q\mathbb{Z}$ as $q \to +\infty$, following Marklof and Strömbergsson [13] (see also [15] for the case of random Cayley graphs of arbitrary finite abelian groups). In [1], Amir and Gurel-Gurevich conjectured the existence of a limiting distribution, as $q \to +\infty$, for $\mathrm{diam}(q, d)/q^{1/d}$ where $\mathrm{diam}(q, d)$ denotes the diameter of the Cayley graph of $\mathbb{Z}/q\mathbb{Z}$ with respect to the subset $\{\pm a_1, \ldots, \pm a_d\}$ where $(a_1, \ldots, a_d)$ is chosen uniformly at random from $\mathcal{R}_q$. Following the method expounded in [13], the existence of this limiting distribution is a consequence of the main theorem in [5]. By the same token, our Theorem 1.1 implies the following result:

Corollary 1.1. *For every $d \geq 3$, there exists a continuous non-increasing function $\Psi_d : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $\Psi_d(0) = 1$ and $\lim_{R \to \infty} \Psi_d(R) = 0$, and a constant $\eta_d > 0$ such that for every $\varepsilon > 0$ and every $R \geq 0$, we have*

$$\mathrm{Prob}\left(\frac{\mathrm{diam}(q, d)}{q^{1/d}} \geq R\right) = \Psi_d(R) + O(q^{-\eta_d + \varepsilon}), \tag{1.7}$$

where the implicit constant depends on R and ε.

We state a more precise version of the above corollary as Corollary 5.1, which also contains an explicit description of the limiting distribution in terms of the space of d-dimensional unimodular lattices. At this point, we do however note that the decay of Ψ_d as $R \to +\infty$ is known: it is proved in [13, Section 3.3] that for $d \geq 2$,

$$\Psi_d(R) = \frac{1}{2\zeta(d)R^d} + O_d\left(\frac{1}{R^{d+1+\frac{1}{d-1}}}\right). \tag{1.8}$$

In order to deduce this corollary, which we do in Section 5, the explicit dependence on f in the error term of Theorem 1.1 is required.

Our strategy to prove Theorem 1.1 is based on harmonic analysis and Weil's bound for Kloosterman sums, more precisely:

- in Section 2, which contains the main novelty of our approach, we avoid the need to obtain an explicit solution to a (non-linear) system of equations modulo q – as was done for $d = 2$ in [9] – by introducing a helpful parametrisation of $\mathcal{R}_q$;

- we then use Fourier analysis on the space of affine lattices in order to estimate the sum we are interested in – this follows a strategy introduced by Strömbergsson [17] for the space of shifted lattices in two dimensions and we extend the required Fourier tools to any dimension in Section 3;

- these estimates are carried out in Section 4: to get to the main term, the key ingredient is a deep result of Clozel, Oh and Ullmo [3]; to bound the error terms, we use estimates for Ramanujan and Kloosterman sums, combined with various counting arguments.

2. Primitive rational points on horospheres

Let $d \geq 1$, $G = SL_{d+1}(\mathbb{R})$ and $\Gamma = SL_{d+1}(\mathbb{Z})$. For any $g \in G$, we write $g = \begin{pmatrix} A & b \\ {}^t c & D \end{pmatrix}$ where $A \in M_d(\mathbb{R})$, $b, c \in \mathbb{R}^d$ and $D \in \mathbb{R}$. Let I_k be the $k \times k$ identity matrix.

For a positive integer q, we define the following congruence subgroup of $SL_d(\mathbb{Z})$:

$$\Gamma_{0,d}(q) = \left\{ \gamma \in SL_d(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} * & * \\ {}^t 0 & u \end{pmatrix} \pmod{q} \right\}. \tag{2.1}$$

Note that for any $\gamma \in \Gamma_{0,d}(q)$ satisfying $\gamma \equiv \begin{pmatrix} * & * \\ {}^t 0 & u \end{pmatrix} \pmod{q}$, we have $\gcd(u, q) = 1$.

We record the formula for the index of $\Gamma_{0,d}(q)$ inside $SL_d(\mathbb{Z})$.

Proposition 2.1. *For every $d \geq 2$ and $q \geq 1$, we have*

$$[SL_d(\mathbb{Z}) : \Gamma_{0,d}(q)] = q^{d-1} \prod_{p | q} \frac{1 - p^{-d}}{1 - p^{-1}}. \tag{2.2}$$

Proof (sketch). It is a standard fact (for an explicit reference see, for instance, [7, Corollary 2.9]) that

$$\# GL_d(\mathbb{Z}/q\mathbb{Z}) = q^{d^2} \prod_{p | q} \left(1 - \frac{1}{p^d} \right) \left(1 - \frac{1}{p^{d-1}} \right) \cdots \left(1 - \frac{1}{p} \right), \tag{2.3}$$

from which it follows that

$$\# SL_d(\mathbb{Z}/q\mathbb{Z}) = q^{d^2-1} \prod_{p | q} \left(1 - \frac{1}{p^d} \right) \left(1 - \frac{1}{p^{d-1}} \right) \cdots \left(1 - \frac{1}{p^2} \right). \tag{2.4}$$

This last cardinality is precisely the index of the principal congruence subgroup

$$\Gamma_d(q) = \{ M \in SL_d(\mathbb{Z}) : M \equiv I_d \pmod{q} \} \tag{2.5}$$

inside $SL_d(\mathbb{Z})$ (for a reference about the surjectivity of the reduction map, see [16, proof of Lemma 1.38] for instance). We note the inclusions $\Gamma_d(q) \subset \Gamma_{0,d}(q) \subset SL_d(\mathbb{Z})$ and therefore use the identity

$$[SL_d(\mathbb{Z}) : \Gamma_d(q)] = [SL_d(\mathbb{Z}) : \Gamma_{0,d}(q)][\Gamma_{0,d}(q) : \Gamma_d(q)] \tag{2.6}$$

to conclude. All that remains is to compute $[\Gamma_{0,d}(q) : \Gamma_d(q)]$ and it is easy to see that it is equal to $q^{d-1} \# GL_{d-1}(\mathbb{Z}/q\mathbb{Z})$. The desired formula follows. ∎

For a positive integer q, recall that

$$\mathcal{R}_q = \{r \in (\mathbb{Z} \cap (0,q])^d : \gcd(r,q) = 1\}. \tag{2.7}$$

We now give a simple formula and a lower bound for the size of this set.

Lemma 2.1. *For $d \geq 1$ and $q \geq 1$, we have*

$$\#\mathcal{R}_q = \sum_{\delta|q} \mu(\delta)(q/\delta)^d. \tag{2.8}$$

Remark 2.1. Note that when $d = 1$, that is $\varphi(q)$, Euler's totient function, as it should be.

Proof of Lemma 2.1. By partitioning all d-tuples $r \in (\mathbb{Z} \cap [1,q])^d$ according to the value of $\gcd(r,q)$, we see that

$$q^d = \sum_{\delta|q} \#\mathcal{R}_{q/\delta}. \tag{2.9}$$

The claim follows by Möbius inversion. ∎

We note the following trivial corollary.

Corollary 2.1. *For $d \geq 1$ and $q \geq 1$,*

$$\#\mathcal{R}_q = q^d \prod_{p|q} \left(1 - \frac{1}{p^d}\right). \tag{2.10}$$

In particular, for $d \geq 2$ and $q \geq 1$,

$$\#\mathcal{R}_q > \frac{1}{\zeta(d)} q^d. \tag{2.11}$$

Remark 2.2. The above inequality generalises [9, (2.2)], whose proof has an unfortunate mistake (see the first inequality in [9, (2.6)]).

Recall the following subgroup of G:

$$H = \left\{ \begin{pmatrix} A & v \\ {}^t0 & 1 \end{pmatrix} : A \in SL_d(\mathbb{R}),\ v \in \mathbb{R}^d \right\}. \tag{2.12}$$

For a positive integer q, we also recall that

$$D(q) = \begin{pmatrix} q^{1/d} I_d & 0 \\ {}^t0 & q^{-1} \end{pmatrix}. \tag{2.13}$$

By (1.4) we see that, for every $r \in \mathcal{R}_q$, there exist $A \in SL_d(\mathbb{R})$ and $x \in \mathbb{R}^d$ such that

$$\Gamma n_+(q^{-1}r)D(q) = \Gamma \begin{pmatrix} A & x \\ {}^t0 & 1 \end{pmatrix}. \tag{2.14}$$

This is equivalent to the existence of $A \in \mathrm{SL}_d(\mathbb{R})$ and $x \in \mathbb{R}^d$, uniquely determined modulo Γ, satisfying

$$\begin{pmatrix} A & x \\ {}^t\mathbf{0} & 1 \end{pmatrix} (n_+(q^{-1}r)D(q))^{-1} = \begin{pmatrix} A & x \\ {}^t\mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} q^{-1/d}I_d & \mathbf{0} \\ {}^t\mathbf{0} & q \end{pmatrix} \begin{pmatrix} I_d & \mathbf{0} \\ -q^{-1}\,{}^tr & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{q^{1-1/d}A-qx\,{}^tr}{q} & qx \\ -{}^tr & q \end{pmatrix} \in \Gamma. \tag{2.15}$$

Let $s = qx$ and $B = q^{\frac{d-1}{d}}A$. By the above relation,

$$s \in \mathbb{Z}^d, \quad \frac{1}{q}(B - s\,{}^tr) \in \mathrm{M}_d(\mathbb{Z}) \quad \text{and} \quad \det(B) = q^{d-1}\det(A) = q^{d-1}. \tag{2.16}$$

So

$$B \in \mathrm{M}_d(\mathbb{Z}) \quad \text{and} \quad B \equiv s\,{}^tr \pmod{q}. \tag{2.17}$$

Since

$$\begin{pmatrix} \frac{B-s\,{}^tr}{q} & s \\ -{}^tr & q \end{pmatrix} \in \Gamma, \tag{2.18}$$

we get $\gcd(s, q) = 1$ (see also [5, Lemma 2.4]).

We now come to the goal of this section, which is to parametrise $\mathcal{R}_q$ in terms of $\Gamma_{0,d}(q)\backslash\mathrm{SL}_d(\mathbb{Z})$ and $(\mathbb{Z}/q\mathbb{Z})^\times$.

Let $\mathcal{B}_q$ be a set of representatives for $\Gamma_{0,d}(q)\backslash\mathrm{SL}_d(\mathbb{Z})$.

Lemma 2.2. *We have*

$$\mathcal{R}_q = \left\{ {}^t\gamma \begin{pmatrix} \mathbf{0} \\ u \end{pmatrix} \pmod{q} : \gamma \in \mathcal{B}_q, \, u \in (\mathbb{Z}/q\mathbb{Z})^\times \right\}. \tag{2.19}$$

Proof. For any $\gamma \in \mathcal{B}_q$ and $u \in (\mathbb{Z}/q\mathbb{Z})^\times$, there exists $r \in (\mathbb{Z} \cap (0,q])^d$ such that

$$u\,{}^t\gamma e_d = {}^t\gamma \begin{pmatrix} \mathbf{0} \\ u \end{pmatrix} \equiv r \pmod{q}, \tag{2.20}$$

where e_d is the last vector of the canonical basis of $\mathbb{R}^d$. We claim that $\gcd(r, q) = 1$. Let ${}^ta$ be the last row of γ, that is, $a = {}^t\gamma e_d$. If $\gcd(r, q) \neq 1$, then since $ua - r \equiv 0 \pmod{q}$, this implies that $\gcd(ua, q) \neq 1$, so $\gcd(a) \neq 1$. This contradicts the fact that $\gamma \in \mathrm{SL}_d(\mathbb{Z})$.

Note that, using (2.10) and Proposition 2.1, it follows that $\#\mathcal{R}_q = \#\mathcal{B}_q \cdot \varphi(q)$. Therefore, we only need to prove that ${}^t\gamma u e_d \not\equiv {}^t\gamma'u'e_d \pmod{q}$ if $(\gamma, u \pmod{q}) \neq (\gamma', u' \pmod{q})$ for $\gamma, \gamma' \in \mathcal{B}_q$.

Indeed, suppose ${}^t\gamma u e_d \equiv {}^t\gamma'u'e_d \pmod{q}$. Then

$$ {}^t(\gamma(\gamma')^{-1})u e_d \equiv ({}^t\gamma')^{-1}\,{}^t\gamma u e_d \equiv u'e_d \pmod{q},$$

that is, $\gamma(\gamma')^{-1} \in \Gamma_{0,d}(q)$. Since $\gamma, \gamma' \in \mathcal{B}_q$, we get $\gamma = \gamma'$. Using ${}^t\gamma u e_d \equiv {}^t\gamma'u'e_d \pmod{q}$ again, we obtain $u \equiv u' \pmod{q}$. This proves the lemma. ∎

Let $B_0 = \begin{pmatrix} qI_{d-1} & \\ & 1 \end{pmatrix}$; for every $\gamma \in \Gamma_{0,d}(q)\backslash\mathrm{SL}_d(\mathbb{Z})$ and every $u \in (\mathbb{Z}/q\mathbb{Z})^\times$, if we set

$$r \equiv u\,{}^t\gamma e_d \pmod{q}, \tag{2.21}$$

$$s = \bar{u}e_d, \quad u\bar{u} \equiv 1 \pmod{q}, \tag{2.22}$$

$$B = B_0\gamma, \tag{2.23}$$

then by Lemma 2.2 we have $r \in \mathcal{R}_q$. Moreover, $\det(B) = q^{d-1}$ and $B \equiv s\,{}^t r \pmod{q}$. One checks that

$$\begin{pmatrix} q^{-1+1/d}B & q^{-1}s \\ {}^t\mathbf{0} & 1 \end{pmatrix} D(q)^{-1}n_+(-q^{-1}r) \in \Gamma, \tag{2.24}$$

which implies (1.4) and the following lemma, which is used in §4.

Lemma 2.3. *For every $\gamma \in \Gamma_{0,d}(q)\backslash\mathrm{SL}_d(\mathbb{Z})$ and $u \in (\mathbb{Z}/q\mathbb{Z})^\times$, if we define r, s and B as in (2.21)–(2.23), then*

$$\Gamma n_+(q^{-1}r)D(q) = \Gamma \begin{pmatrix} q^{-1+1/d}B & q^{-1}s \\ {}^t\mathbf{0} & 1 \end{pmatrix}. \tag{2.25}$$

3. Fourier analysis on the space of lattice translates

In this section, we generalise the results given in [17, Section 4] and [9, Section 3] to an arbitrary dimension. When comparing with [17], one should keep in mind that he uses a different representation for $\mathrm{ASL}_2(\mathbb{R})$.

For $d \geq 2$, we define

$$\mathrm{ASL}_d(\mathbb{R}) := \mathrm{SL}_d(\mathbb{R}) \ltimes \mathbb{R}^d. \tag{3.1}$$

For $M_1, M_2 \in \mathrm{SL}_d(\mathbb{R})$ and $v_1, v_2 \in \mathbb{R}^d$, the multiplication law on $\mathrm{ASL}_d(\mathbb{R})$ is given by

$$(M_1, v_1) \cdot (M_2, v_2) = (M_1 M_2, M_1 v_2 + v_1). \tag{3.2}$$

The discrete subgroup $\mathrm{ASL}_d(\mathbb{Z})$ is defined similarly.

Let $\mathfrak{g}$ be the Lie algebra of $\mathrm{ASL}_d(\mathbb{R})$, which we identify with $\mathfrak{sl}_d(\mathbb{R}) \oplus \mathbb{R}^d$. We pick the following basis of $\mathfrak{g}$:

$$Y_{i,j} = (E_{i,j}, \mathbf{0}), \quad 1 \leq i \neq j \leq d, \tag{3.3}$$

$$Y_i = (E_{i,i} - E_{1,1}, \mathbf{0}), \quad i \geq 2, \tag{3.4}$$

$$X_i = (\mathbf{0}, e_i), \quad 1 \leq i \leq d, \tag{3.5}$$

where $E_{i,j} \in \mathrm{M}_d(\mathbb{R})$ has a 1 at the (i, j)th entry and zeros elsewhere, and the e_i are the canonical basis of $\mathbb{R}^d$.

Each $(E, \boldsymbol{y}) \in \mathfrak{g}$ yields a left-invariant differential operator on a function on $\mathrm{ASL}_d(\mathbb{R})$ in the following way:

$$
\begin{aligned}
((E, \boldsymbol{y})F)(g, \boldsymbol{x}) &= \frac{\partial}{\partial t} F((g, \boldsymbol{x}) \exp(tE, t\boldsymbol{y})) \Big|_{t=0} \\
&= \frac{\partial}{\partial t} F((g, \boldsymbol{x})((I_d, \boldsymbol{0}) + t(E, \boldsymbol{y}))) \Big|_{t=0}.
\end{aligned}
\tag{3.6}
$$

In particular, for $X_{i_0} = (\boldsymbol{0}, \boldsymbol{e}_{i_0})$, $1 \le i_0 \le d$, by the chain rule, we get

$$
(X_{i_0} F)(g, \boldsymbol{x}) = \sum_{i=1}^{d} g_{i,i_0} \left(\frac{\partial}{\partial x_i} F \right)(g, \boldsymbol{x}),
\tag{3.7}
$$

where $g = (g_{i,j})_{1 \le i,j \le d}$.

Let $\mathrm{C}_b^k(\mathrm{ASL}_d(\mathbb{Z}) \backslash \mathrm{ASL}_d(\mathbb{R}))$ denote the space of k times continuously differentiable functions with all derivatives bounded. For $F \in \mathrm{C}_b^k(\mathrm{ASL}_d(\mathbb{Z}) \backslash \mathrm{ASL}_d(\mathbb{R}))$ we set

$$
\|F\|_{\mathrm{C}_b^k} = \sum_{0 \le \ell \le k} \sum_{\substack{E_i \in \{Y_{i_0,j_0}, X_{i_0}, Y_{i_0}\} \\ 1 \le i \le \ell}} \|E_1 \circ \cdots \circ E_\ell F\|_{L^\infty}.
\tag{3.8}
$$

Let F be a function on $\mathrm{ASL}_d(\mathbb{Z}) \backslash \mathrm{ASL}_d(\mathbb{R})$. From now on, we implicitly identify functions on $\mathrm{ASL}_d(\mathbb{Z}) \backslash \mathrm{ASL}_d(\mathbb{R})$ with $\mathrm{ASL}_d(\mathbb{Z})$-invariant functions on $\mathrm{ASL}_d(\mathbb{R})$. For any $\boldsymbol{m} \in \mathbb{Z}^d$,

$$
F(A, \boldsymbol{x} + \boldsymbol{m}) = F((I_d, \boldsymbol{m})(A, \boldsymbol{x})) = F(A, \boldsymbol{x})
\tag{3.9}
$$

for $(A, \boldsymbol{x}) \in \mathrm{ASL}_d(\mathbb{R})$. So we have the following Fourier expansion of F:

$$
F(A, \boldsymbol{x}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^d} \widehat{F}(A, \boldsymbol{m}) e^{2\pi i \, {}^t\boldsymbol{m}\boldsymbol{x}},
\tag{3.10}
$$

where

$$
\widehat{F}(A, \boldsymbol{m}) = \int_{(\mathbb{R}/\mathbb{Z})^d} F(A, \boldsymbol{t}) e^{-2\pi i \, {}^t\boldsymbol{m}\boldsymbol{t}} \, d\boldsymbol{t}
\tag{3.11}
$$

Here $d\boldsymbol{t}$ denotes integration with respect to the Haar measure induced by the Lebesgue measure on $\mathbb{R}^d$.

Lemma 3.1. *For any $\gamma \in \mathrm{SL}_d(\mathbb{Z})$ we have*

$$
\widehat{F}(\gamma A, \boldsymbol{m}) = \widehat{F}(A, {}^t\gamma \boldsymbol{m}).
\tag{3.12}
$$

In particular, when $\boldsymbol{m} = \boldsymbol{0}$, $\widehat{F}(A, \boldsymbol{0})$ is an automorphic function on $\mathrm{SL}_d(\mathbb{Z}) \backslash \mathrm{SL}_d(\mathbb{R})$.

Proof. For any $\gamma \in \mathrm{SL}_d(\mathbb{Z})$, we get

$$
\begin{aligned}
\widehat{F}(\gamma A, \boldsymbol{m}) &= \int_{(\mathbb{R}/\mathbb{Z})^d} F(\gamma A, \boldsymbol{t}) e^{-2\pi i \, {}^t\boldsymbol{m}\boldsymbol{t}} \, d\boldsymbol{t} = \int_{(\mathbb{R}/\mathbb{Z})^d} F((\gamma, \boldsymbol{0})(A, \gamma^{-1}\boldsymbol{t})) e^{-2\pi i \, {}^t\boldsymbol{m}\boldsymbol{t}} \, d\boldsymbol{t} \\
&= \int_{(\mathbb{R}/\mathbb{Z})^d} F(A, \boldsymbol{t}) e^{-2\pi i \, {}^t({}^t\gamma \boldsymbol{m})\boldsymbol{t}} \, d\boldsymbol{t} = \widehat{F}(A, {}^t\gamma \boldsymbol{m}).
\end{aligned}
\tag{3.13}
$$

Here in the third identity we use the fact that F is left $\mathrm{ASL}_d(\mathbb{Z})$-invariant and $t \mapsto \gamma t$ is a diffeomorphism of $(\mathbb{R}/\mathbb{Z})^d$ preserving the volume measure dt. ∎

Set $A = (a_{i,j})_{1 \le i,j \le d}$. For each $1 \le i_0 \le d$, by applying integration by parts, we get

$$
\begin{aligned}
\widehat{(X_{i_0}F)}(A, \boldsymbol{m}) &= \int_{(\mathbb{R}/\mathbb{Z})^d} (X_{i_0}F)(A, t)e^{-2\pi\mathrm{i}^{\mathrm{t}}\boldsymbol{m}t}\, dt \\
&= \sum_{i=1}^{d} a_{i,i_0} \int_{(\mathbb{R}/\mathbb{Z})^d} \frac{\partial}{\partial t_i} F(A, t)e^{-2\pi\mathrm{i}^{\mathrm{t}}\boldsymbol{m}t}\, dt \\
&= \Big(\sum_{i=1}^{d} a_{i,i_0} 2\pi\mathrm{i}m_i\Big) \int_{(\mathbb{R}/\mathbb{Z})^d} F(A, t)e^{-2\pi\mathrm{i}^{\mathrm{t}}\boldsymbol{m}t}\, dt \\
&= 2\pi\mathrm{i}\Big(\sum_{i=1}^{d} m_i a_{i,i_0}\Big)\widehat{F}(A, \boldsymbol{m}).
\end{aligned}
\tag{3.14}
$$

So for $k \in \mathbb{Z}_{\ge 1}$,

$$
\int_{(\mathbb{R}/\mathbb{Z})^d} (X_{i_0}^k F)(A, t)e^{-2\pi\mathrm{i}^{\mathrm{t}}\boldsymbol{m}t}\, dt = \Big(2\pi\mathrm{i}\sum_{i=1}^{d} m_i a_{i,i_0}\Big)^k \widehat{F}(A, \boldsymbol{m}),
\tag{3.15}
$$

and we get

$$
\begin{aligned}
(2\pi)^k \Big|\sum_{i=1}^{d} m_i a_{i,i_0}\Big|^k |\widehat{F}(A, \boldsymbol{m})| &\le \int_{(\mathbb{R}/\mathbb{Z})^d} |(X_{i_0}^k F)(A, t)|\, dt \\
&\le \|X_{i_0}^k F\|_\infty.
\end{aligned}
\tag{3.16}
$$

For $\boldsymbol{b} \in \mathbb{R}^d$, let $\|\boldsymbol{b}\|_\infty := \max_{1 \le i \le d} |b_i|$. Then

$$
\max_{1 \le i_0 \le d} \Big|\sum_{i=1}^{d} m_i a_{i,i_0}\Big|^k = \|{}^{\mathrm{t}}A\boldsymbol{m}\|_\infty^k,
\tag{3.17}
$$

and we have

$$
\begin{aligned}
(2\pi \|{}^{\mathrm{t}}A\boldsymbol{m}\|_\infty)^k |\widehat{F}(A, \boldsymbol{m})| &= \max_{1 \le i_0 \le d} (2\pi)^k \Big|\sum_{i=1}^{d} a_{i,i_0} m_i\Big|^k |\widehat{F}(A, \boldsymbol{m})| \\
&\le \max_{1 \le i_0 \le d} \|X_{i_0}^k F\|_\infty \le \|F\|_{C_b^k}.
\end{aligned}
\tag{3.18}
$$

So for $\boldsymbol{m} \ne \boldsymbol{0}$, we have

$$
|\widehat{F}(A, \boldsymbol{m})| \le \frac{\|F\|_{C_b^k}}{(2\pi \|{}^{\mathrm{t}}A\boldsymbol{m}\|_\infty)^k}.
\tag{3.19}
$$

4. Proof of the main theorem

In this section, we prove Theorem 1.1.

For $f \in C_b^k(\Gamma \backslash \Gamma H \times (\mathbb{R}/\mathbb{Z})^d)$, using the fact that $\Gamma \backslash \Gamma H$ is diffeomorphic to $\mathrm{ASL}_d(\mathbb{Z}) \backslash \mathrm{ASL}_d(\mathbb{R})$, we set, similarly to (3.8),

$$
\|f\|_{C_b^k} = \sum_{0 \le \ell \le k} \sum_{\substack{E_i \in \{Y_{i_0, j_0}, X_{i_0}, Y_{i_0}\} \\ 1 \le i \le \ell}} \sum_{\substack{\ell_1, \dots, \ell_d \ge 0 \\ \ell_1 + \cdots + \ell_d + \ell \le k}} \left\| E_1 \circ \cdots \circ E_\ell \frac{\partial^{\ell_1}}{\partial x_1^{\ell_1}} \cdots \frac{\partial^{\ell_d}}{\partial x_d^{\ell_d}} f \right\|_{L^\infty}.
$$
(4.1)

We have the Fourier expansion

$$
f(g, \boldsymbol{x}) = \sum_{\boldsymbol{n} \in \mathbb{Z}^d} \hat{f}_{\boldsymbol{n}}(g) e^{2\pi i \,{}^t \boldsymbol{n} \boldsymbol{x}}
$$
(4.2)

where

$$
\hat{f}_{\boldsymbol{n}}(g) = \int_{(\mathbb{R}/\mathbb{Z})^d} f(g, \boldsymbol{x}) e^{-2\pi i \,{}^t \boldsymbol{n} \boldsymbol{x}} \, d\boldsymbol{x}.
$$
(4.3)

By using integration by parts repeatedly, for $\boldsymbol{n} \ne \boldsymbol{0}$, we have

$$
\sup_{g \in \Gamma \backslash \Gamma H} |\hat{f}_{\boldsymbol{n}}(g)| \ll_k \|f\|_{C_b^k} \|\boldsymbol{n}\|_\infty^{-k}.
$$
(4.4)

By Lemma 2.3, we have (recall that $\mathcal{B}_q$ is a set of representatives for $\Gamma_{0,d}(q) \backslash \mathrm{SL}_d(\mathbb{Z})$):

$$
\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{r} \in \mathcal{R}_q} f\left(\Gamma n_+\left(\frac{1}{q}\boldsymbol{r}\right) D(q), \frac{1}{q}\boldsymbol{r} \right)
$$

$$
= \frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{u \in (\mathbb{Z}/q\mathbb{Z})^\times} f\left(\begin{pmatrix} q^{-1+1/d} B_0 \gamma & q^{-1} \overline{u} \boldsymbol{e}_d \\ {}^t \boldsymbol{0} & 1 \end{pmatrix}, q^{-1} u \,{}^t \gamma \boldsymbol{e}_d \right)
$$

$$
= \frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{u \in (\mathbb{Z}/q\mathbb{Z})^\times} \sum_{\boldsymbol{n} \in \mathbb{Z}^d} \hat{f}_{\boldsymbol{n}}\left(\begin{pmatrix} q^{-1+1/d} B_0 \gamma & q^{-1} \overline{u} \boldsymbol{e}_d \\ {}^t \boldsymbol{0} & 1 \end{pmatrix} \right) e^{2\pi i \frac{{}^t \boldsymbol{n} u \,{}^t \gamma \boldsymbol{e}_d}{q}}.
$$
(4.5)

We first note that we can truncate $\boldsymbol{n}$-sum at $\|\boldsymbol{n}\|_\infty \le q^{\vartheta_1}$ for some small $0 < \vartheta_1 < \frac{1}{2}$. Indeed, by (4.4), we know that the contribution from the terms with $\|\boldsymbol{n}\|_\infty > q^{\vartheta_1}$ is

$$
\ll_k \frac{\|f\|_{C_b^k}}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{u \in (\mathbb{Z}/q\mathbb{Z})^\times} \sum_{\substack{\boldsymbol{n} \in \mathbb{Z}^d \\ \|\boldsymbol{n}\|_\infty > q^{\vartheta_1}}} \frac{1}{\|\boldsymbol{n}\|_\infty^k} \ll_k \|f\|_{C_b^k} \sum_{\substack{\boldsymbol{n} \in \mathbb{Z}^d \\ \|\boldsymbol{n}\|_\infty > q^{\vartheta_1}}} \frac{1}{\|\boldsymbol{n}\|_\infty^k}.
$$
(4.6)

Note that $\|\boldsymbol{n}\|_\infty \le \|\boldsymbol{n}\|_2 \le \sqrt{d} \, \|\boldsymbol{n}\|_\infty$. It is a standard fact that

$$
\sum_{\substack{\boldsymbol{n} \in \mathbb{Z}^d \\ \|\boldsymbol{n}\|_\infty > q^{\vartheta_1}}} \frac{1}{\|\boldsymbol{n}\|_\infty^k} \ll_{d,k} (q^{\vartheta_1})^{d-k} = q^{-\vartheta_1(k-d)}.
$$
(4.7)

So we have

$$\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{r}\in\mathcal{R}_q} f\left(\Gamma n_+\left(\frac{1}{q}\boldsymbol{r}\right)D(q),\frac{1}{q}\boldsymbol{r}\right)$$

$$= \frac{1}{\#\mathcal{R}_q} \sum_{\gamma\in\mathcal{B}_q} \sum_{u\in(\mathbb{Z}/q\mathbb{Z})^\times} \sum_{\substack{\boldsymbol{n}\in\mathbb{Z}^d \\ \|\boldsymbol{n}\|_\infty\leq q^{\vartheta_1}}} \widehat{f_{\boldsymbol{n}}}\left(\begin{pmatrix} q^{-1+1/d}B_0\gamma & q^{-1}\bar{u}\boldsymbol{e}_d \\ {}^{\mathrm{t}}\boldsymbol{0} & 1 \end{pmatrix}\right)e^{2\pi\mathrm{i}\frac{{}^{\mathrm{t}}\boldsymbol{n}u\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d}{q}} \tag{4.8}$$

$$+ O_{d,k}(\|f\|_{\mathrm{C}_b^k}q^{-\vartheta_1(k-d)}).$$

For $A\in\mathrm{SL}_d(\mathbb{R})$ and $\boldsymbol{y}\in\mathbb{R}^d$, let

$$F_{\boldsymbol{n}}(A,\boldsymbol{y}) = \widehat{f_{\boldsymbol{n}}}\left(\begin{pmatrix} A & \boldsymbol{y} \\ {}^{\mathrm{t}}\boldsymbol{0} & 1 \end{pmatrix}\right). \tag{4.9}$$

Then $F_{\boldsymbol{n}}$ is a function on $\mathrm{ASL}_d(\mathbb{Z})\backslash\mathrm{ASL}_d(\mathbb{R})$ and as such has the Fourier expansion

$$F_{\boldsymbol{n}}(A,\boldsymbol{y}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^d} \widehat{F}_{\boldsymbol{n}}(A,\boldsymbol{m})e^{2\pi\mathrm{i}\,{}^{\mathrm{t}}\boldsymbol{m}\boldsymbol{y}}. \tag{4.10}$$

Here

$$\widehat{F}_{\boldsymbol{n}}(A,\boldsymbol{m}) = \int_{(\mathbb{R}/\mathbb{Z})^d} F_{\boldsymbol{n}}(A,\boldsymbol{t})e^{-2\pi\mathrm{i}\,{}^{\mathrm{t}}\boldsymbol{m}\boldsymbol{t}}\,d\boldsymbol{t}. \tag{4.11}$$

Recall that $B_0 = \left(\begin{smallmatrix} qI_{d-1} & \\ & 1 \end{smallmatrix}\right)$. By (4.8) and (4.10), we get

$$\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{r}\in\mathcal{R}_q} f\left(\Gamma n_+\left(\frac{1}{q}\boldsymbol{r}\right)D(q),\frac{1}{q}\boldsymbol{r}\right)$$

$$= \frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{n}\in\mathbb{Z}^d} \sum_{\boldsymbol{m}\in\mathbb{Z}^d} \sum_{\gamma\in\mathcal{B}_q} \widehat{F}_{\boldsymbol{n}}(q^{-1+1/d}B_0\gamma,\boldsymbol{m}) \sum_{u\in(\mathbb{Z}/q\mathbb{Z})^\times} e^{2\pi\mathrm{i}\frac{{}^{\mathrm{t}}\boldsymbol{m}\bar{u}\boldsymbol{e}_d+{}^{\mathrm{t}}\boldsymbol{n}u\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d}{q}}$$

$$= \frac{1}{\#\mathcal{R}_q} \sum_{\substack{\boldsymbol{n}\in\mathbb{Z}^d \\ \|\boldsymbol{n}\|_\infty\leq q^{\vartheta_1}}} \sum_{\boldsymbol{m}\in\mathbb{Z}^d} \sum_{\gamma\in\mathcal{B}_q} \widehat{F}_{\boldsymbol{n}}(q^{-1+1/d}B_0\gamma,\boldsymbol{m})S(m_d,{}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d;q)$$

$$+ O_{d,k}(\|f\|_{\mathrm{C}_b^k}q^{-\vartheta_1(k-d)}). \tag{4.12}$$

Here $S(a,b;q) = \sum_{u\in(\mathbb{Z}/q\mathbb{Z})^\times} e^{2\pi\mathrm{i}\frac{a\bar{u}+bu}{q}}$ is the classical Kloosterman sum.

Note that by (3.19), for $\boldsymbol{m}\neq\boldsymbol{0}$, we have

$$|\widehat{F}_{\boldsymbol{n}}(q^{-1+1/d}B_0\gamma,\boldsymbol{m})| \leq \frac{\|F\|_{\mathrm{C}_b^k}}{(2\pi\|q^{-1+1/d}\,{}^{\mathrm{t}}\gamma B_0\boldsymbol{m}\|_\infty)^k}. \tag{4.13}$$

We bound (4.12) by considering the following four distinct cases in the subsequent four propositions:

- $\boldsymbol{m}=\boldsymbol{n}=\boldsymbol{0}$;
- $\boldsymbol{n}\neq\boldsymbol{0}$ and $\boldsymbol{m}=\boldsymbol{0}$;

- $n, m \neq 0$ and ${}^{\mathrm{t}}\gamma q^{-1+1/d} B_0 m$ "small";
- $n, m \neq 0$ and ${}^{\mathrm{t}}\gamma q^{-1+1/d} B_0 m$ "large".

We prove these propositions in the subsections below, from which Theorem 1.1 follows as we explain after the statements of the propositions.

Proposition 4.1. *For every $\varepsilon > 0$ and every integer $k \geq d^2$, we have*

$$\frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \widehat{F}_{\mathbf{0}}(q^{-1+1/d} B_0 \gamma, \mathbf{0}) S(0, 0; q)$$
$$= \int_{\mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})} \widehat{F}_{\mathbf{0}}(A, \mathbf{0}) \, d\mu(A) + O_\varepsilon(\|\widehat{F}_{\mathbf{0}}(*, \mathbf{0})\|_{\mathrm{C}_b^k} q^{-1/2+\varepsilon}). \quad (4.14)$$

Proposition 4.2. *For each $\mathbf{0} \neq n \in \mathbb{Z}^d$ with $\|n\|_\infty \leq q^{\vartheta_1}$, we have*

$$\mathcal{E}_1 = \frac{1}{\#\mathcal{R}_q} \left| \sum_{\gamma \in \mathcal{B}_q} \widehat{F}_{\boldsymbol{n}}(q^{-1+1/d} B_0 \gamma, \mathbf{0}) S(0, {}^{\mathrm{t}}n\, {}^{\mathrm{t}}\gamma e_d; q) \right| \ll_\varepsilon \|F\|_{\mathrm{C}_b^0} q^{-1+\vartheta_1+\varepsilon}. \quad (4.15)$$

Proposition 4.3. *For each $n \in \mathbb{Z}^d$ with $\|n\|_\infty \leq q^{\vartheta_1}$ and $0 < \vartheta_2 < \frac{1}{2d}$, we have*

$$\mathcal{E}_2 = \frac{1}{\#\mathcal{R}_q} \left| \sum_{\substack{m \in \mathbb{Z}^d \backslash \{\mathbf{0}\}, \\ \|{}^{\mathrm{t}}\gamma q^{-1+1/d} B_0 m\|_\infty \leq q^{\vartheta_2}}} \sum_{\gamma \in \mathcal{B}_q} \widehat{F}_{\boldsymbol{n}}(q^{-1+1/d} B_0 \gamma, m) S(m_d, {}^{\mathrm{t}}n\, {}^{\mathrm{t}}\gamma e_d; q) \right|$$
$$\leq \|F\|_{\mathrm{C}_b^0} q^{-1/2+d\vartheta_2} \frac{\sigma_0(q)^2}{\prod_{p|q}(1 - p^{-1})}. \quad (4.16)$$

Here $\sigma_0(q)$ is the number of positive divisors of q.

Proposition 4.4. *For each $n \in \mathbb{Z}^d$ with $\|n\|_\infty \leq q^{\vartheta_1}$ and $0 < \vartheta_2 < \frac{1}{2d}$, we have*

$$\mathcal{E}_3 = \frac{1}{\#\mathcal{R}_q} \left| \sum_{\substack{m \in \mathbb{Z}^d \backslash \{\mathbf{0}\}, \\ \|{}^{\mathrm{t}}\gamma q^{-1+1/d} B_0 m\|_\infty > q^{\vartheta_2}}} \sum_{\gamma \in \mathcal{B}_q} \widehat{F}_{\boldsymbol{n}}(q^{-1+1/d} B_0 \gamma, m) S(m_d, {}^{\mathrm{t}}n\, {}^{\mathrm{t}}\gamma e_d; q) \right|$$
$$\ll_{d,k,\vartheta_2} \|F\|_{\mathrm{C}_b^k} q^{-1/2+d\vartheta_2}, \quad (4.17)$$

provided k is an integer such that $k \geq \frac{2d-1}{2\vartheta_2}$.

Proof of Theorem 1.1. By (4.12) and Propositions 4.1–4.4, we have

$$\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{r} \in \mathcal{R}_q} f\left(\Gamma n_+ \left(\frac{1}{q} \boldsymbol{r}\right) D(q), \frac{1}{q} \boldsymbol{r}\right) = \int_{\mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})} \widehat{F}_{\mathbf{0}}(A, \mathbf{0}) \, d\mu(A)$$
$$+ O(\|f\|_{\mathrm{C}_b^k} q^{-\vartheta_1(k-d)})$$
$$+ O(\|f\|_{\mathrm{C}_b^k} q^{-1/2+d(\vartheta_1+\vartheta_2)+\varepsilon}), \quad (4.18)$$

for any $\varepsilon > 0$, $0 < \vartheta_1 < \frac{1}{2}$, $0 < \vartheta_2 < \frac{1}{2d}$ and $k \geq \max\{\frac{2d-1}{2\vartheta_2}, d^2\}$. Note that by (4.11), (4.9), and (4.3), we have

$$\int_{\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})} \widehat{F}_{\mathbf{0}}(A, \mathbf{0}) \, d\mu(A) = \int_{\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})} \int_{(\mathbb{R}/\mathbb{Z})^d} \widehat{f}_0\left(\begin{pmatrix} A & t \\ {}^t\mathbf{0} & 1 \end{pmatrix}\right) dt \, d\mu(A)$$

$$= \int_{\Gamma\backslash\Gamma_{\mathrm{H}}\times\mathbb{T}^d} f \, d\mu_{\mathrm{H}} \, d\mathbf{x}. \quad (4.19)$$

Taking $\vartheta_2 = \frac{2d-1}{2k}$ and $\vartheta_1 = \frac{1/2-d\vartheta_2}{k}$, we see that $\vartheta_2 = \frac{2d-1}{2k} < \frac{1}{2k}$ and $0 < \vartheta_1 = \frac{1/2-d\vartheta_2}{k}$ if and only if $k \geq 2d^2 - d + 1$ (which is at least d^2, so that Proposition 4.1 applies). This proves Theorem 1.1. ∎

4.1. The main term: effective equidistribution of Hecke points

Recall that $\mathscr{B}_q$ is a set of representatives for $\Gamma_{0,d}(q)\backslash\mathrm{SL}_d(\mathbb{Z})$. Throughout this section, we let $B_0 = \begin{pmatrix} qI_{d-1} & \\ & 1 \end{pmatrix}$.

Lemma 4.1. *We have*

$$\mathrm{SL}_d(\mathbb{Z}) B_0 \, \mathrm{SL}_d(\mathbb{Z}) = \dot{\bigcup_{\delta \in \mathscr{B}_q}} \mathrm{SL}_d(\mathbb{Z})(B_0\delta). \quad (4.20)$$

Proof. We first check that the decomposition on the right hand side is disjoint. For $\delta_1, \delta_2 \in \mathscr{B}_q$, if $\gamma B_0\delta_1 = B_0\delta_2$ for some $\gamma \in \mathrm{SL}_d(\mathbb{Z})$, then $B_0^{-1}\gamma B_0 = \delta_2\delta_1^{-1} \in \mathrm{SL}_d(\mathbb{Z})$. Note that in this case

$$\delta_2\delta_1^{-1} = B_0^{-1}\gamma B_0 = \begin{pmatrix} q^{-1}I_{d-1} & \\ & 1 \end{pmatrix} \gamma \begin{pmatrix} qI_{d-1} & \\ & 1 \end{pmatrix} \in \Gamma_{0,d}(q). \quad (4.21)$$

So we get $\delta_2 \in \Gamma_{0,d}(q)\delta_1$.

From the construction, it is clear that

$$\mathrm{SL}_d(\mathbb{Z}) B_0 \, \mathrm{SL}_d(\mathbb{Z}) \supset \dot{\bigcup_{\delta \in \mathscr{B}_q}} \mathrm{SL}_d(\mathbb{Z})(B_0\delta). \quad (4.22)$$

Let $\tau = \begin{pmatrix} T & t \\ {}^t\mathbf{s} & t \end{pmatrix} \in \Gamma_{0,d}(q)$ for $\mathbf{s} \equiv \mathbf{0} \pmod{q}$. Then

$$B_0\tau B_0^{-1} = \begin{pmatrix} qI_{d-1} & \\ & 1 \end{pmatrix} \begin{pmatrix} T & t \\ {}^t\mathbf{s} & t \end{pmatrix} \begin{pmatrix} q^{-1}I_{d-1} & \\ & 1 \end{pmatrix} = \begin{pmatrix} T & qt \\ q^{-1}\,{}^t\mathbf{s} & t \end{pmatrix} \in \mathrm{SL}_d(\mathbb{Z}), \quad (4.23)$$

so $B_0\Gamma_{0,d}(q)B_0^{-1} \subset \mathrm{SL}_d(\mathbb{Z})$. Take $\gamma_1, \gamma_2 \in \mathrm{SL}_d(\mathbb{Z})$. There exists $\delta_2 \in \mathscr{B}_q$ such that $\gamma_2 \in \Gamma_{0,d}(q)\delta_2$. We have

$$\gamma_1 B_0\gamma_2 \in \gamma_1 B_0\Gamma_{0,d}(q)\delta_2 = \gamma_1(B_0\Gamma_{0,d}(q)B_0^{-1})B_0\delta_2 \subset \mathrm{SL}_d(\mathbb{Z})B_0\delta_2 \quad (4.24)$$

and this implies that

$$\mathrm{SL}_d(\mathbb{Z}) B_0 \, \mathrm{SL}_d(\mathbb{Z}) \subset \dot{\bigcup_{\delta \in \mathscr{B}_q}} \mathrm{SL}_d(\mathbb{Z})(B_0\delta), \quad (4.25)$$

as claimed. ∎

Note that $\det(q^{-\frac{d-1}{d}}B_0) = 1$ so $q^{-\frac{d-1}{d}}B_0 \in \mathrm{SL}_d(\mathbb{R})$. For a function $F : \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R}) \to \mathbb{C}$, following [3], the Hecke operator for B_0 is defined as

$$(T_{B_0}F)(g) = \frac{1}{\#(\mathcal{B}_q)} \sum_{\delta \in \mathcal{B}_q} F(q^{-\frac{d-1}{d}}B_0\delta g). \tag{4.26}$$

Assume that $F \in L^2(\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R}))$. Following the argument in [3, Section 1] with [3, Theorem 1.1] and the formula from [3, p. 346], we get

$$\left\| T_{B_0}F - \int_{\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})} F(g)\,d\mu(g) \right\|_2 \ll_\varepsilon q^{-1/2+\varepsilon}\|F\|_2, \tag{4.27}$$

for any $\varepsilon > 0$. Note that the implicit constant only depends on ε. By [4, Proposition 8.2], we find that this L^2-convergence implies the same rate for pointwise convergence: for every integer $k \geq d^2$, if $F \in \mathrm{C}_b^k(\mathrm{ASL}_d(\mathbb{Z})\backslash\mathrm{ASL}_d(\mathbb{R}))$, then we get

$$\left| T_{B_0}F(I_d) - \int_{\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})} F(g)\,d\mu(g) \right| \ll_\varepsilon q^{-1/2+\varepsilon}\|F\|_{\mathrm{C}_b^k}. \tag{4.28}$$

Proof of Proposition 4.1. Since f is bounded, $\widehat{F}_0(*, \mathbf{0}) \in L^2(\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R}))$, where the invariance under $\mathrm{SL}_d(\mathbb{Z})$ follows from Lemma 3.1.

For $S(0, 0; q) = \varphi(q)$, we get

$$\frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \widehat{F}_0(q^{-1+1/d}B_0\gamma, \mathbf{0})S(0, 0; q)$$

$$= \frac{1}{\#\mathcal{B}_q} \sum_{\gamma \in \mathcal{B}_q} \widehat{F}_0(q^{-1+1/d}B_0\gamma, \mathbf{0}) = T_{B_0}\widehat{F}_0(I_d, \mathbf{0}). \tag{4.29}$$

By (4.28), for any integer $k \geq d^2$, we get

$$\left| T_{B_0}\widehat{F}_0(I_d, \mathbf{0}) - \int_{\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})} \widehat{F}_0(g, \mathbf{0})\,d\mu(g) \right| \ll_\varepsilon q^{-1/2+\varepsilon}\|\widehat{F}_0(*, \mathbf{0})\|_{\mathrm{C}_b^k}. \tag{4.30}$$

This completes the proof of Proposition 4.1. ∎

4.2. The first error term

Proof of Proposition 4.2. Note that $\widehat{F}_{\boldsymbol{n}}(q^{-1+1/d}B_0\gamma, \mathbf{0}) \ll \|F\|_{\mathrm{C}_b^0}$ and

$$|S(0, {}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d; q)| = \left| \mu\left(\frac{q}{\gcd(q, {}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d)} \right) \right| \frac{\varphi(q)}{\varphi\left(\frac{q}{\gcd(q, {}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d)} \right)} \leq \gcd(q, {}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d). \tag{4.31}$$

The first equality holds since $S(0, {}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d; q)$ is a Ramanujan sum. Hence

$$\mathcal{E}_1 \ll \|F\|_{\mathrm{C}_b^0}\frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \gcd(q, {}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d) \ll \|F\|_{\mathrm{C}_b^0}\frac{1}{\#\mathcal{R}_q} \sum_{\ell|q} \ell \sum_{\substack{\gamma \in \mathcal{B}_q \\ \gcd({}^{\mathrm{t}}\boldsymbol{n}\,{}^{\mathrm{t}}\gamma\boldsymbol{e}_d, q)=\ell}} 1. \tag{4.32}$$

For each $\ell \mid q$, let

$$\mathcal{S}_\ell = \{\gamma \in \mathcal{B}_q : \gcd({}^t\boldsymbol{n}\,{}^t\gamma\boldsymbol{e}_d, q) = \ell\}. \tag{4.33}$$

Then

$$\mathcal{E}_1 \ll \|F\|_{C_b^0} \frac{1}{\#\mathcal{R}_q} \sum_{\ell \mid q} \ell \#\mathcal{S}_\ell. \tag{4.34}$$

Since $\boldsymbol{n} = \begin{pmatrix} n_1 \\ \vdots \\ n_d \end{pmatrix} \neq \boldsymbol{0}$, there exists $1 \leq j_0 \leq d$ such that $n_{j_0} \neq 0$. For $\gamma \in \mathcal{S}_\ell$, let $\boldsymbol{a} = {}^t\gamma\boldsymbol{e}_d$ be the last row of γ. Then

$$^t\boldsymbol{n}({}^t\gamma\boldsymbol{e}_d) = {}^t\boldsymbol{n}\boldsymbol{a} = n_1 a_1 + \cdots + n_d a_d \equiv 0 \pmod{\ell} \tag{4.35}$$

and this implies that

$$n_{j_0} a_{j_0} \equiv - \sum_{1 \leq j \leq d,\, j \neq j_0} n_j a_j \pmod{\ell}. \tag{4.36}$$

Consequently, $\gcd(n_{j_0}, \ell) \mid \sum_{1 \leq j \leq d,\, j \neq j_0} n_j a_j$ and we get

$$a_{j_0} \equiv -\tilde{n}_{j_0} \frac{\sum_{1 \leq j \leq d,\, j \neq j_0} n_j a_j}{\gcd(n_{j_0}, \ell)} \pmod{\ell/\gcd(n_{j_0}, \ell)}, \tag{4.37}$$

where $\tilde{n}_{j_0} \frac{n_{j_0}}{\gcd(n_{j_0}, \ell)} \equiv 1 \pmod{\ell/\gcd(n_{j_0}, \ell)}$. We further note that, since $\gamma \in \mathrm{SL}_d(\mathbb{Z})$, $\gcd(\boldsymbol{a}) = 1 = \gcd(\boldsymbol{a}, q)$.

For $\ell \mid q$, let $\ell_0 = \gcd(n_{j_0}, \ell)$ and define

$$\mathcal{A}_{\ell_0} = \left\{ \begin{array}{l} \boldsymbol{a} \in (\mathbb{Z}/q\mathbb{Z})^d \\ \gcd(\boldsymbol{a}, q) = 1 \end{array} : \begin{array}{l} \ell_0 \mid \sum_{1 \leq j \leq d,\, j \neq j_0} n_j a_j, \\ a_{j_0} \equiv -\tilde{n}_{j_0} \frac{1}{\ell_0} \sum_{1 \leq j \leq d,\, j \neq j_0} n_j a_j \pmod{\ell/\ell_0} \end{array} \right\}. \tag{4.38}$$

By the above arguments, we deduce that for each $\gamma \in \mathcal{S}_\ell$, there exists $\boldsymbol{a} \in \mathcal{A}_{\ell_0}$ such that $\boldsymbol{a} \equiv {}^t\gamma\boldsymbol{e}_d \pmod{q}$. Note that $(\mathbb{Z}/q\mathbb{Z})^\times$ acts on the set $\mathcal{A}_{\ell_0}$ by scalar multiplication and whenever $u \in (\mathbb{Z}/q\mathbb{Z})^\times$, $\mathcal{A}_{\ell_0} = u\mathcal{A}_{\ell_0}$. If ${}^t\gamma\boldsymbol{e}_d \equiv u\,{}^t\gamma'\boldsymbol{e}_d \pmod{q}$ for some $u \in (\mathbb{Z}/q\mathbb{Z})^\times$, then $\gamma' \in \Gamma_{0,d}(q)\gamma$. So the map $\gamma \mapsto {}^t\gamma\boldsymbol{e}_d \pmod{q}$ is an injection into the set of orbits of $\mathcal{A}_{\ell_0}$ under the action of $(\mathbb{Z}/q\mathbb{Z})^\times$ and hence

$$\#\mathcal{S}_\ell \leq \frac{1}{\varphi(q)} \#\mathcal{A}_{\ell_0}. \tag{4.39}$$

The number of elements in $\mathcal{A}_{\ell_0}$ can, by its definition, be bounded as follows:

$$\#\mathcal{A}_{\ell_0} \leq q^{d-1} \frac{q}{\ell/\ell_0}. \tag{4.40}$$

Combining both of the above inequalities yields

$$\#\mathcal{S}_\ell \leq \frac{q^d}{\varphi(q)} \frac{\gcd(n_{j_0}, \ell)}{\ell}. \tag{4.41}$$

For $n_{j_0} \neq 0$ and $|n_{j_0}| \leq \|\boldsymbol{n}\|_\infty \leq q^{\vartheta_1}$, we have $\gcd(n_{j_0}, \ell) \leq q^{\vartheta_1}$ and inserting (4.41) into (4.34), we get

$$\mathcal{E}_1 \ll \|F\|_{C_b^0} \frac{1}{\#\mathcal{R}_q} \frac{q^d}{\varphi(q)} \sum_{\ell \mid q} \gcd(n_{j_0}, \ell) \ll \|F\|_{C_b^0} \frac{1}{\#\mathcal{R}_q} \frac{q^d}{\varphi(q)} q^{\vartheta_1} \sigma_0(q)$$

$$\ll_{d,\varepsilon} \|F\|_{C_b^0} q^{-1+\vartheta_1+\varepsilon}, \tag{4.42}$$

where we use (2.11) for the last bound. This completes the proof of Proposition 4.2. ∎

4.3. The second error term

Proof of Proposition 4.3. By (4.13),

$$\mathcal{E}_2 \leq \|F\|_{C_b^0} \frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{\substack{\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|{}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty \leq q^{1-1/d+\vartheta_2}}} |S(m_d, {}^{\mathrm{t}}\boldsymbol{n}\, {}^{\mathrm{t}}\gamma \boldsymbol{e}_d; q)|. \tag{4.43}$$

By Weil's bound for Kloosterman sums [6],

$$|S(m_d, {}^{\mathrm{t}}\boldsymbol{n}\, {}^{\mathrm{t}}\gamma \boldsymbol{e}_d; q)| \leq \sqrt{q}\, \gcd(m_d, {}^{\mathrm{t}}\boldsymbol{n}\, {}^{\mathrm{t}}\gamma \boldsymbol{e}_d, q)^{1/2} \sigma_0(q)$$

$$\leq \sqrt{q}\, \gcd(m_d, q)^{1/2} \sigma_0(q). \tag{4.44}$$

We thus have

$$\mathcal{E}_2 \leq \|F\|_{C_b^0} \frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{\substack{\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|{}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty \leq q^{1-1/d+\vartheta_2}}} \sqrt{q}\, \gcd(m_d, q)^{1/2} \sigma_0(q). \tag{4.45}$$

For $\gamma \in \mathcal{B}_q$ and $\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$ with $\gcd(q, m_d) = \ell$, we have

$$ {}^{\mathrm{t}}\gamma B_0 \boldsymbol{m} = {}^{\mathrm{t}}\gamma \begin{pmatrix} q m_1 \\ \vdots \\ q m_{d-1} \\ m_d \end{pmatrix} = \ell\, {}^{\mathrm{t}}\gamma \begin{pmatrix} \frac{q}{\ell} m_1 \\ \vdots \\ \frac{q}{\ell} m_{d-1} \\ \frac{m_d}{\ell} \end{pmatrix}. \tag{4.46}$$

Note that $\gcd(q/\ell, m_d/\ell) = 1$. Set

$$ {}^{\mathrm{t}}\gamma \frac{1}{\ell} B_0 \boldsymbol{m} = \boldsymbol{x} \in \mathbb{Z}^d. \tag{4.47}$$

Since $\gamma \in \mathrm{SL}_d(\mathbb{Z})$, $\boldsymbol{x} = \boldsymbol{0}$ if and only if $\boldsymbol{m} = \boldsymbol{0}$. Assume that $\|{}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty \leq q^{1-1/d+\vartheta_2}$. Then $\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$ and $\|\boldsymbol{x}\|_\infty \leq \frac{q^{1-1/d+\vartheta_2}}{\ell}$. Moreover, since $\|\boldsymbol{x}\|_\infty < 1$ if and only if $\boldsymbol{x} = \boldsymbol{0}$, we only consider $\ell \mid q$ such that $\frac{q^{1-1/d+\vartheta_2}}{\ell} \geq 1$.

Summarising, for each given $\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$ with $\|\boldsymbol{x}\|_\infty \leq \frac{q^{1-1/d+\vartheta_2}}{\ell}$, we count the number of $\gamma \in \mathcal{B}_q$ such that ${}^{\mathrm{t}}\gamma \boldsymbol{m} = \boldsymbol{x}$ has an integral solution $\boldsymbol{m} \in \mathbb{Z}^d$ satisfying $\frac{q}{\ell} \mid m_j$

for $1 \le j \le d-1$ and $\gcd(q/\ell, m_d) = 1$. Moreover, the solution $\boldsymbol{m}$ is uniquely determined since $\boldsymbol{m} = {}^t\gamma^{-1}\boldsymbol{x}$. So we can write

$$
\frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{\substack{\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|{}^t\gamma B_0 \boldsymbol{m}\|_\infty \le q^{1-1/d+\vartheta_2}}} \sqrt{q} \, \gcd(m_d, q)^{1/2} \sigma_0(q)
$$

$$
= \frac{1}{\#\mathcal{R}_q} \sqrt{q} \, \sigma_0(q) \sum_{\substack{\ell \mid q \\ \ell \le q^{1-x1/d+\vartheta_2}}} \ell^{1/2} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty \le \frac{q^{1-1/d+\vartheta_2}}{\ell}}} \sum_{\gamma \in \mathcal{B}_q} \sum_{\substack{\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \frac{q}{\ell} \mid m_j, \, 1 \le j \le d-1 \\ \gcd(m_d, q/\ell) = 1 \\ {}^t\gamma \boldsymbol{m} = \boldsymbol{x}}} 1. \quad (4.48)
$$

For $\ell \mid q$ satisfying $\ell \le q^{1-1/d+\vartheta_2}$, and for each $\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$ with $\|\boldsymbol{x}\|_\infty \le q^{1-1/d+\vartheta_2}/\ell$, let

$$
S_\ell(\boldsymbol{x}) = \left\{ \gamma \in \mathcal{B}_q : \begin{array}{l} \exists \boldsymbol{m} \in \mathbb{Z}^d \text{ satisfies } {}^t\gamma \boldsymbol{m} = \boldsymbol{x}, \\ \frac{q}{\ell} \mid m_j, \, 1 \le j \le d-1, \, \gcd(m_d, q/\ell) = 1 \end{array} \right\}. \quad (4.49)
$$

Then (4.48) is equal to

$$
\frac{1}{\#\mathcal{R}_q} \sqrt{q} \, \sigma_0(q) \sum_{\substack{\ell \mid q \\ \ell \le q^{1-1/d+\vartheta_2}}} \ell^{1/2} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty \le q^{1-1/d+\vartheta_2}/\ell}} \#S_\ell(\boldsymbol{x}). \quad (4.50)
$$

We claim that

$$
\#S_\ell(\boldsymbol{x}) \le \frac{[\mathrm{SL}_d(\mathbb{Z}) : \Gamma_{0,d}(q)]}{[\mathrm{SL}_d(\mathbb{Z}) : \Gamma_{0,d}(q/\ell)]} = \ell^{d-1} \prod_{p \mid \ell, \, p \nmid q/\ell} \frac{1-p^{-d}}{1-p^{-1}}. \quad (4.51)
$$

Indeed, for $\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$, consider γ and $\widetilde{\gamma}$ in $\mathrm{SL}_d(\mathbb{Z})$ such that there exist $\boldsymbol{m}$ and $\boldsymbol{n}$ satisfying: ${}^t\gamma \boldsymbol{m} = \boldsymbol{x}, {}^t\widetilde{\gamma} \boldsymbol{n} = \boldsymbol{x}$ with $\frac{q}{\ell} \mid m_i$ and $\frac{q}{\ell} \mid n_i$ for $1 \le i \le d-1$, while $\gcd(\frac{q}{\ell}, m_d) = \gcd(\frac{q}{\ell}, n_d) = 1$. It follows that

$$
\boldsymbol{m} = {}^t\gamma^{-1}\boldsymbol{x} = {}^t\gamma^{-1} \, {}^t\widetilde{\gamma}\boldsymbol{n} = {}^t(\widetilde{\gamma}\gamma^{-1})\boldsymbol{n}. \quad (4.52)
$$

Upon reducing modulo $\frac{q}{\ell}$, we get

$$
{}^t(\widetilde{\gamma}\gamma^{-1}) \begin{pmatrix} \boldsymbol{0} \\ n_d \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{0} \\ m_d \end{pmatrix} \pmod{q/\ell} \quad (4.53)
$$

with n_d and m_d both invertible modulo q/ℓ. This means $\widetilde{\gamma}\gamma^{-1} \in \Gamma_{0,d}(q/\ell)$. Now fix one matrix $\gamma \in \mathrm{SL}_d(\mathbb{Z})$ satisfying the condition in the definition of $S_\ell(\boldsymbol{x})$ (if no such γ exists, then $\#S_\ell(\boldsymbol{x}) = 0$ and the claim is proved). Every matrix $\widetilde{\gamma} \in S_\ell(\boldsymbol{x})$ is then of the form $\widetilde{\gamma} = \delta\gamma$ for some $\delta \in \Gamma_{0,d}(q/\ell)$. Hence $\#S_\ell(\boldsymbol{x})$ is bounded by the number of distinct $\Gamma_{0,d}(q)$-cosets of the form $\Gamma_{0,d}(q)\delta\gamma$ with $\delta \in \Gamma_{0,d}(q/\ell)$. One can check that for any $\delta, \delta' \in \Gamma_{0,d}(q/\ell)$, we have $\Gamma_{0,d}(q)\delta\gamma = \Gamma_{0,d}(q)\delta'\gamma$ if and only if $\delta'\delta^{-1} \in \Gamma_{0,d}(q)$. Therefore

$$
\#S_\ell(\boldsymbol{x}) \le \#(\Gamma_{0,d}(q) \backslash \Gamma_{0,d}(q/\ell)) = [\Gamma_{0,d}(q/\ell) : \Gamma_{0,d}(q)], \quad (4.54)
$$

which is precisely the inequality in (4.51). The equality follows from Proposition 2.1.

Then we have

$$
\frac{1}{\#\mathcal{R}_q} \sqrt{q}\, \sigma_0(q) \sum_{\ell \mid q} \ell^{1/2} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty \leq \frac{q^{1-1/d+\vartheta_2}}{\ell}}} \#S_\ell(\boldsymbol{x})
$$

$$
\leq \frac{1}{\#\mathcal{R}_q} \sqrt{q}\, \sigma_0(q) \sum_{\substack{\ell \mid q \\ \ell \leq q^{1-1/d+\vartheta_2}}} \ell^{1/2} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty \leq \frac{q^{1-1/d+\vartheta_2}}{\ell}}} \ell^{d-1} \prod_{p \mid \ell,\, p \nmid q/\ell} \frac{1-p^{-d}}{1-p^{-1}}
$$

$$
\ll \frac{1}{\#\mathcal{R}_q} \sqrt{q}\, \sigma_0(q) \prod_{p \mid q} \frac{1-p^{-d}}{1-p^{-1}} \sum_{\substack{\ell \mid q \\ \ell \leq q^{1-1/d+\vartheta_2}}} \ell^{1/2} \left(\frac{q^{1-1/d+\vartheta_2}}{\ell} \right)^d \ell^{d-1}
$$

$$
= \frac{1}{\#\mathcal{R}_q} q^{d-1/2+\vartheta_2 d} \sigma_0(q) \prod_{p \mid q} \frac{1-p^{-d}}{1-p^{-1}} \sum_{\substack{\ell \mid q \\ \ell \leq q^{1-1/d+\vartheta_2}}} \ell^{-1/2}
$$

$$
\leq \frac{1}{\#\mathcal{R}_q} q^{d-1/2+\vartheta_2 d} \sigma_0(q)^2 \prod_{p \mid q} \frac{1-p^{-d}}{1-p^{-1}}. \tag{4.55}
$$

Using (2.10), we get

$$
\frac{1}{\#\mathcal{R}_q} \sqrt{q}\, \sigma_0(q) \sum_{\ell \mid q} \ell^{1/2} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty \leq \frac{q^{1-1/d+\vartheta_2}}{\ell}}} \#S_\ell(\boldsymbol{x})
$$

$$
\leq q^{-1/2+\vartheta_2 d} \frac{\sigma_0(q)^2}{\prod_{p \mid q}(1-p^{-1})}. \tag{4.56}
$$

This proves Proposition 4.3. ∎

4.4. The third error term

Proof of Proposition 4.4. By (4.13), for any integer $k \geq 0$, we have

$$
\mathcal{E}_3 \leq \frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{\substack{\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|{}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty > q^{1-1/d+\vartheta_2}}} \frac{\|F\|_{C_b^k} |S(m_d, {}^{\mathrm{t}}\boldsymbol{n}({}^{\mathrm{t}}\gamma \boldsymbol{e}_d); q)|}{(2\pi \|q^{-1+1/d}\, {}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty)^k}. \tag{4.57}
$$

By the trivial bound for the Kloosterman sum $|S(m_d, {}^{\mathrm{t}}\boldsymbol{n}({}^{\mathrm{t}}\gamma \boldsymbol{e}_d); q)| \leq \varphi(q)$, we have

$$
\mathcal{E}_3 \leq \|F\|_{C_b^k} \frac{\varphi(q)}{(2\pi)^k} \frac{1}{\#\mathcal{R}_q} \sum_{\gamma \in \mathcal{B}_q} \sum_{\substack{\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|{}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty > q^{1-1/d+\vartheta_2}}} \frac{q^{k(1-1/d)}}{\|{}^{\mathrm{t}}\gamma B_0 \boldsymbol{m}\|_\infty^k}. \tag{4.58}
$$

For each $\gamma \in \mathcal{B}_q$ and $\boldsymbol{m} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$, let ${}^t\gamma B_0 \boldsymbol{m} = \boldsymbol{x}$. Then we have

$$\mathcal{E}_3 \leq \|F\|_{C_b^k} \frac{1}{(2\pi)^k} \frac{\varphi(q)\#\mathcal{B}_q}{\#\mathcal{R}_q} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty > q^{1-1/d+\vartheta_2}}} \frac{q^{k(1-1/d)}}{\|\boldsymbol{x}\|_\infty^{k-d-\vartheta_2}} \frac{1}{\|\boldsymbol{x}\|_\infty^{d+\vartheta_2}}$$

$$\leq \|F\|_{C_b^k} \frac{1}{(2\pi)^k} q^{d-1-\vartheta_2(k-d-1+1/d-\vartheta_2)} \sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty > q^{1-1/d+\vartheta_2}}} \frac{1}{\|\boldsymbol{x}\|_\infty^{d+\vartheta_2}}. \qquad (4.59)$$

By the same argument used to obtain (4.7), we have

$$\sum_{\substack{\boldsymbol{x} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \|\boldsymbol{x}\|_\infty > q^{1-1/d+\vartheta_2}}} \frac{1}{\|\boldsymbol{x}\|_\infty^{d+\vartheta_2}} \ll_{d,\vartheta_2} (q^{1-1/d+\vartheta_2})^{-\vartheta_2}. \qquad (4.60)$$

Thus, by (4.59), we have

$$\mathcal{E}_3 \ll_{d,\vartheta_2,k} \|F\|_{C_b^k} q^{d-1-\vartheta_2(k-d)} \ll_{d,\vartheta_2,k} \|F\|_{C_b^k} q^{-1/2+\vartheta_2 d},$$

provided that $k \geq \frac{2d-1}{2\vartheta_2}$. This proves Proposition 4.4. ∎

5. An application: diameters of random circulant graphs

In this section, we denote by X the space of unimodular lattices in $\mathbb{R}^d$.

We abuse notation and still denote by $C_b^k(X)$ the space of k times continuously differentiable functions f from X to $\mathbb{R}$ such that for every left-invariant differential operator D on $\mathrm{SL}_d(\mathbb{R})$ of order at most k, $\|Df\|_\infty$ is finite. Likewise, for a function $f \in C_b^k(X)$, we still denote by $\|f\|_{C_b^k}$ the obvious analogue of (3.8).

Define, for $q \geq 2$ and $d \geq 2$, the $(d+1)$-dimensional lattice $\Lambda_q = \mathbb{Z}^d \times q\mathbb{Z}$. For $\boldsymbol{a} \in (\mathbb{Z} \cap [1, q])^d$ with $\gcd(\boldsymbol{a}, q) = 1$ (meaning $\boldsymbol{a} \in \mathcal{R}_q$), define

$$n(\boldsymbol{a}) = \begin{pmatrix} I_d & \boldsymbol{a} \\ {}^t\boldsymbol{0} & 1 \end{pmatrix} \in \mathrm{SL}_{d+1}(\mathbb{Z}). \qquad (5.1)$$

Consider $\Lambda_q(\boldsymbol{a})_0 = \Lambda_q n(\boldsymbol{a}) \cap (\mathbb{R}^d \times \{0\})$. Finally, define $D_q = q^{-1/d} I_d \in \mathrm{GL}_d(\mathbb{R})$, so that $\det(D_q) = q^{-1}$. Consider the d-dimensional lattice $L_{q,\boldsymbol{a}} = \Lambda_q(\boldsymbol{a})_0 D_q$. Following the steps used to prove [13, Theorem 3], with Theorem 1.1 replacing the use of [13, Theorem 4], we see that Theorem 1.1 implies:

Theorem 5.1. *For every $d \geq 3$, every $\varepsilon > 0$ and every function $f \in C_b^k(X)$, with an integer $k \geq 2d^2 - d + 1$, we have*

$$\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{a} \in \mathcal{R}_q} f(L_{q,\boldsymbol{a}}) = \int_X f \, d\mu + O\big(\|f\|_{C_b^k} q^{-1/2+\frac{d^2(2k-2d+1)}{2k^2}+\varepsilon}\big). \qquad (5.2)$$

Remark 5.1. When $d = 2$, a version of this theorem follows from [9, Theorem 1.3] (see also Remark 1.2).

As explained in the introduction, we can use this theorem to deduce the following rate of convergence for the limiting distribution of the appropriately rescaled diameters of random circulant graphs.

To help understand what follows, we briefly summarise the key steps in the relevant parts of Marklof and Strömbergsson's paper [13]. The first step (see [13, Section 2.2] for more details) is to identify the circulant graph $C_q(\boldsymbol{a})$ – that is, the Cayley graph of $\mathbb{Z}/q\mathbb{Z}$ with respect to the a_i – with a lattice graph on a torus:

(1) consider the graph LG_d whose vertices are the points of the lattice $\mathbb{Z}^d$ and whose edges are of the form $(\boldsymbol{k}, \boldsymbol{k} + \boldsymbol{e}_i)$ for some $\boldsymbol{k} \in \mathbb{Z}^d$, where $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d)$ is the canonical basis of $\mathbb{R}^d$;

(2) introduce a metric m on LG_d by defining the distance between two vertices $\boldsymbol{k}$ and $\boldsymbol{l}$ in $\mathbb{Z}^d$ to be $m(\boldsymbol{k}, \boldsymbol{l}) = \sum_{i=1}^{d} |k_i - l_i|$;

(3) extend this metric in the obvious way to a metric on $\mathbb{Z}^d/\Lambda$ where Λ is a sublattice of $\mathbb{Z}^d$;

(4) [13, Lemma 2] is the assertion that $LG_d/\Lambda_q(\boldsymbol{a})_0$ and $C_q(\boldsymbol{a})$ are isomorphic as metric graphs.

The next step is to relate the diameter of $LG_d/\Lambda_q(\boldsymbol{a})_0$ – which, by the first step, is exactly the diameter $\mathrm{diam}(q, d)$ we are interested in – to the diameter of $\mathbb{R}^d/L_{q,\boldsymbol{a}}$ (where the distance on the torus is the ℓ^1 distance): [13, Proposition 1] asserts that

$$q^{1/d} \mathrm{diam}(\mathbb{R}^d/L_{q,\boldsymbol{a}}) - d/2 \leq \mathrm{diam}(LG_d/\Lambda_q(\boldsymbol{a})_0) \leq q^{1/d} \mathrm{diam}(\mathbb{R}^d/L_{q,\boldsymbol{a}}). \quad (5.3)$$

The final step [13, Lemma 4] connects the diameter of a torus $\mathbb{R}^d/L$ to the covering radius of the d-orthoplex with respect to the lattice $L \subset \mathbb{R}^d$:

$$\mathrm{diam}(\mathbb{R}^d/L) = \rho(\mathfrak{B}, L). \quad (5.4)$$

We recall that the latter quantity is defined to be

$$\rho(\mathfrak{B}, L) = \inf\{r > 0 : r\mathfrak{B} + L = \mathbb{R}^d\} \quad (5.5)$$

and that for $d \geq 2$, the d-orthoplex is the polytope

$$\mathfrak{B} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_1 \leq 1\}. \quad (5.6)$$

We can now state the consequence of Theorem 5.1 pertaining to the diameters of random circulant graphs.

Corollary 5.1. *For every $d \geq 3$, there exists a continuous non-increasing function $\Psi_d \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $\Psi_d(0) = 1$ and $\lim_{R \to \infty} \Psi_d(R) = 0$ such that for every $\varepsilon > 0$ and every $R \geq 0$, we have*

$$\mathrm{Prob}\left(\frac{\mathrm{diam}(q, d)}{q^{1/d}} \geq R\right) = \Psi_d(R) + O_{R,\varepsilon}(q^{-\eta_d + \varepsilon}), \quad (5.7)$$

where

$$\eta_d = \frac{2d^2 - 2d + 1}{2(2d^2 - d + 1)^2(2d^2 - d + 2)}.$$

Moreover, for $R \geq 0$, Ψ_d is explicitly given by

$$\Psi_d(R) = \mu(\{L \in X : \rho(\mathfrak{P}, L) \geq R\}) \tag{5.8}$$

where μ is the Haar probability measure on X.

It should be clear from the discussion preceding the above corollary that its proof requires an approximation argument to pass from the smooth functions in Theorem 5.1 to characteristic functions. We borrow the following definition from Li's paper [10, Definition 1.3]:

Definition 5.1. A subset of X is said to have *thin boundary* if its boundary is contained in the union of finitely many connected smooth submanifolds of X, all of which have codimension at least 1.

We also borrow (in a slightly modified form) the following technical lemma from a paper by Strömbergsson and Venkatesh [18, Lemma 1]. For a set $S \subset X$, we denote by $\chi_S \colon X \to \{0, 1\}$ its characteristic function.

Lemma 5.1. *If $S \subset X$ has thin boundary, then for each $\delta \in (0, 1)$, there exist functions f_- and f_+ in $C^\infty(X)$ such that for every $k \geq 1$,*

(1) $0 \leq f_- \leq \chi_S \leq f_+ \leq 1$;

(2) $\|f_-\|_{C_b^k} \ll_S \delta^{-k}$ *and* $\|f_+\|_{C_b^k} \ll_S \delta^{-k}$;

(3) $\|f_- - \chi_S\|_{L^1} \ll_S \delta$ *and* $\|f_+ - \chi_S\|_{L^1} \ll_S \delta$.

We can finally proceed with the proof of Corollary 5.1.

Proof of Corollary 5.1. Define, for $R \geq 0$, the following subset of d-dimensional unimodular lattices:

$$S_R = \{L \in X : \rho(\mathfrak{P}, L) \geq R\} \tag{5.9}$$

where ρ is the covering radius and $\mathfrak{P}$ is the d-orthoplex.

In order to deduce Corollary 5.1, we wish to apply Theorem 5.1 to χ_{S_R} for each $R \geq 0$. To do so, we make use of Lemma 5.1 to approximate this characteristic function by smooth functions. For this, we first need to show that, for each $R \geq 0$, the set S_R has thin boundary according to Definition 5.1. However, this follows from the proof of [13, Lemma 7]. We therefore find smooth functions f_- and f_+ as in Lemma 5.1. Applying Corollary 5.1 to each of those and using their properties, we conclude that for every $\delta \in (0, 1)$, every $\varepsilon > 0$ and every $k \geq 2d^2 - d + 1$,

$$\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{a} \in \mathcal{R}_q} \chi_{S_R}(L_{q,\boldsymbol{a}}) = \int_X \chi_{S_R} \, d\mu + O_R(\delta + \delta^{-k} q^{-1/2 + \vartheta + \varepsilon}) \tag{5.10}$$

with $\vartheta = \frac{d^2(2k-2d+1)}{2k^2}$. If we now choose $\delta = q^{\frac{-1/2+\vartheta}{k+1}}$, we see that for every $\varepsilon > 0$,

$$\frac{1}{\#\mathcal{R}_q} \sum_{\boldsymbol{a}\in\mathcal{R}_q} \chi_{S_R}(L_{q,\boldsymbol{a}}) = \int_X \chi_{S_R} d\mu + O_R(q^{\kappa_{d,k}+\varepsilon}) \tag{5.11}$$

with $\kappa_{d,k} = \frac{-k^2+d^2(2k-2d+1)}{2k^2(k+1)}$. Finally, picking $k = 2d^2 - d + 1$ we get the desired error term with $\eta_d \sim \frac{1}{8d^4}$ as claimed. ∎

References

[1] Amir, G., Gurel-Gurevich, O.: The diameter of a random Cayley graph of $\mathbb{Z}_q$. Groups Complex. Cryptol. **2**, 59–65 (2010) Zbl 1194.05054 MR 2672553

[2] Browning, T., Vinogradov, I.: Effective Ratner theorem for $SL(2,\mathbb{R}) \ltimes \mathbb{R}^2$ and gaps in $\sqrt{n}$ modulo 1. J. London Math. Soc. (2) **94**, 61–84 (2016) Zbl 1368.37008 MR 3532164

[3] Clozel, L., Oh, H., Ullmo, E.: Hecke operators and equidistribution of Hecke points. Invent. Math. **144**, 327–351 (2001) Zbl 1144.11301 MR 1827734

[4] Clozel, L., Ullmo, E.: Équidistribution des points de Hecke. In: Contributions to Automorphic Forms, Geometry, and Number Theory, Johns Hopkins Univ. Press, Baltimore, MD, 193–254 (2004) Zbl 1068.11042 MR 2058609

[5] Einsiedler, M., Mozes, S., Shah, N., Shapira, U.: Equidistribution of primitive rational points on expanding horospheres. Compos. Math. **152**, 667–692 (2016) Zbl 1382.37032 MR 3484111

[6] Estermann, T.: On Kloosterman's sum. Mathematika **8**, 83–86 (1961) Zbl 0114.26302 MR 126420

[7] Han, J.: The general linear group over a ring. Bull. Korean Math. Soc. **43**, 619–626 (2006) Zbl 1111.20040 MR 2264921

[8] Kim, H. H.: Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2. J. Amer. Math. Soc. **16**, 139–183 (2003) Zbl 1018.11024 MR 1937203

[9] Lee, M., Marklof, J.: Effective equidistribution of rational points on expanding horospheres. Int. Math. Res. Notices **2018**, 6581–6610 Zbl 1442.37052 MR 3873538

[10] Li, H.: Effective limit distribution of the Frobenius numbers. Compos. Math. **151**, 898–916 (2015) Zbl 1351.37020 MR 3347994

[11] Marklof, J.: The asymptotic distribution of Frobenius numbers. Invent. Math. **181**, 179–207 (2010) Zbl 1200.11022 MR 2651383

[12] Marklof, J.: Horospheres, Farey fractions and Frobenius numbers. Oberwolfach Reports **29**, 28–32 (2010) Zbl 1291.37043

[13] Marklof, J., Strömbergsson, A.: Diameters of random circulant graphs. Combinatorica **33**, 429–466 (2013) Zbl 1340.05063 MR 3133777

[14] Oh, H.: Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants. Duke Math. J. **113**, 133–192 (2002) Zbl 1011.22007 MR 1905394

[15] Shapira, U., Zuck, R.: Asymptotic metric behavior of random Cayley graphs of finite abelian groups. Combinatorica **39**, 1133–1148 (2019) Zbl 1449.05136 MR 4039604

[16] Shimura, G.: Introduction to the Arithmetic Theory of Automorphic Functions. Publ. Math. Soc. Japan 11, Princeton Univ. Press, Princeton, NJ (1994) Zbl 0221.10029 MR 1291394

[17] Strömbergsson, A.: An effective Ratner equidistribution result for $\mathrm{SL}(2, \mathbb{R}) \ltimes \mathbb{R}^2$. Duke Math. J. **164**, 843–902 (2015) Zbl 1351.37014 MR 3332893

[18] Strömbergsson, A., Venkatesh, A.: Small solutions to linear congruences and Hecke equidistribution. Acta Arith. **118**, 41–78 (2005) Zbl 1082.11061 MR 2136220