ON THE LOCAL-GLOBAL PRINCIPLE FOR INTEGRAL APOLLONIAN 3-CIRCLE PACKINGS

XIN ZHANG

Abstract. In this paper we study the integral properties of Apollonian-3 circle packings, which are variants of the standard Apollonian circle packings. Specifically, we study the reduction theory, formulate a local-global conjecture, and prove a density one version of this conjecture. Along the way, we prove a uniform spectral gap for the congruence towers of the symmetry group.

1. Introduction

Apollonian circle packings are well-known planar fractal sets. Starting with three mutually tangent circles, we inscribe one circle into each curvilinear triangle. Repeat this process ad infinitum and we get an Apollonian circle packing. Soddy first observed the existence of some Apollonian packings with all circles having integer curvatures, and we call these packings *integral*. The systematic study of the integers from such packings was initiated by Graham, Lagarias, Mallows, Wilks, and Yan [9] [10]. We first briefly review what is known for integral Apollonian packings. Fix an integral Apollonian packing $\mathcal{P}$, and let $\mathcal{K}$ be the set of curvatures from $\mathcal{P}$. Without loss of generality we can assume $\mathcal{P}$ is *primitive* (i.e. the *gcd* of $\mathcal{K}$ is 1). We say an integer n is *admissible* if it passes all local obstructions (i.e. for any q, we can find $\kappa \in \mathcal{K}$ such that $n \equiv \kappa \pmod{q}$). Finally, let Γ be the orientation-preserving symmetry group acting on $\mathcal{P}$, which is an infinite co-volume Kleinian group. We have:

(1) *The reduction theorem*: Fuchs in her thesis [7] proved that an integer is admissible if and only if it passes the local obstruction at 24.

(2) *The local-global conjecture*: Graham, Lagarias, Mallows, Wilks, Yan [9] conjectured that every sufficiently large admissible integer is actually a curvature.

(3) *A congruence subgroup*: Sarnak [20] observed that there is a real congruence subgroup lying in Γ. As a consequence, some curvatures can be represented by certain shifted quadratic forms.

(4) *The congruence towers of* Γ *has a spectral gap* (See Page 3 for definition): This fact was proved by Varjü in the appendix of [4], using Theorem 1.2 of [3].

(5) *A density one theorem*: Building on the works of Sarnak [20], Fuchs [7], and Fuchs-Bourgain [1], Bourgain and Kontorovich [4] proved that almost every admissible integer is a curvature, which is a step towards the local-global conjecture.

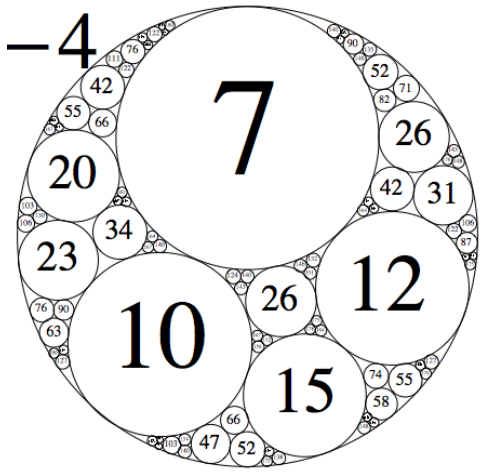

FIGURE 1. An integral Apollonian-3 circle packing

In this paper we generalize the above results to the type of circle packings illustrated in Figure 1. To construct such a packing, we begin with three mutually tangent circles. We iteratively inscribe three circles into curvilinear triangles, and obtain a circle packing, which we call an *Apollonian 3-circle packing*, or Apollonian 3-packing. (By comparison, if we inscribe one circle in each gap, we obtain a standard Apollonian packing.) As shown in Figure 1, there also exist integral Apollonian-3 packings. This was first observed by Guettler-Mallows [11].

We carry over the notations $\mathcal{P}, \mathcal{K}, \Gamma$ to our Apollonian-3 setting. We fix a primitive Apollonian-3 packing $\mathcal{P}$, let $\mathcal{K}$ be the set of curvatures from $\mathcal{P}$, and Γ be the orientation-preserving symmetry group acting on $\mathcal{P}$. We first state a reduction theorem for $\mathcal{P}$.

Theorem 1.1. *(Reduction Theorem) An integer n is admissible by $\mathcal{P}$ if and only if it passes the local obstruction at 8.*

Let $A_{\mathcal{P}}$ be the set of admissible integers of $\mathcal{P}$. In the case of Figure 1,

$$A_{\mathcal{P}} = \{n \in \mathbb{Z} | n \equiv 2, 4, 7 (\text{mod } 8)\}.$$

A general result from Weisfeiler [24] implies the existence of a number Q which completely determines the local obstruction. However in practice it's a hard problem to determine Q. In our case $Q = 8$. Technically, we will prove the following lemma, which directly implies Theorem 1.1. Let $\mathcal{K}_d$ be the reduction of $\mathcal{K}$ (mod d), and ρ_{p^m} be the natural projection from $\mathbb{Z}/p^{m+1}\mathbb{Z}$ to $\mathbb{Z}/p^m\mathbb{Z}$. Write $d = \prod_i p_i^{n_i}$, then we have

Lemma 1.2.
(1) $\mathcal{K}_q \cong \prod_i \mathcal{K}_{p_i^{n_i}}$,
(2) $\mathcal{K}_{p^m} = \mathbb{Z}/p^m\mathbb{Z}$ *for* $p \geq 3$ *and* $m \geq 0$,
(3) $\rho_{2^{m+1}}^{-1}(\mathcal{K}_{2^m}) = \mathcal{K}_{2^{m+1}}$ *for* $p = 2$ *and* $m \geq 3$.

Based on Theorem 1.1, we formulate the following local-global conjecture:

Conjecture 1.3. *(Local-global Conjecture) Every sufficiently large admissible integer from $\mathcal{P}$ is a curvature. Or equivalently,*

$$\#\{n \in \mathcal{K} | n \leq N\} = \#\{n \in A_{\mathcal{P}} | 0 < n \leq N\} + O(1).$$

However, it seems that the current technology is not enough to deal with this conjecture. Instead, we prove a density one theorem:

Theorem 1.4. *(Density One Theorem) There exists $\eta > 0$ such that*

$$\#\{n \in \mathcal{K} | n \leq N\} = \#\{n \in A_{\mathcal{P}} | 0 < n \leq N\} + O(N^{1-\eta}).$$

.

To deduce Theorems 1.1 and 1.4, we need to study the symmetry group $\tilde{\Gamma}$, or more conveniently its orientation preserving subgroup Γ. The group $\tilde{\Gamma} \subset \mathrm{Isom}(\mathbb{H}^3)$ is generated by eight reflections corresponding to eight mutually disjoint hemispheres, and our Apollonian-3 packing can be realized as the limit set of a point orbit under $\tilde{\Gamma}$ (see Figure 2). Therefore Γ is geometric finite. It is clear that $\Gamma\backslash\mathbb{H}^3$ has infinite volume, so Γ is a thin subgroup of $SL(2,\mathbb{C})$. The local structure of Γ will lead to Theorem 1.1. Here we exploit a crucial fact that Γ contains a real congruence subgroup, which is the analogue of Sarnak's observation for the Apollonian group [20]. This congruence subgroup also implies that some curvatures can be represented by certain shifted binary quadratic forms (See Theorem 3.1), which is a key starting point for proving Theorem 1.4.

Another crucial ingredient for Theorem 1.4 is a (geometric) *spectral gap* for Γ, as we explain now. For any positive integer q, Let $\Gamma(q)$ be the principle congruence subgroup of Γ at q (i.e. $\Gamma(q) = \{\gamma \in \Gamma | \gamma \equiv I (\mathrm{mod}\ q)\}$). Let Δ be the hyperbolic Laplacian operator associated to the metric $ds^2 = \frac{dx^2+dy^2+dz^2}{z^2}$ on $\mathbb{H}^3$:

$$\Delta = -z^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}) + z\frac{\partial}{\partial z}$$

The operator Δ is symmetric and positive definite on $L^2(\Gamma(q)\backslash\mathbb{H}^3)$ with the standard inner product. From Larman [16] we know that the Hausdorff dimension δ of our packing $\mathcal{P}$ is > 1. Hence Patterson-Sullivan theory [19][22], together with Lax-Phillips[17] tell us that for each q, there are only finitely many exceptional eigenvalue for Δ acting on $L^2(\Gamma(q)\backslash\mathbb{H}_3)$, and the base (smallest) eigenvalue $\lambda_0(q)$ of Δ on $L^2(\Gamma(q)\backslash\mathbb{H}^3)$ is equal to $\delta(2-\delta)$.

However, a priori the second smallest eigenvalue $\lambda_1(q)$ might get arbitrarily close to $\lambda_0(q)$. But in the case of Γ, this phenomenon does not happen:

Theorem 1.5. *(Spectral Gap) There exists $\delta_0 > 0$ such that for all q,*

$$\lambda_1(q) - \lambda_0(q) \geq \delta_0$$

For the modular group $SL(2,\mathbb{Z})$, the celebrated Selberg $\frac{3}{16}$ Theorem says that $\delta_0 \geq \frac{3}{16}$. For an arbitrary finitely generated subgroup of $SL(2,\mathbb{Z})$, a spectral gap when q is ranging over square free numbers was obtained by Bourgain-Gamburd-Sarnak [3]. Recently this result was extended to much more general groups by Golsefidy-Varjú[8], again over squarefree numbers . But for our need, we need to require q to exhaust all integers.

We then follow the strategy in [4] to prove Theorem 1.4. The main approach is the Hardy-Littlewood circle method. The spectral gap given in Theorem 1.5, together with the bisector counting result from Vinogradov [23], allows us to do various (thin) lattice point counting restricted to certain regions of $SL(2,\mathbb{C})$, effectively and with uniform rates over the

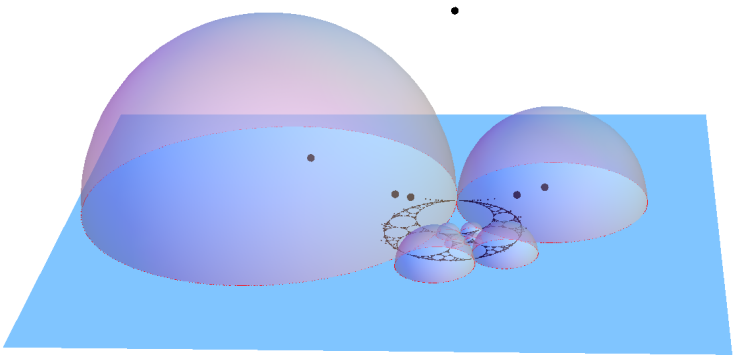

FIGURE 2. The fundamental domain for $\tilde{\Gamma}$ and the orbit of an point under $\tilde{\Gamma}$

congruence towers $\Gamma(q)$ and their cosets. All these are encoded in Lemma 5.2, Lemma 5.3 and Lemma 5.4 from Bourgain and Kontorovich's work on Apollonian packings [4]. These Lemmas can be modified word by word to fit our setting. Another ingredient which appears in the minor arc analysis is the elementary $\frac{3}{4}$ bound for the Kloosterman sums.

Plan for the paper : In §2 we discuss the local properties of $\mathcal{K}$, these properties are revealed by Γ and its subgroups. Theroems 1.1 and 1.5 are proved at the end of this section. The main goal of §3 is to prove Theorem 1.4. In §3.1 we introduce the main exponential sum and give an outline of the proof of Theorem 1.4. In §3.2 we analyze the major arcs, and from §3.3 to §3.5 we give bounds for three parts of the minor-arc integrals. Finally in §3.6 we conclude our proof.

Notation: We adopt the following standard notations. We write $e^{2\pi ix}$ as $e(x)$, and $e^{\frac{2\pi ix}{q}}$ as $e_q(x)$. The relation $f \ll g$ means that $f = O(g)$, and $f \asymp g$ means $f \ll g$ and $g \ll f$. The Greek letter ϵ denotes an arbitrary small positive number, and η denotes a small positive number which appears in several contexts. We assume that each time when η appears, we let η not only satisfy the current claim, but also satisfy the claims in all previous contexts. The symbols p and p_i always denote a prime. The relation $p^j||n$ means $p^j|n$ and $p^{j+1} \nmid n$. The expression $\sum_{r(q)}^{'}$ means sum over all $r(\mathrm{mod}\ q)$ where $(r,q) = 1$. For a finite set Z, its cardinality is denoted by $|Z|$ or $\#Z$. For an algebraic group Γ (or $\mathcal{A}$, $\tilde{\mathcal{A}}$) over $\mathbb{Z}$, $\Gamma(q)$ (or $\mathcal{A}(q)$, $\tilde{\mathcal{A}}(q)$) denotes its principle congruence subgroup of level q. Without further mentioning, all the implied constants depend at most on the given packing.

2. Local Property

2.1. **Apollonian 3-Group and Its Subgroups.** We start with three mutually tangent circles C_1, C_2, C_3 (suppose C_1 is bounding the other two). In each of the two gaps formed by these three circles, there's a unique way to inscribe three more circles, in a way that each of these six circles is tangent to four other circles and disjoint to the last one. Let's say $C_{1'}, C_{2'}, C_{3'}$ is one such inscription (see Figure 3). It is known that their curvatures

$\kappa_1, \kappa_2, \kappa_3, \kappa_{1'}, \kappa_{2'}, \kappa_{3'}$ satisfy the following algebraic relations [11]:

$$\kappa_1 + \kappa_{1'} = \kappa_2 + \kappa_{2'} = \kappa_3 + \kappa_{3'} := 2w \tag{1}$$

$$Q(\kappa_1, \kappa_2, \kappa_3, w) = w^2 - 2w(\kappa_1 + \kappa_2 + \kappa_3) + \kappa_1^2 + \kappa_2^2 + \kappa_3^2 = 0 \tag{2}$$

The Möbius inversion via the dual circle of C_1, C_2, C_3 takes $C_{1'}, C_{2'}, C_{3'}$ to three other circles $C_{1''}, C_{2''}, C_{3''}$, which gives the other way of inscribing. There are two solutions for w in (2), which correspond exactly to two ways of filling.

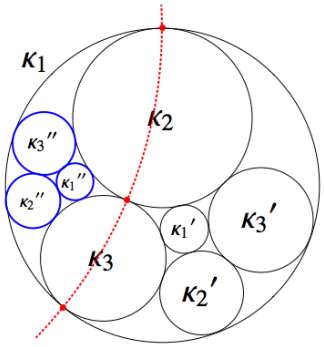

FIGURE 3. Reflection via the dual circle of C_1, C_2, C_3

We associate a quadruple $\mathbf{r} = \langle \kappa_1, \kappa_2, \kappa_3, w \rangle^T$ to the six circles $C_1, C_2, C_3, C_{1'}, C_{2'}, C_{3'}$, which we call the *root circles*. There are eight gaps formed by circular triangles. Each gap corresponds to one Möbius inversion, which takes three of the six root circles to three new circles and fixes the rest three. We associate a vector $\mathbf{v} = \langle x, y, z, w' \rangle^T$ to this new collection of six circles, where x, y, z are the curvatures of the circles which are the images of C_1, C_2, C_3 under the reflection, and w' is the sum of any pair of disjoint circles from this new collection, as w in (1). From (1) and (2) it follows that x, y, z, w' has linear dependance on $\kappa_1, \kappa_2, \kappa_3, w$. Eight gaps correspond to eight linear transformations which take $\mathbf{r}$ to $\mathbf{v}$:

$$S_{123} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & -1 \end{pmatrix}, \qquad S_{1'23} = \begin{pmatrix} -3 & 4 & 4 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 2 & 2 & 3 \end{pmatrix},$$

$$S_{12'3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 4 & -3 & 4 & 4 \\ 0 & 0 & 1 & 0 \\ 2 & -2 & 2 & 3 \end{pmatrix}, \qquad S_{123'} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 4 & 4 & -3 & 4 \\ 2 & 2 & -2 & 3 \end{pmatrix},$$

$$S_{1'2'3} = \begin{pmatrix} -3 & -4 & 4 & 12 \\ -4 & -3 & 4 & 12 \\ 0 & 0 & 1 & 0 \\ -2 & -2 & 2 & 7 \end{pmatrix}, \qquad S_{1'23'} = \begin{pmatrix} -3 & 4 & -4 & 12 \\ 0 & 1 & 0 & 0 \\ -4 & 4 & -3 & 12 \\ -2 & 2 & -2 & 7 \end{pmatrix},$$

$$S_{12'3'} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 4 & -3 & -4 & 12 \\ 4 & -4 & -3 & 12 \\ 2 & -2 & -2 & 7 \end{pmatrix}, \qquad S_{1'2'3'} = \begin{pmatrix} -3 & -4 & -4 & 20 \\ -4 & -3 & -4 & 20 \\ -4 & -4 & -3 & 20 \\ -2 & -2 & -2 & 11 \end{pmatrix}. \tag{3}$$

The subtitles of the above notations keep track of the circles forming the triangular gap. For example, $S_{1'2'3}$ denotes the reflection via the dual circle of $C_{1'}, C_{2'}, C_3$. The group generated by these eight matrices is called Apollonian 3-group, denoted by $\tilde{\mathcal{A}}$:

$$\tilde{\mathcal{A}} = \langle S_{123}, S_{1'23}, S_{12'3}, S_{123'}, S_{1'2'3}, S_{1'2'3}, S_{1'23'}, S_{1'2'3'} \rangle \tag{4}$$

Then we have

$$\mathcal{K} = \{\langle \mathbf{e}_i, \tilde{\mathcal{A}} \cdot \mathbf{r} \rangle | i = 1, 2, 3\} \cup \{\langle \mathbf{e}_i, \tilde{\mathcal{A}} \cdot \mathbf{r}^{'} \rangle | i = 1, 2, 3\} \tag{5}$$

where $\mathbf{r}^{'} = \langle \kappa_{1'}, \kappa_{2'}, \kappa_{3'}, w \rangle$. It then follows that if the initial six circles have integral curvatures, then $\mathcal{P}$ is integral.

In light of (7), we reduce studying $\mathcal{K}$ to studying the group $\tilde{\mathcal{A}}$ which acts on some quadruples containing full information of $\mathcal{K}$. $\mathcal{A}$ is a Coxeter group with the only relations

$$S_{123}^2 = S_{1'23}^2 = \ldots = I.$$

It preserves the quadratic 3-1 form Q, so $\tilde{A}a \subseteq O_Q(\mathbb{Z})$. Furthermore, we pass to its orientation-preserving subgroup $\mathcal{A} = \mathcal{A} \cap SO_Q(\mathbb{Z})$, which is an index-2 subgroup of $\tilde{\mathcal{A}}$ and a free group generated by

$$S_{123}S_{1'23}, S_{123}S_{12'3}, S_{123}S_{123'}, S_{123}S_{1'2'3}, S_{123}S_{1'23'}, S_{123}S_{12'3'}, S_{123}S_{1'2'3'}. \tag{6}$$

From (5) we also have

$$\mathcal{K} = \{\langle \mathbf{e}_i, \mathcal{A} \cdot \mathbf{r} \rangle | i = 1, 2, 3\} \cup \{\langle \mathbf{e}_i, \mathcal{A} \cdot \mathbf{r}^{'} \rangle | i = 1, 2, 3\} \tag{7}$$

This is because if a word from $\tilde{\mathcal{A}}$ consists of odd number of reflections, we can always pre-add S_{123} (or $S_{1'2'3'}$) without changing $\mathbf{r}$ (or $\mathbf{r}^{'}$). The augmented word is even, thus lies in $\mathcal{A}$.

Recall the spin homomorphism $\rho_0 : SL(2, \mathbb{C}) \longrightarrow SO_{Q_0}$, where $\tilde{Q}_0(x, y, z, t) = t^2 - x^2 - y^2 - z^2$ is the standard $3 - 1$ form (see [6]):

$$\rho_0\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} \Re(a\bar{d} + b\bar{c}) & \Im(a\bar{d} - b\bar{c}) & \Re(-a\bar{c} + b\bar{d}) & \Re(a\bar{c} + b\bar{d}) \\ \Im(-a\bar{d} - b\bar{c}) & \Re(a\bar{d} - b\bar{c}) & \Im(a\bar{c} - b\bar{d}) & \Im(-a\bar{c} - b\bar{d}) \\ \Re(-a\bar{b} + c\bar{d}) & \Im(-a\bar{b} + c\bar{d}) & \frac{|a|^2 - |b|^2 - |c|^2 + |d|^2}{2} & \frac{-|a|^2 - |b|^2 + |c|^2 + |d|^2}{2} \\ \Re(a\bar{b} + c\bar{d}) & \Im(a\bar{b} + c\bar{d}) & \frac{-|a|^2 + |b|^2 - |c|^2 + |d|^2}{2} & \frac{|a|^2 + |b|^2 + |c|^2 + |d|^2}{2} \end{pmatrix} \tag{8}$$

The isomorphism between SO_{Q_0} and SO_Q is given by

$$A \longrightarrow J^{-1}AJ,$$

where

$$J = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & \sqrt{2} \end{pmatrix}$$

The spin homomorphism that we use is ρ, defined from $SL(2, \mathbb{C})$ to $SO_{\tilde{Q}}$ as

$$\rho(\gamma) = J^{-1}\rho_0\left(\begin{pmatrix} 1+i & -\sqrt{2} \\ \sqrt{2} & 1+i \end{pmatrix} \gamma \begin{pmatrix} 1+i & -\sqrt{2} \\ \sqrt{2} & 1+i \end{pmatrix}^{-1}\right) J \tag{9}$$

The good thing about conjugating γ with $\begin{pmatrix} 1+i & -\sqrt{2} \\ \sqrt{2} & 1+i \end{pmatrix}$ is that the preimage of the generators in (6) is

$$M_1 = \begin{pmatrix} 1 & 2 \\ -2 & -3 \end{pmatrix}, M_2 = \begin{pmatrix} 1-2\sqrt{2}i & 2 \\ 2+4\sqrt{2}i & -3+2\sqrt{2}i \end{pmatrix}, M_3 = \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}, M_4 = \begin{pmatrix} -1+2\sqrt{2}i & -4 \\ -4\sqrt{2}i & 7-2\sqrt{2}i \end{pmatrix}$$
$$M_5 = \begin{pmatrix} -1 & 2 \\ 2 & -5 \end{pmatrix}, M_6 = \begin{pmatrix} 1+2\sqrt{2}i & -2 \\ -6-4\sqrt{2}i & 5-2\sqrt{2}i \end{pmatrix}, M_7 = \begin{pmatrix} -1-2\sqrt{2}i & 4 \\ 4+4\sqrt{2}i & -9+2\sqrt{2}i \end{pmatrix}, \tag{10}$$

which all lie in $SL(2, \mathbb{Z}[\sqrt{2}i])$, and we let $\Gamma = \langle M_1, M_2, M_3, M_4, M_5, M_6, M_7 \rangle$.

If we write $a = a_1 + a_2\mathbf{i}, b = b_1 + b_2\mathbf{i}, c = c_1 + c_2\mathbf{i}, d = d_1 + d_2\mathbf{i}$, one can verify (with the aid of computer) that ρ maps the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to a 4×4 matrix, each entry of which is a homogenous quadratic polynomial of $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$, with half-integer coefficients. Therefore, ρ can descend to a homomorphism from $\Gamma/\Gamma(q)$ to $A/A(q)$ for any q that does not contain a power of 2.

The group Γ contains a real subgroup $\Gamma_{C_3} = \langle M_1, M_3, M_5 \rangle$. Geometrically, Γ_{C_3} fixes the circle C_3. It turns out that Γ_{C_3} is a congruence subgroup:

Proposition 2.1. *The group Γ is a congruence subgroup of level 4. Explicitly,*

$$\Gamma_{C_3} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) | a \equiv d \equiv 1 (\text{mod } 2), \text{b} \equiv \text{c} \equiv 0 \text{ or } 2 (\text{mod } 4) \right\} \tag{11}$$

Proof. We notice that the $\subseteq$ direction is straightforward, then we can prove the proposition by explicitly constructing the fundamental domain (See Figure 4). Indeed once we show that the fundamental domain of Γ_{C_3} is as shown in Figure 4, we can compute the covolume of Γ_{C_3} to be 8π, which coincides with the covolume of the group described by the righthand side of (11), thus the proposition is established.

First we replace the generators M_1, M_3, M_5 of Γ_{C_3} by three parabolic generators $M_1, M_3^{-1} = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}, M_3^{-1}M_5 = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}$ which fix -1,0,1 respectively. We denote the corresponding parabolic subgroups by B_{-1}, B_0, B_1. We have $M_1(\infty) = -\frac{1}{2}, M_3^{-1}(-\frac{1}{2}) = \frac{1}{2}$ and $M_3^{-1}M_5(\frac{1}{2}) = \infty$. It turns out that the open region $\mathcal{F}_{C_3}$ bounded by the closed loop $\infty \to -1 \to -\frac{1}{2} \to 0 \to \frac{1}{2} \to 1 \to \infty$ is the fundamental domain for Γ_{C_3}.

Associate the open regions I, II, III (See Figure 4) to B_{-1}, B_0, B_1, then B_{-1} maps II, III to I, B_0 maps I, III to II and B_1 maps I, II to III. We can apply the Pingpong Lemma to show that Γ is freely generated by these three elements. To show that $\mathcal{F}_{C_3}$ is a fundamental domain, one needs to show

(i)$\gamma(\mathcal{F}_{C_3}) \cap \mathcal{F}_{C_3} = \emptyset$ if $\gamma \neq I$.

(ii)$\overline{\Gamma_{C_3}(\mathcal{F}_{C_3})} = \mathbb{H}$.

For (i), first write $\gamma = T_1T_2\cdots T_m$, where each T_i comes from one of the parabolic subgroups B_{-1}, B_0 or B_1. We say the *length* of this word is m. We assume the length of the word is minimal so that T_i, T_{i+1} are not in a same parabolic subgroup. Then one can prove that $\gamma(\mathcal{F}_{C_3})$ lies in one of the regions from I, II, III, which is determined by T_1. Since I, II, III are disjoint from $\mathcal{F}_{C_3}$, (i) is thus proved.

For (ii), suppose $z \in \overline{\Gamma_{C_3}(\mathcal{F}_{C_3})} = \mathbb{H}$, we want to show that z lies also in the interior of $\overline{\Gamma(\mathcal{F}_{C_3})}$. First one can check that for each side of $\mathcal{F}_{C_3}$, there's one element γ from $M_1, M_1^{-1}, M_3, M_3^{-1}, M_3M_5^{-1}, M_3^{-1}M_5$ such that $\gamma(\mathcal{F}_{C_3})$ and $\mathcal{F}_{C_3}$ share this given side. Now we place a ball of radius ϵ sitting at each of the cusps $-1, 0, 1$ and we say the complement of these balls to $\mathcal{F}_{C_3}$ the *compact part* of $\mathcal{F}_{C_3}$, denoted by $\mathcal{F}_{C_3}^{c,\epsilon}$. We define the compact part of $\gamma(\mathcal{F}_{C_3})$ simply by $\gamma(\mathcal{F}_{C_3}^{c,\epsilon})$. Then there exists a universal constant $l(\epsilon)$ such that if z lies within the $l(\epsilon)$ distance of some $\gamma(\mathcal{F}_{C_3}^{c,\epsilon})$, then z lies either within $\gamma(\mathcal{F}_{C_3})$ itself, or some $\gamma'(\mathcal{F}_{C_3})$ next to $\gamma(\mathcal{F}_{C_3})$, or on the common boundary of these two domains. In both cases z is an inner point of $\overline{\Gamma_{C_3}(\mathcal{F}_{C_3})}$.

It's an elementary geometric exercise to check that any $\gamma \in \Gamma_{C_3}$ will send these ϵ-balls to balls with radii no greater than ϵ (by induction on the minimal length of word). This means that if we choose $\epsilon = \frac{\mathrm{Im}z}{10}$ and some $l(\epsilon) < \frac{\mathrm{Im}z}{10}$, and some γ_ϵ such that $d(\gamma_\epsilon(\mathcal{F}_{C_3})), z) < \min\{\frac{\mathrm{Im}z}{10}, l(\epsilon)\}$. In other words, z is very close to the compact part of the fundamental domain $\gamma_\epsilon(\mathcal{F}_{C_3})$. Therefore z is an inner point. Since $\overline{\Gamma(\mathcal{F}_{C_3})}$ is both open and closed, $\overline{\Gamma(\mathcal{F}_{C_3})} = \mathbb{H}$. □

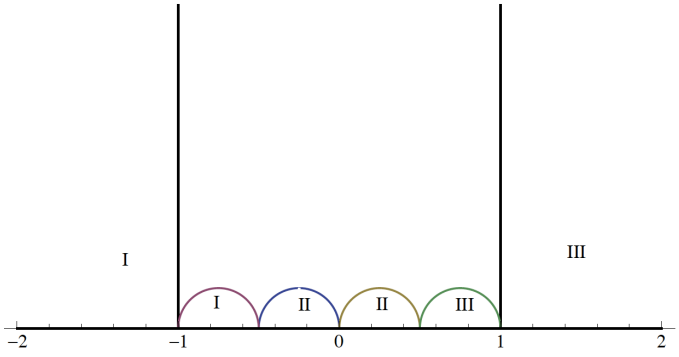

FIGURE 4. The fundamental domain for Γ_{C_3}

Conjugating Γ_{C_3} by $\begin{pmatrix} 1 & 0 \\ \sqrt{2}i & 1 \end{pmatrix}$, one gets $\Gamma_{C_1} = \begin{pmatrix} 1 & 0 \\ \sqrt{2}i & 1 \end{pmatrix} \Gamma_{C_3} \begin{pmatrix} 1 & 0 \\ -\sqrt{2}i & 1 \end{pmatrix} = \langle M_2, M_3, M_6 \rangle$, which is a subgroup of Γ fixing C_1. Similarly,

$$\Gamma_{C_{3'}} = \begin{pmatrix} -1 & 1+\sqrt{2}i \\ -1 & -1+\sqrt{2}i \end{pmatrix} \Gamma_{C_3} \begin{pmatrix} -1 & 1+\sqrt{2}i \\ -1 & -1+\sqrt{2}i \end{pmatrix}^{-1} =< M_7^{-1}M_3, M_7^{-1}M_5, M_7^{-1}M_6 >,$$

which is a subgroup fixing $C_{3'}$.

Let

$$A_k(q) = \{g_1h_1j_1 \dots g_kh_kj_k : g_1, \dots, g_k \in \Gamma_{C_3}, h_1, \dots, h_k \in \Gamma_{C_1}, j_1, \dots, j_k \in \Gamma_{C_{3'}}\} \tag{12}$$

We have the following proposition:

Proposition 2.2. *Let $q = \prod_i p_i^{n_i}$, then $\Gamma/\Gamma(q) = A_{10^9}(q)$.*

Before proving Proposition 2.2, we prove a few lemmas first.

Lemma 2.3. *If $p \geq 5$, then $A_{54}(p^m) = \Gamma/\Gamma(p^m)$.*

Proof. Since Γ_{C_1} is a congruence subgroup of level 4, we have $\Gamma_{C_1}/\Gamma_{C_1}(p^m) = SL(2, \mathbb{Z}/p^m\mathbb{Z})$. We also have

$$\begin{pmatrix} b^{-1} & 0 \\ 0 & b \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{2} & 0 \\ -\frac{7}{4} & -2 \end{pmatrix} \cdot M_2^2 \cdot \begin{pmatrix} 1 & 0 \\ -\frac{1}{4} & 1 \end{pmatrix} \cdot M_2^{-1} \cdot \begin{pmatrix} b & 0 \\ 0 & b^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3\sqrt{2}b^2 i & 1 \end{pmatrix}.$$

Now we show that $\forall M > 1$, we can find at most four elements $a, b, c, d \in \mathbb{Z}/p^m\mathbb{Z})$ such that

$$a^2 + b^2 + c^2 + d^2 \equiv M (\text{mod } p^m)$$

This is true for $m = 1$ by the Lagrange's Four Square Theorem, which states that every integer can be written as a sum of at most four squares of integers. Choose $M^{'} \equiv M(p)$ with $0 < M^{'} \leq p$, then we can choose $a^{'}, b^{'}, c^{'}, d^{'}$ such that

$$a^{'2} + b^{'2} + c^{'2} + d^{'2} = M^{'}, \tag{13}$$

Necessarily all $a^{'}, b^{'}, c^{'}, d^{'}$ have to be strictly less than p, and at least one of them is not zero, thus invertible in $\mathbb{Z}/p\mathbb{Z}$. So when mod p, $(a^{'}, b^{'}, c^{'}, d^{'})$ is a regular point on the curve

$$x^2 + y^2 + z^2 + w^2 \equiv M^{'} (\text{mod } p). \tag{14}$$

The general case follows from Hensel's lemma by lifting the solution $(a^{'}, b^{'}, c^{'}, d^{'})$ of (14) to a solution (a, b, c, d) of

$$x^2 + y^2 + z^2 + w^2 = M (\text{mod } p^m)$$

This shows that

$$\begin{pmatrix} 1 & 0 \\ a\sqrt{2}i & 1 \end{pmatrix} \in A_9(p^m)$$

Multiplying the above matrix by $\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}, b \in \mathbb{Z}/(p^m)$, which can be found in Γ_{C_1} since it contains $\begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}$, we have

$$\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \in A_9(p^m)$$

for any $c \in \mathbb{Z}[\sqrt{2}i]/(p^m)$. Conjugating the above element by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, which is also congruent to some element in $\Gamma_{C_3} (\text{mod } p^m)$, we have

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \in A_{12}(p^m)$$

for any $c \in \mathbb{Z}[\sqrt{2}i]/(p^m)$. Now

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+ab & a+c+abc \\ b & 1+bc \end{pmatrix}.$$

This shows that

$$\begin{pmatrix} a^{'} & b^{'} \\ c^{'} & d^{'} \end{pmatrix} \in \Gamma/\Gamma(p^m)$$

for any $c^{'}$ invertible in $\mathbb{Z}[\sqrt{2}i]/(p^m)$ and $a^{'}d^{'} - b^{'}c^{'} = 1$. There are $p^{3m-1}(p-1)$ such elements. The size of $SL(2, \mathbb{Z}[\sqrt{2}i]/(p^m))$ is $p^{3m-3}(p-1)(p^2-1)$, which is strictly less than twice of $p^{3m-1}(p-1)$, this means that $A_{54}(p^m)$ has to be full of the group $SL(2, \mathbb{Z}[\sqrt{2}i]/(p^m))$. □

Lemma 2.4. *$A_{10^7}(2^m) = \Gamma/\Gamma(2^m)$, and $A_{10^7}(3^m) = \Gamma/\Gamma(3^m)$*

Proof. We prove the case when $p = 2$ and explain the difference when $p = 3$. For $p = 2$, we first prove the following claim by induction:

Claim: For every $m \geq 6$ and $g \in \Gamma(2^6)/\Gamma(2^m)$, we can find $g_1, g_2, g_3 \in \Gamma_{C_3}(2^3)/\Gamma_{C_3}(2^m)$ such that

$$g = g_1 M_2 g_2 M_2^{-1} M_2^2 g_3 M_2^{-2}.$$

For $m = 6$ we can choose $g_1 = g_2 = g_3 = 1$. For $m > 6$, we now assume this holds for $m-1$. By the induction hypothesis, there exists $h_1, h_2, h_3 \in \Gamma(2^3)$ such that

$$g = h_1 M_2 h_2 M_2^{-1} M_2^2 h_3 M_2^{-2} + 2^{m-1}x (\text{mod } 2^m)$$

Now we choose some $x_i \in \text{Mat}(2, \mathbb{Z})$ such that $x_i \equiv 0(2^{m-3})$ and $\text{tr}(x_i) \equiv 0 (\text{mod } 2^m)$ for $i = 1, 2, 3$. We have

$$\begin{aligned} g \equiv (h_1 + x_1) M_2 (h_2 + x_2) M_2^{-1} M_2^2 (h_2 + x_3) m_2^{-2} \\ - (x_1 + M_2 x_2 M_2^{-1} + M_2^2 x_2 M_2^{-1}) + 2^{m-1}x (\text{mod } 2^m) \end{aligned}$$

Since $\text{Det}(x_i + h_i) = 1(2^m)$ and $x_i + h_i \equiv I(2^3)$, $x_i + h_i$ is congruent to some element $g_i \in \Gamma_{C_3} (\text{mod } 2^m)$ using the congruence property of Γ_{C_3}. The matrices x_1, x_2, x_3 can be chosen as a suitable linear combination of the matrices in the following calculations to cancel the term $2^{m-1}x$:

$$2^{m-1}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + M_2 0 M_2^{-1} + m_2^2 0 M_2^{-2} \equiv 2^{m-1}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} (\text{mod } 2^m)$$

$$2^{m-1}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + M_2 0 M_2^{-1} + m_2^2 0 M_2^{-2} \equiv 2^{m-1}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} (\text{mod } 2^m)$$

$$2^{m-1}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + M_2 0 M_2^{-1} + m_2^2 0 M_2^{-2} \equiv 2^{m-1}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} (\text{mod } 2^m)$$

$$2^{m-3}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix} + M_2 2^{m-3}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 2^{m-1}\begin{pmatrix} 0 & \sqrt{2}i \\ 0 & 0 \end{pmatrix} (\text{mod } 2^m)$$

$$2^{m-3}\begin{pmatrix} -2 & 4 \\ -1 & 2 \end{pmatrix} + M_2 2^{m-3}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 2^{m-1}\begin{pmatrix} \sqrt{2}i & 0 \\ \sqrt{2}i & -\sqrt{2}i \end{pmatrix} (\text{mod } 2^m)$$

$$2^{m-3}\begin{pmatrix} 4 & 0 \\ 1 & 4 \end{pmatrix} + M_2 0 M_2^{-1} + M_2^2 2^{m-3}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} M_2^{-2} \equiv 2^{m-1}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} (\text{mod } 2^m)$$

Thus we showed that

$$A_3(2^m) \supseteq \Gamma(2^6)/\Gamma(2^m).$$

Now since the index of $\Gamma(2^6)/\Gamma(2^m)$ in $\Gamma/\Gamma(2^m)$ is $|\Gamma/\Gamma(2^6)| = 2^{26}$, this implies that

$$A_{10^7}(2^m) = \Gamma/\Gamma(2^m) \tag{15}$$

For the case $p = 3$, the proof goes in the same way. We choose the linear combinations of the following:

$$3^{m-1}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} + M_2 0 M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 3^{m-1}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} \pmod{3^m}$$

$$3^{m-1}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} + M_2 0 M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 3^{m-1}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} \pmod{3^m}$$

$$3^{m-1}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} + M_2 0 M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 3^{m-1}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \pmod{3^m}$$

$$3^{m-1}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} + M_2 3^{m-1}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 3^{m-1}\begin{pmatrix}0 & -\sqrt{2}i\\ -\sqrt{2}i & 0\end{pmatrix} \pmod{3^m}$$

$$3^{m-1}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} + M_2 3^{m-1}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} M_2^{-1} + M_2^2 0 M_2^{-2} \equiv 3^{m-1}\begin{pmatrix}\sqrt{2}i & 0\\ 0 & -\sqrt{2}i\end{pmatrix} \pmod{3^m}$$

$$3^{m-1}\begin{pmatrix}-1 & 1\\ 1 & 1\end{pmatrix} + M_2 0 M_2^{-1} + M_2^2 3^{m-1}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} M_2^{-2} \equiv 3^{m-1}\begin{pmatrix}\sqrt{2}i & \\ -\sqrt{2}i & -\sqrt{2}i\end{pmatrix} \pmod{3^m}$$

The constant 10^7 also works in this case. □

Now we are able to prove Proposition 2.2.

Proof of Proposition 2.2. First we embed $\Gamma/\Gamma(d)$ into $\prod_{p_i^{m_i}||d} \Gamma/\Gamma(p^m)$. For any $x \in \prod_{p_i^{m_i}||} \Gamma/\Gamma(p^m)$, from Lemma 2.3 and Lemma 2.4, we can write

$$x \equiv \prod_{j=1}^{10^7} \gamma_{j,C_3}^{(i)} . \gamma_{j,C_1}^{(i)} . \gamma_{j,C_{3'}}^{(i)} (p_i^{m_i})$$

for each i, where $\gamma_{j,C_3}^{(i)} \in \Gamma_{C_3}, \gamma_{j,C_1}^{(i)} \in \Gamma_{C_1}, \gamma_{j,C_{3'}}^{(i)} \in \Gamma_{C_{3'}}$. Since Γ_{C_3} is a congruence subgroup and $\Gamma_{C_1}, \Gamma_{C_{3'}}$ are conjugate to Γ_{C_1}, we can find $\gamma_1, \gamma_2, \gamma_3$ such that

$$\gamma_1 \equiv \gamma_{j,C_3}^i (\text{mod } p_i^{m_i})$$
$$\gamma_2 \equiv \gamma_{j,C_1}^i (\text{mod } p_i^{m_i})$$
$$\gamma_3 \equiv \gamma_{j,C_{3'}}^i (\text{mod } p_i^{m_i})$$

for each i. So $x = \gamma_1\gamma_2\gamma_3 \in \Gamma/\Gamma(d)$. So we have

$$\prod_{p^m||d} \Gamma/\Gamma(p^m) = \Gamma/\Gamma(d)$$

□

From the above proposition, it follows directly that

Lemma 2.5.
(1) If $q=\prod_i p_i^{m_i}$, *then* $\Gamma/\Gamma(q) \cong \prod_i \Gamma/\Gamma(p_i^{m_i})$,
(2) If $(q,6)=1$, *then* $\Gamma/\Gamma(q) = SL(2,(\mathbb{Z}[\sqrt{2}i]/(q)))$.
(3) If $l \geq 3$, *then the kernel of* $\Gamma/\Gamma(2^l) \longrightarrow \Gamma/\Gamma(8)$ *is the full of the kernel of* $SL(2,\mathbb{Z}[\sqrt{2}i]/(2^l)) \longrightarrow SL(2,\mathbb{Z}[\sqrt{2}i]/(8))$; *If* $l \geq 1$, *then the kernel of* $\Gamma/\Gamma(3^l) \longrightarrow \Gamma/\Gamma(3)$ *is the full of the kernel of* $SL(2,\mathbb{Z}[\sqrt{2}i]/(3^l)) \longrightarrow SL(2,\mathbb{Z}[\sqrt{2}i]/(3))$.

Since $\rho : SL(2,\mathbb{Z}[\sqrt{2}i]/(p_i^{m_i})) \longrightarrow SO_Q(\mathbb{Z}/p_i^{m_i}\mathbb{Z})$ is surjective for each i, the above theorem also holds for $\mathcal{A}$. We state it here:

Lemma 2.6.
(1) If $q=\prod_i p_i^{n_i}$, *then* $\mathcal{A}/\mathcal{A}(q) \cong \prod_i \mathcal{A}/\mathcal{A}(p_i^{n_i})$,
(2) If (q,6)=1, then $\mathcal{A}/\mathcal{A}(q) = SO_Q(\mathbb{Z}/q\mathbb{Z})$.
(3) If $l \geq 3$, *then the kernel of* $\mathcal{A}/\mathcal{A}(2^l) \longrightarrow \mathcal{A}/\mathcal{A}(8)$ *is the full of the kernel of* $SO_Q(\mathbb{Z}/2^l\mathbb{Z}) \longrightarrow SO_Q(\mathbb{Z}/8\mathbb{Z})$; *If* $l \geq 1$, *then the kernel of* $\mathcal{A}/\mathcal{A}(3^l) \longrightarrow \mathcal{A}/\mathcal{A}(3)$ *is the full of the kernel of* $SO_Q(\mathbb{Z}/3^l\mathbb{Z}) \longrightarrow SO_Q(\mathbb{Z}/3\mathbb{Z})$.

Now we can study the local obstruction of $\mathcal{P}$. We let V be the set of vectors $\Gamma \cdot \mathbf{r}$ and V_d be the reduction of $V (\text{mod } d)$. We define C_{p^m} as follows:

- if $p \geq 3$,
$$C_{p^m} = \{\mathbf{v} \in (\mathbb{Z}/p^m\mathbb{Z})^4 | Q(\mathbf{v}) \equiv 0 (\text{mod } p^m)\}$$
- if p=2,
$$C_{2^m} = \{\mathbf{v} \in (\mathbb{Z}/2^m\mathbb{Z})^4 | Q(\mathbf{v}) \equiv 0 (\text{mod } 2^m), \exists \mathbf{w} \equiv \mathbf{v}(2^m), Q(\mathbf{w}) \equiv 0 (\text{mod } 2^{m+1})\}$$

Let
$$\pi_{p^m} : C_{p^{m+1}} \longrightarrow C_{p^m}$$
be the canonical projection. We have following lemmata:

Lemma 2.7. *If* $p \geq 5$, *then*
$$V_{p^m} = C_{p^m}$$

Proof. This follows from Lemma 2.6, and the fact that $SO_Q(\mathbb{Z}/p^m\mathbb{Z})$ acts transitively on C_{p^m}. □

When $p = 2, 3$, the argument in Lemma 2.7 does not work because Γ reduced at these local places is not the full group of SL_2. But in each case the lifting will saturate for some finite m, as shown in Lemma 2.8 and 2.9. In the case $p = 3$, $\Gamma(\mathbb{Z}[\sqrt{2}i]/(3^m))$ is actually big enough to make $V_{3^m} = C_{3^m}$:

Lemma 2.8. *If* $p = 3$, *then*
$$V_{3^m} = C_{3^m}.$$

Proof. Using a program, we can check that $|V_3| = |C_3| = 27$, and moreover, there exist $T_1, \ldots, T_{27} \in \mathcal{A} \cap SO_Q(\mathbb{Z})(3)$ such that all the solutions of $Q(\mathbf{v}) \equiv 0 (\text{mod } 9)$ lying above $\mathbf{r}$ is given by:
$$T_1(\mathbf{r}) = \mathbf{r} + (T_1 - I)\mathbf{r} \ (\text{mod } 3)$$
$$\vdots$$
$$T_{27}(\mathbf{r}) = \mathbf{r} + (T_{27} - I)\mathbf{r} \ (\text{mod } 3).$$

Then for any $m \geq 0$ the liftings from V_{3^m} to $V_{3^{m+1}}$ are given by

$$T_1^{3^m}(\mathbf{r}) = \mathbf{r} + (T_1 - I)^{3^m}\mathbf{r} \pmod{3^{m+1}}$$
$$\vdots$$
$$T_{27}^{3^m}(\mathbf{r}) = \mathbf{r} + (T_{27} - I)^{3^m}\mathbf{r} \pmod{3^{m+1}}$$

We find that $|V_{3^m}| = |C_{3^m}|$. □

Lemma 2.9. *If $p = 2$, then for $m \geq 3$,*

$$\pi_{2^{m+1}}^{-1}(V_{2^m}) = V_{2^{m+1}}$$

Proof. We prove this by effective lifting. This argument is due to Fuchs [7]. For $n \geq 3$, let $W(m) = (S_{1'23}.S_{1'2'3})^{2^{m-3}}, X(m) = (S_{12'3}.S_{12'3'})^{2^{m-3}}, Y(m) = (S_{123'}.S_{1'2'3})^{2^{m-4}}$. Then

$$W(n) = \begin{pmatrix} 1 & 0 & 0 & 2^{m-1} \\ 2^{m-1} & 1+2^{m-1} & 2^{m-1} & 2^{m-1} \\ 0 & 0 & 1 & 0 \\ 0 & 2^{m-1} & 0 & 1+2^{m-1} \end{pmatrix},$$

$$X(m) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2^{m-1} & 2^{m-1} & 1+2^{m-1} & 2^{m-1} \\ 0 & 0 & 2^{m-1} & 1+2^{m-1} \end{pmatrix},$$

$$Y(m) = \begin{pmatrix} 1-2^{m-2} & -2^{m-2} & 2^{m-2} & -2^{m-2} \\ -2^{m-2} & 1-2^{m-2} & 2^{m-2} & -2^{m-2} \\ 2^{m-2} & -2^{m-2} & 1+2^{m-2} & -2^{m-2} \\ -2^{m-2} & -2^{m-2} & 2^{m-2} & 1+2^{m-2} \end{pmatrix}.$$

Then for example if $\mathbf{r} \equiv \langle 3,2,2,3\rangle (\text{mod } 4)$, then

$$I\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 0,0,0,0\rangle (\text{mod } 2^m)$$
$$W(m)\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 1,0,0,1\rangle (\text{mod } 2^m)$$
$$X(m)\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 0,0,0,1\rangle (\text{mod } 2^m)$$
$$Y(m)\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 1,1,1,0\rangle (\text{mod } 2^m)$$
$$W(m)X(m)\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 1,0,0,0\rangle (\text{mod } 2^m)$$
$$W(m)Y(m)I\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 0,1,1,1\rangle (\text{mod } 2^m)$$
$$X(m)Y(m)\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 1,1,1,1\rangle (\text{mod } 2^m)$$
$$W(m)X(m)Y(m)\mathbf{r} \equiv \mathbf{r} + 2^{m-1}\langle 0,1,1,0\rangle (\text{mod } 2^m)$$

□

Collecting the result from Lemma 2.7 to Lemma 2.9, we obtain the following proposition which describes the local structure of V.

Theorem 2.10.
(1) $V_q \cong \prod_i V_{p_i^{n_i}}$,
(2) $\pi_{p^{m+1}}^{-1}(V_{p^m}) = V_{p^{m+1}}$ *for* $p \geq 3$ *and* $m \geq 0$,
(3) $\pi_{2^{m+1}}^{-1}(V_{2^m}) = V_{2^{m+1}}$ *for* $p = 2$ *and* $m \geq 3$.

Lemma 1.2, thus Theorem 1.1 then follow directly from Theorem 2.10 because the first three components of V are curvatures.

Now we prove Theorem 1.5. Bourgain, Gamburd and Sarnak [3] established an equivalence between a geometric spectral gap and a combinatorial spectral gap for a finitely generated Fuchsian group F. Let S be a finite symmetric ($S = S^{-1}$) generating set of F. For each q, we have a Cayley graph of $F/F(q)$ over S. There's a Markov operator (which is a discrete version of Laplacian) on the functions of this Cayley graph. A Combinatorial spectral gap is then a uniform positive lower bound of the distance between the biggest two eigenvalues $\lambda_0^{'}(q) = 1$ and $\lambda_1^{'}(F(q), S)$ of this operator. Later this equivalence is generalized by Kim [13] to Kleinian groups, which applies to our case Γ. From the celebrated Selberg's $\frac{3}{16}$ theorem we know there are geometric spectral gaps for $\Gamma_{C_3}, \Gamma_{C_1}, \Gamma_{C_{3'}}$. It then follows that the combinatorial gaps exist for these groups from [2]. Now we apply Varjü's lemma in the Appendix of [4]:

Lemma 2.11 (Varjü)**.** *Let G be a finite group and $S \subset G$ a finite symmetric generating set. Let $G_1, G_2, ..., G_k$ be subgroups of G such that for every $g \in G$ there are $g_1 \in G_1, \ldots, g_k \in G_k$ such that $g = g_1 \ldots g_k$. Then*

$$1 - \lambda_1^{'}(G, S) \geq \min_{1 \leq i \leq k} \left\{ \frac{|S \cap G_i|}{|S|} . \frac{1 - \lambda_1^{'}(G_i, S \cap G_i)}{2k^2} \right\}$$

In our case G is $\Gamma(\text{mod } q)$, G_i's are $\Gamma_{C_3}, \Gamma_{C_1}$ or $\Gamma_{C_{3'}}(\text{mod } q)$, in light of Proposition 2.2. And we let S to be the union of $M_1, M_2, M_3, M_4, M_5, M_6, M_7^{-1}M_3$ and their inverses. Clearly Lemma 2.11 provides a spectral gap for Γ. This implies a geometric spectral gap for Γ again by [3].

3. Circle Method

In this chapter we are proving Theorem 1.4 via the Hardy-Littlewood circle method. In §3.1 we set up the ensemble for the circle method. In §3.2 we do major arc analysis, where we crucially use the spectral gap property of Γ for several counts. From §3.3 to §3.5 we do minor arc analysis, and several Kloosterman-type sums naturally appear here. §3.6 gathers all the previous results and finishes the proof of Theorem 1.4.

3.1. Setup of the circle method.

Recall that Γ_{C_3} is a congruence subgroup

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}) | a \equiv d \equiv 1(\text{mod } 2), b \equiv c \equiv 0 \text{ or } 2(\text{mod } 4) \right\}.$$

Therefore, for any $x, y \in \mathbb{Z}$ with $(x, 2y) = 1$, we can find an element $\xi_{x,y}$ of the form $\begin{pmatrix} x & 2y \\ * & * \end{pmatrix} \in \Gamma_{C_3}$. Under the spin homomorphism ρ, $\xi_{x,y}$ will be mapped to

$$\begin{pmatrix} x^2 - y^2 & -1 + x^2 + y^2 & 2xy & -2xy + 2y^2 \\ 0 & 1 & 0 & 0 \\ * & * & * & * \\ * & * & * & * \end{pmatrix}.$$

Hence we have the following theorem:

Theorem 3.1. *Let $x, y \in \mathbb{Z}$ with $(x, 2y) = 1$, and take any element $\gamma \in \mathcal{A}$ with the corresponding quadruple*

$$\mathbf{v}_\gamma = \gamma(\mathbf{r}) = \langle a_\gamma, b_\gamma, c_\gamma, d_\gamma \rangle.$$

Then the number

$$\langle \mathbf{e_1}, \xi_{x,y}.\gamma\mathbf{r} \rangle = A_\gamma x^2 + 2B_\gamma xy + C_\gamma y^2 - b_\gamma \tag{16}$$

is the curvature of some circle in $\mathcal{P}$, where

$$\begin{aligned} A_\gamma &:= a_\gamma + b_\gamma \\ B_\gamma &:= c_\gamma - d_\gamma \\ C_\gamma &:= -a_\gamma + b_\gamma + 2d_\gamma \end{aligned} \tag{17}$$

We can view (16) as a shifted quadratic form $\mathfrak{f}(x, y)$ determined by γ with variables x, y. We define

$$\begin{aligned} \mathfrak{f}(x, 2y) &= \langle \mathbf{e}_1, \xi_{x,y}.\gamma(\mathbf{r}) \rangle \\ \tilde{\mathfrak{f}}(x, 2y) &= A_\gamma x^2 + 2B_\gamma xy + C_\gamma y^2 \end{aligned}$$

Then $\mathfrak{f} = \tilde{\mathfrak{f}} - b_\gamma$, and the discriminant of $\tilde{\mathfrak{f}}$ is $-8b_\gamma^2$.

Now we set up our ensemble for the circle method. Let N be the main growing parameter. Write $N = TX^2$, where $T = N^{\frac{1}{200}}$, a small power of N, and $X = N^{\frac{199}{400}}$. We define our ensemble to be a subset of $\mathcal{A}$ (with multiplicity) of Frobenius norm $\asymp N$. The ensemble is a product of a subset $\mathfrak{F}$ of norm T, and a subset $\mathcal{X}$ of norm X^2. We further write $T = T_1T_2$, where $T_2 = T_1^{\mathcal{C}}$ and $\mathcal{C}$ is a large number which is determined in Lemma 3.11 . We define $\mathfrak{F}$ in the following way:

$$\mathfrak{F} = \mathfrak{F}_T = \left\{ \gamma = \gamma_1\gamma_2 : \begin{array}{c} \gamma_1, \gamma_2 \in \mathcal{A} \\ T_1 < ||\gamma_1|| < 2T_1 \\ T_1 < ||\gamma_2|| < 2T_2 \\ < \mathbf{e}_2, \gamma_1\gamma_2\mathbf{r} > \frac{T}{100} \end{array} \right\}$$

Recall that the Hausdorff dimension of the circle packing δ is strictly greater than 1. The size of $\mathfrak{F}$ is $\asymp T^\delta$, which can be seen from [15]. The last condition in the definition of $\mathfrak{F}$ implies that $b_\alpha \asymp T$, which is crucial in our minor arc analysis later. The subset of norm X^2 is the image of some elements of the form $\begin{pmatrix} x & 2y \\ * & * \end{pmatrix}$ in Γ_{C_3}, with $x, y \asymp X$, under the map ρ.

For technical reasons we need to smooth the variables x and y. We fix a smooth, nonnegative function ψ which is supported in $[1, 2]$ and $\int_{\mathbb{R}} \psi(x)\mathrm{d}x = 1$. Our main goal is to study the following representation number

$$\mathcal{R}_N(n) := \sum_{\mathfrak{f} \in \mathfrak{F}_T} \sum_{x,y \in \mathbb{Z}(x,2y)=1} \psi\left(\frac{x}{X}\right) \psi\left(\frac{2y}{X}\right) \mathbf{1}_{\{n=\mathfrak{f}(x,2y)\}} \tag{18}$$

via its Fourier transform:

$$\widehat{\mathcal{R}}_N(\theta) := \sum_{\mathfrak{f} \in \mathfrak{F}_T} \sum_{x,y \in \mathbb{Z}(x,2y)=1} \psi\left(\frac{x}{X}\right) \psi\left(\frac{2y}{X}\right) e(\theta\mathfrak{f}(x, 2y)) \tag{19}$$

$\mathcal{R}_N$ and $\widehat{\mathcal{R}}_N$ is related by

$$\mathcal{R}_N(n) = \int_0^1 \widehat{\mathcal{R}}_N(\theta) e(-n\theta) d\theta.$$

Therefore, $\mathcal{R}_N(n) \neq 0$ implies n is represented. Since $\delta > 1$, one expects roughly that each admissible n is represented by $T^{\delta-1}$ times. One important thing for circle method here is that $T^{\delta-1}$ is a positive power of N, so we have enough solutions to play with.

Another technicality is that we replace the condition $(x, 2y) = 1$ by the Möbius orthogonal relation:

$$\sum_{d|n} \mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

We introduce another parameter U which is a small power of N. It is determined in (56). We then define the corresponding representation function

$$\mathcal{R}_N^U(n) := \sum_{\mathfrak{f} \in \mathfrak{F}_T} \sum_{x,y \in \mathbb{Z}} \sum_{\substack{u | (x,2y) \\ u < U}} \mu(u) \psi\left(\frac{x}{X}\right) \psi\left(\frac{2y}{X}\right) \mathbf{1}_{\{n = \mathfrak{f}(x,2y)\}}$$

and its Fourier transform:

$$\widehat{\mathcal{R}}_N^U(\theta) := \sum_{\mathfrak{f} \in \mathfrak{F}_T} \sum_{x,y \in \mathbb{Z}} \sum_{\substack{u | (x,2y) \\ u < U}} \mu(u) \psi\left(\frac{x}{X}\right) \psi\left(\frac{2y}{X}\right) e(\theta \mathfrak{f}(x, 2y))$$

The ℓ^1 norm of $\mathcal{R}_N$ is $\asymp T^\delta X^2$. We first show that the difference between $\mathcal{R}_N$ and $\mathcal{R}_N^U$ is small in ℓ^1, compared to $T^\delta X^2$:

Lemma 3.2.

$$\sum_{n<N} \left| \mathcal{R}_N(n) - \mathcal{R}_N^U(n) \right| \ll_\epsilon \frac{T^\delta X^{2+\epsilon}}{U}.$$

Proof.

$$\begin{aligned}
\sum_{n<N} |\mathcal{R}_N(n) - \mathcal{R}_N^U(n)| &= \sum_{n<N} \left| \sum_{\mathfrak{f} \in \mathfrak{F}_T} \sum_{(x,2y)=1} \sum_{\substack{u | (x,2y) \\ u \geq U}} \mu(u) \psi(\frac{x}{X}) \psi(\frac{2y}{X}) \mathbf{1}_{\{n = \mathfrak{f}(x,2y)\}} \right| \\
&\leq \sum_{\mathfrak{f} \in \mathfrak{F}} \sum_{x,y \in \mathbb{Z}} \psi(\frac{x}{X}) \psi(\frac{2y}{X}) \left| \sum_{\substack{u | (x,2y) \\ u \geq U}} \mu(u) \right| \\
&\ll \sum_{\mathfrak{f} \in \mathfrak{F}} \sum_{x \ll X} \sum_{\substack{u \geq U \\ u | x}} \sum_{\substack{y \ll X \\ 2y \equiv 0 (u)}} 1 \ll \frac{T^\delta X^{2+\epsilon}}{U}
\end{aligned}$$

□

Now we decompose [0,1] into "major" and "minor' arcs according to the standard Diophantine approximation of real numbers by rationals. Let $M = TX$ be the parameter controlling the depth of the approximation. Write $\alpha = \frac{a}{q} + \beta$. We introduce two parameters Q_0, K_0

such that the major arcs corresponds to $q \leq Q_0, \beta \leq \frac{K_0}{N}$. Both Q_0 and K_0 are small powers of N, and they are determined in (56).

Next we introduce the "hat" function

$$\mathfrak{t} := \min(1+x, 1-x)^+$$

whose Fourier transform is

$$\hat{t}(y) = \left(\frac{\sin(\pi y)}{\pi y}\right)^2.$$

From $\mathfrak{t}$, we construct a spike function $\mathfrak{T}$ which captures the major arcs:

$$\mathfrak{T}(\theta) := \sum_{q \leq Q_0} \sum_{(r,q)=1} \sum_{m \in \mathbb{Z}} \mathfrak{t}\left(\frac{N}{K_0}\left(\theta + m - \frac{a}{q}\right)\right).$$

The "main" term is then defined to be:

$$\mathcal{M}_N(n) := \int_0^1 \mathfrak{T}(\theta)\widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta \tag{20}$$

and the "error" term

$$\mathcal{E}_N(n) := \int_0^1 (1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta. \tag{21}$$

We define $\mathcal{M}_N^U(n)$ and $\mathcal{E}_N^U(n)$ in a similar way.

Now we explain the general strategy to prove the Theorem 1.4.

$$\begin{array}{ccccc} \mathcal{R}_N & = & \mathcal{M}_N & + & \mathcal{E}_N \\ | & & | & & | \\ \mathcal{R}_N^U & = & \mathcal{M}_N^U & + & \mathcal{E}_N^U \end{array} \tag{22}$$

STRATEGY :

(1) The difference between $\mathcal{R}_N$ and $\mathcal{R}_N^U$ is small in ℓ^1. We have shown this in Lemma 3.2.
(2) $\mathcal{M}_N$ is large for each n admissible in the range $(\frac{N}{2}, N)$ (See Theorem 3.5), and the difference of $\mathcal{M}_N$ and $\mathcal{M}_N^U$ is small in ℓ^2 (See Lemma 3.6). This will be done in §3.2.
(3) Step 2 will imply that the difference between $\mathcal{E}_N^U$ and $\mathcal{E}_N$ is also small. §3.3 to §3.5 will show that the $\mathcal{E}_N^U$ is small in l^2 (See Theorem 3.7), which implies that $\mathcal{E}_N$ is small in ℓ^1. This would greatly restrain the size of the set of admissible n's where $\mathcal{R}_N(n) = 0$ in $(\frac{N}{2}, N)$, because each term would contribute large to $\mathcal{E}_N$.

3.2. **Major Arc Analysis.** From (20),

$$\begin{aligned}\mathcal{M}_N(n) &= \int_0^1 \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{m\in\mathbb{Z}} \mathfrak{t}\left(\frac{N}{K_0}\left(\theta+m-\frac{r}{q}\right)\right)\widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta \\ &= \int_{-\infty}^{\infty} \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \mathfrak{t}\left(\frac{N}{K_0}\beta\right)\widehat{\mathcal{R}}_N\left(\beta+\frac{r}{q}\right)e\left(-n\left(\beta+\frac{r}{q}\right)\right)d\beta \\ &= \sum_{\substack{x,y\\(x,2y)=1}} \psi\left(\frac{x}{X}\right)\psi\left(\frac{2y}{X}\right) \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{\mathfrak{f}\in\mathfrak{F}} e\left(\frac{r}{q}(\mathfrak{f}(x,2y)-n)\right)\int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)e(\beta(\mathfrak{f}(x,2y)-n)d\beta\end{aligned} \tag{23}$$

Now we cite Lemma 5.3 from [4] to deal with the $\mathfrak{F}$ sum in (23).

Lemma 3.3 (Bourgain, Kontorovich)**.** *Let $1<K<T_2^{\frac{1}{10}}$, fix $|\beta|<\frac{K}{N}$, and fix $x,y\asymp X$. Then for any $\gamma_0\in\Gamma$, any $q\geq 1$, we have*

$$\sum_{\gamma\in\mathfrak{F}\cap\gamma_0\Gamma(q)} e(\beta\mathfrak{f}_\gamma(x,2y)) = \frac{1}{\Gamma:\Gamma(q)}\sum_{\mathfrak{f}\in\mathfrak{F}} e(\beta\mathfrak{f}_\gamma(x,2y)) + O(T^\Theta K),$$

where $\Theta<\delta$ depends only on the spectral gap for Γ, and the implied constant does not depend on q,γ_0,x or y.

Returning to (23), we can decompose the set $\mathfrak{F}$ as cosets of $\Gamma(q)$. Applying Lemma 3.3 and setting $K=K_0$, we have

$$\begin{aligned}\mathcal{M}_N(n) &= \sum_{\substack{x,y\in\mathbb{Z}\\(x,2y)=1}} \psi\left(\frac{x}{X}\right)\psi\left(\frac{2y}{X}\right) \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{\bar\gamma\in\Gamma/\Gamma(q)} e\left(\frac{r}{q}(\mathfrak{f}_{\bar\gamma}(x,2y)-n)\right) \\ &\quad\times \sum_{\substack{\gamma\in\mathfrak{F}\\ \gamma\equiv\bar\gamma}} \int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)e(\beta(\mathfrak{f}_\gamma(x,2y)-n))d\beta \\ &= \sum_{\substack{x,y\in\mathbb{Z}\\(x,2y)=1}} \psi\left(\frac{x}{X}\right)\psi\left(\frac{2y}{X}\right) \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{\bar\gamma\in\Gamma/\Gamma(q)} e\left(\frac{r}{q}(\mathfrak{f}_{\bar\gamma}(x,2y)-n)\right) \\ &\quad\times \left(\frac{1}{[\Gamma:\Gamma(q)]}\sum_{\gamma\in\mathfrak{F}}\int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)e(\beta(\mathfrak{f}_\gamma(x,2y)-n))d\beta + O\left(\frac{T^\Theta K_0^2}{N}\right)\right) \\ &= \sum_{r(q)}{}'\psi\left(\frac{x}{X}\right)\psi\left(\frac{2y}{X}\right)\mathfrak{S}_{Q_0}(n)\mathfrak{M}(n) + O\left(\frac{T^\Theta X^2K_0^2Q_0^8}{N}\right) \\ &= \sum_{r(q)}{}'\psi\left(\frac{x}{X}\right)\psi\left(\frac{2y}{X}\right)\mathfrak{S}_{Q_0}(n)\mathfrak{M}(n) + O\left(N^{-\eta}\right)\end{aligned} \tag{24}$$

where $\eta_1 > 0$, as can be seen from (56), and

$$\mathfrak{S}_{Q_0}(n) = \mathfrak{S}_{Q_0;x,y}(n) := \sum_{q<Q_0} \sum_{(r,q)=1} \frac{1}{[\Gamma : \Gamma(q)]} \sum_{\bar{\gamma}\in\Gamma/\Gamma(q)} e\left(\frac{r}{q}(\mathfrak{f}_{\bar{\gamma}}(x,2y)) - n\right) \tag{25}$$

$$= \sum_{q<Q_0} \frac{1}{[\Gamma/\Gamma(q)]} \sum_{\bar{\gamma}\in\Gamma/\Gamma(q)} c_q(\mathfrak{f}_{\bar{\gamma}}(x,2y) - n) \tag{26}$$

and

$$\mathfrak{M}(n) := \mathfrak{M}_{x,y}(n) := \sum_{\gamma\in\mathfrak{F}} \int_{-\infty}^{\infty} \mathfrak{t}\left(\frac{N}{K_0}\beta\right) e(\beta(\mathfrak{f}_\gamma(x,2y) - n))d\beta$$

$$= \frac{K_0}{N} \sum_{\gamma\in\mathfrak{F}} \hat{t}\left(\frac{K_0}{N}(\mathfrak{f}(x,2y) - n)\right) \tag{27}$$

The function $c_q(n)$ in (25) is the classical Ramanujan's sum, defined by

$$c_q(n) = \sum_{a(q)}{}' e\left(\frac{an}{q}\right),$$

$c_q(n)$ is multiplicative with respect to q, and

$$c_{p^k}(n) = \begin{cases} 0 & \text{if } p^m||n, m \le k-2, \\ -p^{k-1} & \text{if } p^{k-1}||n, \\ p^{k-1}(p-1) & \text{if } p^k|n. \end{cases}$$

Now $\mathfrak{M}(n) \gg \frac{T^\delta}{N}$ for $\frac{N}{2} < n < N$, which can be seen from the following lemma by Lemma 5.4 in [4]. We record it here:

Lemma 3.4 (Bourgain, Kontorovich)**.** *Fix $N/2 < n < N, 1 < K \le T_2^{\frac{1}{10}}$, and $x, y \asymp X$. Then*

$$\sum_{\gamma\in\mathfrak{F}} \mathbf{1}_{\{|\mathfrak{f}_\gamma(x,2y)-n|<\frac{N}{K}\}} \gg \frac{T^\delta}{K} + T^\Theta,$$

where $\Theta < \delta$ depends only on the spectral gap for Γ. The implied constant is independent of x, y and n.

Now we are at a position to analyze the non-Archimedean part $\mathfrak{S}_{Q_0}$. We push $\mathfrak{S}_{Q_0}(n)$ to infinity, and define

$$\mathfrak{S}(n) := \sum_{q=1}^{\infty} \frac{1}{[\Gamma : \Gamma(q)]} \sum_{\bar{\gamma}\in\Gamma/\Gamma(q)} c_q(\mathfrak{f}_{\bar{\gamma}}(x,2y) - n)$$

$$= \sum_{q=1}^{\infty} \sum_{a\in\mathbb{Z}/q\mathbb{Z}} \tau_q(a) c_q(a-n) := \sum_{q=1}^{\infty} B_q(n),$$

where

$$\tau_q(a) = \frac{\#\{\langle u,v,w\rangle (\mathrm{mod}\ q) | \langle a,u,v,w\rangle \in \mathcal{P}\}}{\#\{\langle x,u,v,w\rangle (\mathrm{mod}\ q) | \langle x,u,v,w\rangle \in \mathcal{P}\}}$$

From Theorem 2.10 we know that $\tau_q(n)$ is multiplicative in the q variable, and so is $B_q(n)$. Therefore, we can formaly write

$$\mathfrak{S}(n) = \prod_p (1 + B_p(n) + B_{p^2}(n) + \dots)$$

For $p \geq 3$, by Theorem 2.10, we can show that

$$B_p(n) = \begin{cases} \frac{-1-p(\frac{-2}{p})}{p^2+(1+(\frac{-2}{p}))p+1} & \text{if} p|n, \\ \frac{p(\frac{-2}{p})+1}{p^3+p(p-1)(\frac{-2}{p})-1} & \text{if} p \nmid n, \end{cases}$$

and $B_{p^k} = 0$ for $k \geq 2$. For $p = 2$, we have $B_{2^m} = 0$ for $m \geq 4$ and

$$1 + B_2(n) + B_4(n) + B_8(n) = \begin{cases} 8 & \text{if } n \equiv \kappa_1 (\text{mod } 8) \\ 0 & \text{otherwise.} \end{cases}$$

Thus we see that $\mathfrak{S}_{Q_0}$ is a non-negative function which is non-zero if and only if $n \equiv \kappa_1 (\text{mod } 8)$, which matches exactly the local obstruction described in Theorem 2.10. For such admissible n's, $\mathfrak{S}_{Q_0}$ satisfies $N^{-\epsilon} \ll_\epsilon \mathfrak{S}_{Q_0}(n) \ll_\epsilon N^\epsilon$.

To analyze $\mathcal{R}_N^U$ we need to extend the definition of $\mathfrak{S}_{x,y}(n)$ restricted to $(x, 2y) = 1$ to all pair of integers x, y. If $(x, 2y) = u > 1$, the same calculation shows that $\mathfrak{S}_{Q_0;x,y}(n)$ has the same local factor for $p \neq 2$, and $B_{2^m} = 0$ for $m \geq 4$. Therefore, $\mathfrak{S}_{x,y}(n) \ll_\epsilon N^\epsilon$ for any $x, y \in \mathbb{Z}$.

The difference between $\mathfrak{S}$ and $\mathfrak{S}_{Q_0}$ is small. In fact, we have

$$|\mathfrak{S}(n) - \mathfrak{S}_{Q_0;x,y}(n)| \leq \sum_{q \geq Q_0} |B_q(n)| \leq \sum_{q_1 : n} |B_{q_1}(n)| \sum_{\substack{(q_2,q_1)=1 \\ q_1 q_2 \geq Q_0}} |B_{q_2}(n)| \tag{28}$$

Here we write $q = q_1 q_2$, where $q_1 : n$ means that q_1 is the product of all primes dividing n. We also know that $B_q(n)$ as a function of q is supported on (almost) square-free numbers (as can be see by the previous paragraphs), we have

$$(28) \ll \sum_{q_1 : n} \frac{1}{q_1} \frac{q_1}{Q_0} \ll \frac{2^{w(n)}}{Q_0}$$

where $w(n)$ denotes the number of primes dividing n. Therefore, we conclude that if n is admissible, then $N^{-\epsilon} \ll \mathfrak{S}_{Q_0}(n) \ll_\epsilon N^\epsilon$. In summary, we have

Theorem 3.5. *For $\frac{N}{2} < n < N$, there exists a function $\mathfrak{S}_{Q_0}(n)$ such that if n is admissible, then*

$$\mathcal{M}_N(n) \gg \mathfrak{S}_{Q_0}(n) T^{\delta - 1},$$

where

$$N^{-\epsilon} \ll_\epsilon \mathfrak{S}_{Q_0}(n) \ll_\epsilon N^\epsilon.$$

Next we show that the difference of $\mathcal{M}_N$ and $\mathcal{M}_N^U$ is small in ℓ^1.

Lemma 3.6.

$$\sum_{\frac{N}{2}<n<N} |\mathcal{M}_N(n) - \mathcal{M}_N^U(n)| \ll_\epsilon \frac{N^\epsilon X^2 T^\delta}{U} + \frac{T^\Theta X^2 K_0^2 Q_0^2}{U},$$

where Θ is the same as in Lemma 3.3.

Proof. Going in the same way as (23) to unfold $\mathcal{M}_N^U(n)$, we have

$$\begin{aligned}
\mathcal{M}_N(n) - \mathcal{M}_N^U(n) &= \sum_{\substack{u\geq U\\ u \text{ odd}}} \mu(u) \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{xu}{X}\right)\psi\left(\frac{2yu}{X}\right) \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{\bar\gamma\in\Gamma/\Gamma(q)} e\left(\frac{r}{q}(\mathfrak{f}_{\bar\gamma}(xu,2yu)-n)\right)\\
&\quad\times \sum_{\substack{\gamma\in\mathfrak{F}\\ \gamma\equiv\bar\gamma (\mathrm{mod}\ \Gamma(q))}} \int_{-\infty}^{\infty} \mathfrak{t}\left(\frac{N}{K_0}\theta\right) e(\theta(\mathfrak{f}_\gamma(xu,2yu)-n))d\theta\\
&\quad+ \sum_{\substack{u\geq U\\ u \text{ even}}} \mu(u) \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right) \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{\bar\gamma\in\Gamma/\Gamma(q)} e\left(\frac{r}{q}(\mathfrak{f}_{\bar\gamma}(xu,yu)-n)\right)\\
&\quad\times \sum_{\substack{\gamma\in\mathfrak{F}\\ \gamma\equiv\bar\gamma (\mathrm{mod}\ \Gamma(q))}} \int_{-\infty}^{\infty} \mathfrak{t}\left(\frac{N}{K_0}\theta\right) e(\theta(\mathfrak{f}_\gamma(xu,yu)-n))d\theta\\
&= \sum_{\substack{u\geq U\\ u \text{ odd}}} \mu(u) \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{xu}{X}\right)\psi\left(\frac{2yu}{X}\right) \mathfrak{S}_{Q_0,(xu,2yu)}(n) \times \sum_{\gamma\in\mathfrak{F}} \frac{K_0}{N}\hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}_\gamma(xu,2yu)-n)\right)\\
&\quad+ \sum_{\substack{u\geq U\\ u \text{ even}}} \mu(u) \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right) \mathfrak{S}_{Q_0,(xu,yu)}(n) \times \sum_{\gamma\in\mathfrak{F}} \frac{K_0}{N}\hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}_\gamma(xu,yu)-n)\right)\\
&\quad+ O\left(\frac{T^\Theta X^2 K_0^2 Q_0^2}{NU}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{\frac{N}{2}<n<N} |\mathcal{M}_N(n) - \mathcal{M}_N^U(n)|\\
&\ll \sum_{\substack{u>U\\ u \text{ odd}}} \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{xu}{X}\right)\psi\left(\frac{2yu}{X}\right)\frac{K_0}{N} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\frac{N}{2}<n<N} \mathfrak{S}_{Q_0}(n)\hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}(xu,2yu)-n)\right)\\
&+ \sum_{\substack{u>U\\ u \text{ even}}} \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)\frac{K_0}{N} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\frac{N}{2}<n<N} \mathfrak{S}_{Q_0}(n)\hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}(xu,yu)-n)\right) + \frac{T^\Theta X^2 K_0^2 Q_0^2}{U}\\
&\ll_\epsilon \frac{X^2 T^\delta N^\epsilon}{U} + \frac{T^\Theta X^2 K_0^2 Q_0^2}{U}
\end{aligned}$$

□

In light of (56), we have

$$\sum_{\frac{N}{2}<n<N} |\mathcal{M}_N(n) - \mathcal{M}_N^U(n)| \ll_\eta T^\delta X^2 N^{-\eta}$$

3.3. **Minor Arc Analysis I.** The rest few sections of the paper is dedicated to proving Theorem 3.7, which shows that $(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U$ is small in L^2. By Plancherel formula this will imply that $\mathcal{E}_N^U$ is small in ℓ^2, fulfilling Step 3 of our strategy.

Theorem 3.7.

$$\int_0^1 \left|(1-\mathfrak{T}(\theta))\mathcal{R}_N^U(\theta)\right|^2 d\theta \ll NT^{2(\delta-1)}N^{-\eta}$$

We divide the integral into three parts.

$$\mathcal{I}_1 = \sum_{q<Q_0}\sideset{}{'}\sum_{r(q)} \int_{\frac{r}{q}-\frac{1}{qM}}^{\frac{r}{q}+\frac{1}{qM}} |(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta \tag{29}$$

$$\mathcal{I}_2 = \sum_{Q_0\le q<X}\sideset{}{'}\sum_{r(q)} \int_{\frac{r}{q}-\frac{1}{qM}}^{\frac{r}{q}+\frac{1}{qM}} |(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta \tag{30}$$

$$\mathcal{I}_3 = \sum_{X\le Q\le M}\sideset{}{'}\sum_{r(q)} \int_{\frac{r}{q}-\frac{1}{qM}}^{\frac{r}{q}+\frac{1}{qM}} |(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta \tag{31}$$

corresponding to different ranges of q. We will show that $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$ are bounded by the same bound as in Theorem 3.7, which immediately implies Theorem 3.7. This section is to deal with $\mathcal{I}_1$, and the next two sections deal with $\mathcal{I}_2, \mathcal{I}_3$ respectively.

First we re-order the sum in $\widehat{\mathcal{R}}_N^U$ according to the u variable:

$$\begin{aligned}
\widehat{\mathcal{R}}_N^U(\theta) &= \sum_{x,y\in\mathbb{Z}}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{u<U}\mu(u)\psi\left(\frac{x}{X}\right)\psi\left(\frac{2y}{X}\right)e(\theta\mathfrak{f}(x,2y)) \\
&= \sum_{u \text{ odd}}\mu(u)\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{x,y\in\mathbb{Z}}\psi\left(\frac{xu}{X}\right)\psi\left(\frac{2yu}{X}\right)e(\theta\mathfrak{f}(xu,2yu)) \\
&+ \sum_{u \text{ even}}\mu(u)\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{x,y\in\mathbb{Z}}\psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)e(\theta\mathfrak{f}(xu,yu)) \\
&:= \sum_{u<U}\mu(u)\sum_{\mathfrak{f}\in\mathfrak{F}}\mathcal{R}_{u,\mathfrak{f}}(\theta)
\end{aligned} \tag{32}$$

For simplicity we restrict our attention to u even. The same argument is applied to u odd. We write $\frac{u^2}{q} = \frac{u_0}{q_0}$ in irreducible form, then we have

$$\begin{aligned}
\mathcal{R}_{u,\mathfrak{f}}\left(\frac{r}{q}+\beta\right) &= \sum_{x,y\in\mathbb{Z}}\mu(u)\psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)e\left(\mathfrak{f}(xu,yu)\left(\frac{r}{q}+\beta\right)\right) \\
&= e\left(-b_{\mathfrak{f}}\left(\frac{r}{q}+\beta\right)\right)\sum_{x_0,y_0(q_0)}e\left(\frac{u_0}{q_0}\tilde{\mathfrak{f}}(x_0,y_0)r\right) \\
&\times\left[\sum_{x\equiv x_0(q_0),y\equiv y_0(q_0)}\psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)e(\tilde{\mathfrak{f}}(xu,yu)\beta)\right]
\end{aligned} \tag{33}$$

Now applying Poisson summation to the bracket, we have

$$\begin{aligned}[\cdot] &= \sum_{\xi,\zeta\in\mathbb{Z}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi\left(\frac{(xq_0+x_0)u}{X}\right)\psi\left(\frac{(yq_0+y_0)u}{X}\right) e\left(\beta\tilde{\mathfrak{f}}((x_0+xq_0)u,(y_0+yq_0)u)-x\xi-y\zeta\right)dxdy \\ &= \frac{X^2}{u^2q_0^2}\sum_{\xi,\zeta\in\mathbb{Z}} e\left(\frac{x_0\xi}{q_0}+\frac{y_0\zeta}{q_0}\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(\tilde{\mathfrak{f}}(x,y)X^2\beta-\frac{X\xi}{uq_0}x-\frac{X\zeta}{uq_0}y\right)dxdy \end{aligned} \tag{34}$$

Putting (34) back to (32), we have

$$\mathcal{R}_{u,\mathfrak{f}}\left(\frac{r}{q}+\beta\right) = \frac{X^2}{u^2}e\left(-b_{\mathfrak{f}}\left(\frac{r}{q}+\beta\right)\right)\sum_{\xi,\zeta\in\mathbb{Z}}\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta)\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta),$$

where

$$\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta) := \frac{1}{q_0^2}\sum_{x_0,y_0(q_0)} e\left(\frac{u_0r}{q_0}\tilde{\mathfrak{f}}(x_0,y_0)+\frac{x_0\xi}{q_0}+\frac{y_0\zeta}{q_0}\right),$$

and

$$\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta) := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(\tilde{\mathfrak{f}}(x,y)X^2\beta-\frac{X\xi}{uq_0}x-\frac{X\zeta}{uq_0}y\right)dxdy.$$

We can compute $\mathcal{S}_{\mathfrak{f}}$ explicitly. For simplicity we assume q_0 is odd, and $A_{\mathfrak{f}}$ is invertible in $\mathbb{Z}/q_0\mathbb{Z}$. We record a standard fact of exponential sum:

$$\sum_{a\in\mathbb{Z}/q\mathbb{Z}} e_q(x^2) = i^{\epsilon(q)}q^{\frac{1}{2}},$$

where $\epsilon(q)=0$ if $q\equiv 1(4)$ and $\epsilon(q)=1$ if $q\equiv 3(4)$. From this, one can get

$$\sum_{r\in\mathbb{Z}/q\mathbb{Z}} e_q(rx^2) = \left(\frac{r}{q}\right)i^{\epsilon(q)}q^{\frac{1}{2}} \tag{35}$$

if $(r,q)=1$. Now complete square of $\mathcal{S}_{\mathfrak{f}}$ and apply (35) to $\mathcal{S}_{\mathfrak{f}}$, we get

$$\begin{aligned}\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta) &= \frac{1}{q_0^2}\sum_{x_0,y_0(q_0)} e_{q_0}\left(u_0r\tilde{\mathfrak{f}}(x_0,y_0)+x_0\xi+y_0\zeta\right) \\ &= \frac{1}{q_0^2}\sum_{x_0,y_0(q_0)} e\left(u_0rA_{\mathfrak{f}}\left(x_0+B_{\mathfrak{f}}\bar{A}_{\mathfrak{f}}y_0\right)^2+\xi\left(x_0+B_{\mathfrak{f}}\bar{A}_{\mathfrak{f}}y_0\right)+2u_0r\bar{A}_{\mathfrak{f}}b_{\mathfrak{f}}^2y_0^2+\left(\zeta-\xi B_{\mathfrak{f}}\bar{A}_{\mathfrak{f}}\right)y_0\right) \\ &= \frac{1}{q_0^2}q_0^{\frac{1}{2}}i^{\epsilon(q_0)}\left(\frac{u_0rA_{\mathfrak{f}}}{q_0}\right)e_{q_0}\left(-\overline{4u_0rA_{\mathfrak{f}}}\xi^2\right)\sum_{y_0(q_0)} e_{q_0}\left(2u_0r\bar{A}_{\mathfrak{f}}b_{\mathfrak{f}}^2y_0^2+\left(\zeta-\xi B_{\mathfrak{f}}\bar{A}_{\mathfrak{f}}\right)y_0\right)\end{aligned} \tag{36}$$

To deal with the sum in the above expression, we write $\frac{b_{\mathfrak{f}}^2}{q_0} = \frac{b_1}{q_1}$ where $(b_1, q_1) = 1$. Then after a linear change of variables and completing square we obtain

$$\mathcal{S}_{\mathfrak{f}}(q_0, u_0 r, \xi, \zeta) = \frac{i^{\epsilon(q_0)+\epsilon(q_1)}}{q_0^{\frac{1}{2}} q_1^{\frac{1}{2}}} \mathbf{1}_{\{A_{\mathfrak{f}}\zeta \equiv B_{\mathfrak{f}}\xi(\frac{q_0}{q_1})\}} \left(\frac{u_0 r A_{\mathfrak{f}}}{q_0}\right) \left(\frac{2u_0 r b_{\mathfrak{f}} \bar{A}_{\mathfrak{f}}}{q_1}\right)$$
$$\times e_{q_0}\left(-\overline{4u_0 r A_{\mathfrak{f}}}\xi^2\right) e_{q_1}\left(-\overline{8u_0 r b_1 A_{\mathfrak{f}}}\left(\frac{q_1(A_{\mathfrak{f}}\zeta - B_{\mathfrak{f}}\xi)}{q_0}\right)^2\right) \tag{37}$$

From (37) we see trivially that $|\mathcal{S}_{\mathfrak{f}}(q_0, u_0 r, \xi, \zeta)| \leq q_0^{-\frac{1}{2}}$.

Now we deal with $\mathcal{J}_{\mathfrak{f}}$. For this we need standard results from non-stationary phase and stationary phase, and we record them here.

Non-stationary phase: Let ϕ be a smooth compactly supported function on $(-\infty, \infty)$ and f be a function which satisfies $|f^{'}(x)| > A > 0$ in the support of ϕ and $A \geq |f^{(2)}(x)|, ..., f^{(n)}(x)$ in the support of ϕ. Then

$$\int_{-\infty}^{\infty} \phi(x) e(f(x)) dx \ll_{\phi, N} A^{-N}$$

Proof. By partial integration,

$$\int_{-\infty}^{\infty} \phi(x) e(f(x)) = \int_{-\infty}^{\infty} \frac{\phi(x)}{f^{'}(x)} de(f(x))$$
$$= -\int_{-\infty}^{\infty} \left(\frac{\phi}{f^{'}}\right)^{'}(x) e(f(x)) dx = -\int_{-\infty}^{\infty} \frac{\phi^{'}(x)}{f^{'}(x)} + \frac{\phi(x) f^{(2)}(x)}{(f^{'}(x))^2} dx$$

From here, we see already that

$$\int_{-\infty}^{\infty} \phi(x) e(f(x)) dx \ll_{\phi, N} A^{-1}$$

Iterating partial integration N times we can get the A^{-N} bound. □

Stationary phase: Let f be a quadratic polynomial of two variables x and y with discriminant $-D$, where $D > 0$. Let $\phi(x, y)$ be a smooth compactly supported function on $\mathbb{R}^2$, then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x, y) e(f(x, y)) dx dy \ll_{\phi} \frac{1}{\sqrt{D}}.$$

Proof. After using an orthonormal matrix L to change variables we can change the above integral into the from

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(L(x, y)) e\left(-x^2 - \frac{D}{4} y^2\right) dx dy$$

Using Plancherel formula,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(L(x, y)) e\left(-x^2 - \frac{D}{4} y^2\right) dx dy = \frac{1}{i\sqrt{D}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \widehat{\phi \circ L}(u, v) e\left(\frac{u^2}{4} + \frac{v^2}{D}\right) du dv$$

We caution the reader that $e(-x^2-\frac{D}{4}y^2)$ is not in L^2, the above formula is obtained in the following way: first approximate $e^{2\pi i(-x^2-\frac{D}{4})y^2}$ by $e^{(-\epsilon+2\pi i)(-x^2-\frac{D}{r}y^2)}$, where we can apply Plancherel formula, then let $\epsilon\to 0$ and pass the limit. Therefore,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(x,y)e(f(x,y))dxdy\leq\frac{1}{\sqrt{D}}||\widehat{\phi\circ L}||_1\leq\frac{1}{\sqrt{D}}||\phi||_1 \tag{38}$$

□

If either $\xi\gg U\geq TX\beta uq_0$ or $\zeta\gg U\geq TX\beta uq_0$, then the non-stationary phase condition is satisfied, we have $\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta)\ll(\frac{uq_0}{X\xi})^N$ for any N, so these terms are negligible. Now we deal with the case $\xi,\zeta\ll U$. Recall that the discriminant of $\mathfrak{f}$ is $-8b_{\mathfrak{f}}^2$, by the stationary phase, we have

$$\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta)\ll\min\left\{1,\frac{1}{TX^2|\beta|}\right\} \tag{39}$$

With this, one gets

$$\mathcal{R}_{u,\mathfrak{f}}(\frac{r}{q}+\beta)\ll\frac{X^2}{u^2}\sum_{\xi,\zeta\ll u}q_0^{-\frac{1}{2}}\frac{1}{TX^2|\beta|}\ll\frac{u}{q^{\frac{1}{2}}T|\beta|}$$

using the fact that $u^2q_0\geq q$. Therefore, we have

$$\mathcal{R}_N^U\left(\frac{r}{q}+\beta\right)\ll T^{\delta}\sum_{u<U}\frac{u}{q^{\frac{1}{2}}T|\beta|}\ll\frac{T^{\delta-1}U^2}{q^{\frac{1}{2}}|\beta|} \tag{40}$$

Now we are able to bound $\mathcal{I}_1$

Lemma 3.8.

$$\mathcal{I}_1\ll NT^{2(\delta-1)}N^{-\eta}$$

Proof. We divide the integral into three parts:

$$\begin{aligned}\mathcal{I}_1&=\sum_{q<Q_0}\sum_{r(q)}{}'\int_{\frac{r}{q}-\frac{1}{qM}}^{\frac{r}{q}+\frac{1}{qM}}\left|(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)\right|^2d\theta\\&=\sum_{q<Q_0}\sum_{r(q)}{}'\int_{-\frac{K_0}{N}}^{\frac{K_0}{N}}|\cdot|^2d\beta+\int_{\frac{K_0}{N}}^{\frac{1}{qM}}|\cdot|^2d\beta+\int_{-\frac{1}{qM}}^{-\frac{K_0}{N}}|\cdot|^2d\beta\end{aligned}$$

For the first summand, we insert $|1-\mathfrak{T}(\frac{r}{q}+\beta)|^2=\frac{N^2\beta^2}{K_0^2}$ and bound $\widehat{\mathcal{R}}_N^U$ by (40). For the second and the third summands, we trivially bound $|1-\mathfrak{T}(\theta)|^2$ by 1 and $\widehat{\mathcal{R}}_N^U$ by (40). Then we get

$$\mathcal{I}_1\ll\frac{NQ_0T^{2(\delta-1)}U^4}{K_0}\ll_{\eta}T^{2\delta-1}X^2N^{-\eta}, \tag{41}$$

which is a power saving. □

3.4. **Minor Arc Analysis II.** In this section we deal with $\mathcal{I}_2$. We divide the q-sum 2-adically:

$$\mathcal{I}_Q := \sum_{Q\leq q<2Q} {\sum_{r(q)}}' \int_{\frac{r}{q}-\frac{1}{qM}}^{\frac{r}{q}+\frac{1}{qM}} \left|\widehat{\mathcal{R}}_N^U(\theta)\right|^2 d\theta \tag{42}$$

We will show that for all $Q_0 \leq Q < X$, $\mathcal{I}_Q$ has a power saving, in the next section we will show that $\mathcal{I}_Q$ has a power saving for the range $X \leq Q \leq M$. Clearly these will imply Theorem 3.7.

Recall from (32) that

$$\begin{aligned}\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right) &= \sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}} \mathcal{R}_{u,\mathfrak{f}}\left(\frac{r}{q}+\beta\right)\\ &= \sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}} e\left(-b_{\mathfrak{f}}\left(\frac{r}{q}+\beta\right)\right)\frac{X^2}{u^2}\sum_{\xi,\zeta\in\mathbb{Z}}\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta)\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta)\end{aligned} \tag{43}$$

Apply Cauchy-Schwartz inequality to the u variable, we have

$$\begin{aligned}\left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 \leq& X^4\left|\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\xi,\zeta\in\mathbb{Z}} e\left(-b_{\mathfrak{f}}\left(\frac{r}{q}+\beta\right)\right)\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta)\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta)\right|^2\\ =& X^4\sum_{\mathfrak{f},\mathfrak{f}'\in\mathfrak{F}} e\left(-\left(b_{\mathfrak{f}}-b_{\mathfrak{f}'}\right)\frac{r}{q}\right)\sum_{\xi,\zeta\in\mathbb{Z}}\sum_{\xi',\zeta'\in\mathbb{Z}}\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta)\overline{\mathcal{S}_{\mathfrak{f}'}(q_0,u_0r,\xi',\zeta')}\\ &\mathcal{J}_{\mathfrak{f}}(\beta;uq_0,\xi,\zeta)\overline{\mathcal{J}_{\mathfrak{f}'}(\beta;uq_0,\xi',\zeta')}e\left(-(b_{\mathfrak{f}}-b_{\mathfrak{f}'})\beta\right)\end{aligned}$$

Changing variables $\theta = \frac{r}{q}+\beta$ in (42) and putting (44) back to (42), we get

$$\begin{aligned}\mathcal{I}_Q \ll X^4 \sum_{\mathfrak{f},\mathfrak{f}'\in\mathfrak{F}}\sum_{\xi,\zeta\in\mathbb{Z}}\sum_{\xi',\zeta'\in\mathbb{Z}}\sum_{Q\leq q<2Q}\left({\sum_{r(q)}}' e\left(-\left(b_{\mathfrak{f}}-b_{\mathfrak{f}'}\right)\frac{r}{q}\right)\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta)\overline{\mathcal{S}_{\mathfrak{f}'}(q_0,u_0r,\xi',\zeta')}\right)\\ \times\int_{-\frac{1}{qM}}^{\frac{1}{qM}}\mathcal{J}_{\mathfrak{f}}(\beta,uq_0,\xi,\zeta)\overline{\mathcal{J}_{\mathfrak{f}'}(\beta;uq_0,\xi',\zeta')}e((-b_{\mathfrak{f}}+b_{\mathfrak{f}})\beta)d\beta\end{aligned} \tag{44}$$

We again split $\mathcal{I}_Q$ into non-Archimedean and Archimedean pieces. For the Archimedean part, we use (39) to bound $\mathcal{J}$. We have

$$\begin{aligned}&\int_{-\frac{1}{qM}}^{\frac{1}{qM}}\mathcal{J}_{\mathfrak{f}}(\beta,uq_0,\xi,\zeta)\overline{\mathcal{J}_{\mathfrak{f}'}(\beta;uq_0,\xi',\zeta')}e((-b_{\mathfrak{f}}+b_{\mathfrak{f}})\beta)d\beta \ll \int_{-\infty}^{\infty}\min\left\{1,\frac{1}{TX^2|\beta|}\right\}^2 d\beta\\ &\ll \int_{-\frac{1}{TX^2}}^{\frac{1}{TX^2}} 1d\beta + \left(\int_{-\infty}^{-\frac{1}{TX^2}}+\int_{\frac{1}{TX^2}}^{\infty}\right)\frac{1}{T^2X^4\beta^2}d\beta \ll \frac{1}{TX^2}.\end{aligned} \tag{45}$$

Now we analyze the non-Archimedean part. Again for simplicity we only deal with q_0 odd, and $A_{\mathfrak{f}}$, $A_{\mathfrak{f}'}$ invertible in $\mathbb{Z}/q_0\mathbb{Z}$. We set

$$\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}') = {\sum_{r(q)}}' e\left(-\left(b_{\mathfrak{f}}-b_{\mathfrak{f}'}\right)\frac{r}{q}\right)\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,\xi,\zeta)\overline{\mathcal{S}_{\mathfrak{f}'}(q_0,u_0r,\xi',\zeta')} \tag{46}$$

Recall that $\frac{b_{\mathfrak{f}}^2}{q_0}=\frac{b_1}{q_1}$, and similarly we write $\frac{b_{\mathfrak{f}'}^2}{q_0}=\frac{b_1'}{q_1'}$. Plug (37) in(46) , then we obtain

$$\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')=\mathbf{1}_{\substack{A_{\mathfrak{f}}\zeta\equiv B_{\mathfrak{f}}\xi(\frac{q_0}{q_1})\\ A_{\mathfrak{f}'}\zeta'\equiv B_{\mathfrak{f}'}\xi'(\frac{q_0}{q_1'})}}\times\frac{i^{\epsilon(q_1)-\epsilon(q_1')}}{q_0q_1^{\frac{1}{2}}q_1'^{\frac{1}{2}}}\left(\frac{A_{\mathfrak{f}}}{q_0}\right)\left(\frac{A_{\mathfrak{f}'}}{q_0}\right)$$

$$\times\sum_{r(q)}{}'\left(\frac{2u_0rb_1\bar{A}_{\mathfrak{f}}}{q_1}\right)\left(\frac{2u_0rb_1'\bar{A}_{\mathfrak{f}'}}{q_1'}\right)e_q((-b_{\mathfrak{f}}+b_{\mathfrak{f}'})r)e_{q_0}\left(-\overline{8u_0b_1A_{\mathfrak{f}}}\frac{q_0}{q_1}\left(\frac{q_1(A_{\mathfrak{f}}\zeta-B_{\mathfrak{f}}\xi)}{q_0}\right)^2\bar{r}\right)$$

$$\times e_{q_0}\left(\overline{8u_0b_1'A_{\mathfrak{f}'}}\frac{q_0}{q_1'}\left(\frac{q_1'(A_{\mathfrak{f}'}\zeta'-B_{\mathfrak{f}'}\xi')}{q_0}\right)^2\bar{r}\right)e_{q_0}\left(-\overline{4u_0rA_{\mathfrak{f}}}\xi^2+\overline{4u_0rA_{\mathfrak{f}'}}\xi'^2\right)\tag{47}$$

This is a type of Kloosterman sum. For our use in (47) we only need an elementary $\frac{3}{4}$ bound originally due to Kloosterman [14] (compared to the $\frac{1}{2}$ bound implied by the Weil conjecture). We stated it here:

Lemma 3.9. *Let $S(m,n,q,\chi)=\sum_{x(q)}{}' e_q(mx+n\bar{x})\chi(x)$, then we have*

$$S(m,n,q,\chi)\ll_\epsilon \min\{(m,q),(n,q)\}^{\frac{1}{4}}q^{\frac{3}{4}+\epsilon}.$$

We have an extra multiplicative character χ compared to the original paper by Kloosterman [14], but his proof is easily modified to suit our case.

Apply Lemma 3.9 to (47), and recall that $q_1=\frac{q_0}{(q_0,b_{\mathfrak{f}}^2)}, q_1'=\frac{q_0}{(q_0,b_{\mathfrak{f}'}^2)}$, then we obtain

$$|\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')|\ll_\epsilon\left(b_{\mathfrak{f}}-b_{\mathfrak{f}'},q\right)^{\frac{1}{4}}\left(\frac{q}{q_0}\right)^2(q_0,b_{\mathfrak{f}}^2)^{\frac{1}{2}}(q_0,{b_{\mathfrak{f}'}}^2)^{\frac{1}{2}}q^{-\frac{5}{4}+\epsilon}.\tag{48}$$

In the case when $b_{\mathfrak{f}}=b_{\mathfrak{f}'}$ and $\mathfrak{f}(\xi,-\zeta)\neq\mathfrak{f}'(\xi',-\zeta')$, we prove a better bound for $\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')$. This will be needed in the next section.

Lemma 3.10. *If $b_{\mathfrak{f}}=b_{\mathfrak{f}'}$, then*

$$|\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')|\ll_\epsilon(q_0,b_{\mathfrak{f}}^2)q^{-\frac{9}{8}+\epsilon}\left(\frac{q}{q_0}\right)^{\frac{17}{8}}\left|\mathfrak{f}(\xi,-\zeta)-\mathfrak{f}'(\xi',-\zeta')\right|^{\frac{1}{2}}$$

.

Proof. If $b_{\mathfrak{f}}=b_{\mathfrak{f}'}$, then $q_1=q_1'=\frac{q_0}{(q_0,b_{\mathfrak{f}}^2)}$. From (47) we have

$$|\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')|=\frac{(q_0,b_{\mathfrak{f}}^2)}{q_0^2}\cdot\frac{q}{q_0}\Bigg|\sum_{r(q_0)}{}'\left(\frac{A_{\mathfrak{f}}A_{\mathfrak{f}'}}{q_1}\right)$$

$$\times e_{q_0}\left(-\overline{8u_0b_1A_{\mathfrak{f}}}\frac{q_0}{q_1}\left(\frac{q_1(A_{\mathfrak{f}}\zeta-B_{\mathfrak{f}}\xi)}{q_0}\right)^2\bar{r}\right)\times e_{q_0}\left(\overline{8u_0b_1'A_{\mathfrak{f}'}}\frac{q_0}{q_1'}\left(\frac{q_1'(A_{\mathfrak{f}'}\zeta'-B_{\mathfrak{f}'}\xi')}{q_0}\right)^2\bar{r}\right)$$

$$\times e_{q_0}(-\overline{4u_0rA_{\mathfrak{f}}}\xi^2+\overline{4u_0rA_{\mathfrak{f}'}\xi'^2})\Bigg|$$

Clearly the term $|\cdot|$ is multiplicative. We apply the Kloostrman 3/4 bound to $|\cdot|$ using the $\bar{r}$ coefficient:

$$|\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')| \ll_\epsilon \frac{(q_0,b_\mathfrak{f}^2)}{q_0^2}\left(\frac{q}{q_0}\right)q_0^{\frac{3}{4}+\epsilon}$$
$$\times \prod_{p^j||q_0}\left(p^j,-\bar{A}_\mathfrak{f}\xi^2-\overline{2b_1A_\mathfrak{f}}\frac{q_0}{q_1}L^2+\bar{A}_{\mathfrak{f}'}\xi'^2+\overline{2b_1A_{\mathfrak{f}'}}\frac{q_0}{q_1}L'^2\right)^{\frac{1}{4}} \tag{49}$$

where $L=\frac{q_1(A_\mathfrak{f}\xi-B_\mathfrak{f}\zeta)}{q_0}$ and $L'=\frac{q_1'(A_{\mathfrak{f}'}\xi'-B_{\mathfrak{f}'}\zeta')}{q_0}$. Now we divide the set of all the primes dividing q_0 into two sets $\mathcal{P}_1$ and $\mathcal{P}_2$, where $\mathcal{P}_1$ contains primes p such that

$$\bar{A}_\mathfrak{f}\xi^2+\overline{2b_1A_\mathfrak{f}}\frac{q_0}{q_1}L^2\equiv\bar{A}_{\mathfrak{f}'}\xi'^2+\overline{2b_1A_{\mathfrak{f}'}}\frac{q_0}{q_1}L'^2(p^{[j/2]})$$

and $\mathcal{P}_2$ is the complement of $\mathcal{P}_1$.

For $p\in\mathcal{P}_2$, the gcd of p^j and $-\bar{A}_\mathfrak{f}\xi^2-\overline{2b_1A_\mathfrak{f}}\frac{q_0}{q_1}L^2+\bar{A}_{\mathfrak{f}'}\xi'^2+\overline{2b_1A_{\mathfrak{f}'}}\frac{q_0}{q_1}L'^2$ is at most $p^{\frac{j}{2}}$. Therefore,

$$\prod_{p\in\mathcal{P}_2}\left(p^j,-\bar{A}_\mathfrak{f}\xi^2-\overline{2b_1A_\mathfrak{f}}\frac{q_0}{q_1}L^2+\bar{A}_{\mathfrak{f}'}\xi'^2+\overline{2b_1A_{\mathfrak{f}'}}\frac{q_0}{q_1}L'^2\right)\le\prod_{p\in\mathcal{P}_2}p^{\frac{j}{2}}\le q_0^{\frac{1}{2}} \tag{50}$$

For $p\in\mathcal{P}_1$, we have

$$\bar{A}_\mathfrak{f}\xi^2+\overline{2b_1A_\mathfrak{f}}\frac{q_0}{q_1}L^2\equiv\bar{A}_\mathfrak{f}\xi^2+\overline{2b_\mathfrak{f}^2}(A_\mathfrak{f}\zeta-B_\mathfrak{f}\xi)^2\equiv\overline{2b_\mathfrak{f}^2}\tilde{\mathfrak{f}}(\xi,-\zeta)(\mathrm{mod}\ p^{[\frac{j}{2}]}). \tag{51}$$

Similarly,

$$\bar{A}_{\mathfrak{f}'}\xi'^2+\overline{2b_1A_{\mathfrak{f}'}}\frac{q_0}{q_1}L^2\equiv\overline{2b_{\mathfrak{f}'}^2}\tilde{\mathfrak{f}'}(\xi',-\zeta')(\mathrm{mod}\ p^{[\frac{j}{2}]}) \tag{52}$$

Since $b_\mathfrak{f}=b_{\mathfrak{f}'}$, we have $\mathfrak{f}(\xi,-\zeta)\equiv\mathfrak{f}'(\xi',-\zeta')(\mathrm{mod}\ p^{\frac{j}{2}})$ for every $p\in\mathcal{P}_1$. Thus we have

$$\prod_{p\in\mathcal{P}_1}\left(p^j,-\bar{A}_\mathfrak{f}\xi^2-\overline{2b_1A_\mathfrak{f}}\frac{q_0}{q_1}L^2+\bar{A}_{\mathfrak{f}'}\xi'^2+\overline{2b_1A_{\mathfrak{f}'}}\frac{q_0}{q_1}L'^2\right)\le\prod_{p\in\mathcal{P}_1}p^j\ll|\mathfrak{f}(\xi,-\zeta)-\mathfrak{f}'(\xi',-\zeta')|^2 \tag{53}$$

Plugging (50) and (53) back into (49) we obtain our lemma. □

Now we go back to $\mathcal{I}_Q$. Again by non-stationary phase the sum is supported on the terms $\xi,\xi',\zeta,\zeta'\ll U$. Using (48) we have

$$\mathcal{I}_Q\ll_\epsilon\frac{N^\epsilon X^4U^4}{TX^2}\sum_{\mathfrak{f},\mathfrak{f}'\in\mathfrak{F}}\sum_{Q\le q\le 2Q}(b_\mathfrak{f}-b_{\mathfrak{f}'},q)^{\frac{1}{4}}\left(\frac{q}{q_0}\right)^2(q_0,b_\mathfrak{f}^2)^{\frac{1}{2}}(q_0,b_{\mathfrak{f}'}^2)^{\frac{1}{2}}q^{-\frac{5}{4}}$$
$$\ll_\epsilon\frac{N^\epsilon X^2U^8}{T}\sum_{\mathfrak{f},\mathfrak{f}'\in\mathfrak{F}}\sum_{Q\le q\le 2Q}(b_\mathfrak{f}-b_{\mathfrak{f}'},q)^{\frac{1}{4}}(q_0,b_\mathfrak{f}^2)^{\frac{1}{2}}(q_0,b_{\mathfrak{f}'}^2)^{\frac{1}{2}}q^{-\frac{5}{4}} \tag{54}$$

We further split (54) into two parts according to $b_\mathfrak{f}=b_{\mathfrak{f}'}$ or not:

$$\mathcal{I}_Q\le\mathcal{I}_Q^{(=)}+\mathcal{I}_Q^{(\neq)}.$$

We first deal with $\mathcal{I}_Q^{(=)}$. Noticing that $\frac{q}{q_0} \leq U$, we have

$$\begin{aligned}\mathcal{I}_Q^{(=)} &\ll_\epsilon \frac{N^\epsilon X^2 U^8}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{Q\leq q\leq 2Q} \frac{(q, b_{\mathfrak{f}}^2)}{q} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}} 1 \\ &\ll_\epsilon \frac{N^\epsilon X^2 U^8}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{a|b_{\mathfrak{f}}^2} a \sum_{Q\leq q\leq 2Q} \mathbf{1}_{a|q} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}} 1 \\ &\ll_\epsilon \frac{N^\epsilon X^2 U^8}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}} 1 \end{aligned} \tag{55}$$

For the last sum above, we introduce Lemma 5.2 from [4]:

Lemma 3.11 (Bourgain, Kontorovich)**.** *There exists a positive constant $\mathcal{C}$ and there exists some $\eta_0 > 0$ which only depend on the spectral gap of Γ such that for any $1 \leq q < N$ and any $r(mod\ q)$,*

$$\sum_{\gamma\in\mathfrak{F}} \mathbf{1}_{\langle e_1, \gamma\mathbf{r}\rangle \equiv r(mod\ q)} \ll \frac{T^\delta}{q^{\eta_0}}.$$

The implied constant is independent of r.

Now we can finally determine K_0, Q_0 and U. We set

$$Q_0 = T^{\frac{\delta-\Theta}{20}}, K_0 = Q_0^2, U = Q_0^{\frac{\eta_0^2}{100}}. \tag{56}$$

Apply Lemma 3.11 to (55), then we get

$$\mathcal{I}_Q^{(=)} \ll_\epsilon N^{-\eta_0+\epsilon} T^{2\delta-1} X^2 U^9 \ll_\eta T^{2\delta-1} X^2 N^{-\eta} \tag{57}$$

which is a power saving.

Now we deal with $\mathcal{I}_Q^{(\neq)}$. We introduce a parameter H which is small power of N. We further split $\mathcal{I}_Q^{(\neq)}$ into $\mathcal{I}_Q^{\neq,>} + \mathcal{I}_Q^{(\neq,\leq)}$ according to $(b_{\mathfrak{f}}, b_{\mathfrak{f}'}) > H$ or not. We first handle big gcd.

Lemma 3.12.

$$\mathcal{I}_Q^{(\neq,>)} \ll_\eta N T^{2(\delta-1)} N^{-\eta}$$

Proof. Apply (54) and replace $(q_0, b_{\mathfrak{f}}^2)$ by $(q, b_{\mathfrak{f}}^2)$ and $(b_{\mathfrak{f}} - b_{\mathfrak{f}'}, q)$ by q:

$$\begin{aligned}\mathcal{I}_Q^{(\neq,>)} &\ll_\epsilon \frac{N^\epsilon X^2 U^8}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\(b_{\mathfrak{f}},b_{\mathfrak{f}'})>H}} \sum_{Q\le q\le 2Q} \frac{(q,b_{\mathfrak{f}}^2)^{\frac12}(q,b_{\mathfrak{f}'}^2)^{\frac12}}{q}\\ &\ll_\epsilon \frac{N^\epsilon X^2 U^8}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\substack{h\mid b_{\mathfrak{f}}^2\\ h>H}} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}\equiv 0(h)}} \sum_{Q\le q\le 2Q} \frac{(q,b_{\mathfrak{f}}^2)^{\frac12}(q,b_{\mathfrak{f}'}^2)^{\frac12}}{q}\\ &\ll_\epsilon \frac{N^\epsilon X^2 U^8}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\substack{h\mid b_{\mathfrak{f}}^2\\ h>H}} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}\equiv 0(h)}} \sum_{\tilde{q_1}\mid b_{\mathfrak{f}}^2} \sum_{\tilde{q_1}'\mid b_{\mathfrak{f}'}^2} (\tilde{q_1}\tilde{q_1}')^{\frac12} \sum_{\substack{Q\le q\le 2Q\\ [\tilde{q_1},\tilde{q_1}']\mid q}} \mathbf{1} \end{aligned} \tag{58}$$

Now since $[\tilde{q_1}, \tilde{q_1}'] > (\tilde{q_1}\tilde{q_1}')^{\frac12}$, the above

$$\ll_\epsilon \frac{N^\epsilon X^2 U^9}{T} \sum_{\mathfrak{f}\in\mathfrak{F}} \sum_{\substack{h\mid b_{\mathfrak{f}}^2\\ h>H}} \sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}\equiv 0(h)}} \sum_{\tilde{q_1}\mid b_{\mathfrak{f}}^2} \sum_{\tilde{q_1}'\mid b_{\mathfrak{f}'}^2} 1 \tag{59}$$

From Lemma 3.11, we have $\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}\equiv 0(h)}} 1 \ll \frac{T^\delta}{H^{\eta_0}}$. We set $H = Q_0^{\frac{\eta_0}{10}}$. Therefore,

$$(59) \ll_\epsilon \frac{N^\epsilon X^2 U^9 T^{2\delta}}{T H^{\eta_0}} \ll_\epsilon \frac{N^\epsilon T^{2\delta-1} X^2 U^9}{H^{\eta_0}} \ll_\eta T^{2\delta-1} X^2 N^{-\eta}$$

which is a power saving. □

Next we deal with small gcd. We write $(b_{\mathfrak{f}}, b_{\mathfrak{f}'}) = h$ and $b_{\mathfrak{f}} = hg_1, b_{\mathfrak{f}'} = hg_2, b_{\mathfrak{f}} - b_{\mathfrak{f}'} = hg_3$. Then g_1, g_2, g_3 are mutually relatively prime. We have

Lemma 3.13.

$$\mathcal{I}_Q^{(\neq,\le)} \ll_\eta N^{1-\eta} T^{2(\delta-1)}$$

Proof. From (54),

$$\begin{aligned}
\mathcal{I}_Q^{(\neq,\leq)} &\ll_\epsilon \frac{N^\epsilon X^2U^8}{T}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\(b_{\mathfrak{f}'},b_{\mathfrak{f}})\leq H}}\sum_{Q\leq q\leq 2Q}\frac{(q_0,b_{\mathfrak{f}}^2)^{\frac12}(q_0,b_{\mathfrak{f}'}^2)^{\frac12}(b_{\mathfrak{f}}-b_{\mathfrak{f}'},q)^{\frac14}}{q^{\frac54}}\\
&\ll_\epsilon \frac{N^\epsilon X^2U^8}{TQ^{\frac54}}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\(b_{\mathfrak{f}'},b_{\mathfrak{f}})\leq H}}\sum_{Q\leq q\leq 2Q}(q_0,b_{\mathfrak{f}})(q_0,b_{\mathfrak{f}'})(b_{\mathfrak{f}}-b_{\mathfrak{f}'},q)^{\frac14}\\
&\ll_\epsilon \frac{N^\epsilon X^2U^8}{TQ^{\frac54}}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\(b_{\mathfrak{f}'},b_{\mathfrak{f}})\leq H}}\sum_{h|(b_{\mathfrak{f}},b_{\mathfrak{f}'})}h^{\frac94}\sum_{g_1|b_{\mathfrak{f}}}\sum_{g_2|b_{\mathfrak{f}'}}\sum_{\substack{g_3|b_{\mathfrak{f}}-b_{\mathfrak{f}'}\\g_3\ll Q}}g_1g_2g_3^{\frac14}\sum_{\substack{Q\leq q\leq 2Q\\[hg_1,hg_2,hg_3]|q}}1\\
&\ll_\epsilon \frac{N^\epsilon X^2U^8H^{\frac94}}{TQ^{\frac54}}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\(b_{\mathfrak{f}'},b_{\mathfrak{f}})\leq H}}\sum_{g_1|b_{\mathfrak{f}}}\sum_{g_2|b_{\mathfrak{f}'}}\sum_{\substack{g_3|b_{\mathfrak{f}}-b_{\mathfrak{f}'}\\g_3\ll Q}}g_1g_2g_3^{\frac14}\frac{q}{g_1g_2g_3}\\
&\ll_\epsilon \frac{N^\epsilon X^2U^8H^{\frac94}}{TQ^{\frac14}}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\(b_{\mathfrak{f}'},b_{\mathfrak{f}})\leq H}}\sum_{\substack{g_3|b_{\mathfrak{f}}-b_{\mathfrak{f}'}\\g_3\ll Q}}g_3^{-\frac34}\\
&\ll_\epsilon \frac{N^\epsilon X^2U^8H^{\frac94}}{TQ^{\frac14}}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{g_3\ll Q}g_3^{-\frac34}\sum_{\substack{\mathfrak{f}\in\mathfrak{F}\\b_{\mathfrak{f}'}\equiv b_{\mathfrak{f}}(g_3)}}1
\end{aligned}$$

By Lemma 3.11, $\sum_{\mathfrak{f}'\in\mathfrak{F}}\mathbf{1}_{b_{\mathfrak{f}'}\equiv b_{\mathfrak{f}}(g_3)}\ll\frac{T^\delta}{g_3^{\eta_0}}$. Therefore,

$$\mathcal{I}_Q^{(\neq,\leq)}\ll_\epsilon \frac{N^\epsilon X^2U^9H^{\frac94}}{TQ^{\frac14}}\sum_{\mathfrak{f}\in\mathfrak{F}}T^\delta Q^{\frac14-\eta_0}\ll_\epsilon N^\epsilon T^{2\delta-1}X^2U^9H^{\frac94}Q_0^{-\eta_0}\ll_\eta T^{2\delta-1}X^2N^{-\eta}$$

Again we have a power savings for $\mathcal{I}_Q^{(\neq,\leq)}$. □

In summary, we have

Lemma 3.14.

$$\mathcal{I}_2\ll_\eta N^{1-\eta}T^{2(\delta-1)}$$

3.5. **Minor Arc Analysis III.** In this section we deal with the last part of the integral, which is on the minor arcs corresponding to $X<q<M$, namely $\mathcal{I}_3$. We keep all the notations from the previous sections. Return to (32), and again for simplicity we restrict our attention on the summands of $\mathcal{R}_N^U$ where u even:

$$\begin{aligned}
\mathcal{R}_{u,\mathfrak{f}}\left(\frac{r}{q}+\beta\right)&=\sum_{x,y\in\mathbb{Z}}\psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)e\left(\mathfrak{f}(xu,yu)\left(\frac{r}{q}+\beta\right)\right)\\
&=e\left(-\left(\frac{r}{q}+\beta\right)b_{\mathfrak{f}}\right)\sum_{x,y\in\mathbb{Z}}\psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)e\left(\frac{ru_0\tilde{\mathfrak{f}}(x,y)}{q_0}\right)e(\tilde{\mathfrak{f}}(x,y)u^2\beta)
\end{aligned}\tag{60}$$

Now we rewrite $e_{q_0}(ru_0\tilde{\mathfrak{f}}(x,y))$ into its Fourier expansion. We have

$$\begin{aligned} e_{q_0}(ru_0\tilde{\mathfrak{f}}(x,y)) &= \frac{1}{q_0^2}\sum_{m(q_0)}\sum_{n(q_0)}\sum_{l(q_0)}\sum_{t(q_0)} e_{q_0}(ru_0\tilde{\mathfrak{f}}(l,t)+lm+tn)e_{q_0}(-mx-ny) \\ &= \sum_{m(q_0)}\sum_{n(q_0)} \mathcal{S}_{\mathfrak{f}}(q_0,u_0r,m,n)e_{q_0}(-mx-ny) \end{aligned}$$

Therefore,

$$\mathcal{R}_{u,\mathfrak{f}}\left(\frac{r}{q}+\beta\right) = e_q(-rb_{\mathfrak{f}})\sum_{m(q_0)}\sum_{n(q_0)}\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,m,n)\lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right),$$

where

$$\lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right) := \sum_{x,y\in\mathbb{Z}}\psi\left(\frac{xu}{X}\right)\psi\left(\frac{yu}{X}\right)e\left(-\frac{mx}{q_0}\right)e\left(-\frac{ny}{q_0}\right)e(\mathfrak{f}(xu,yu)\beta).$$

We apply the Cauchy-Schwarz inequality to the u variable for $\mathcal{I}_Q$:

$$\begin{aligned} \mathcal{I}_Q &= \sum_{Q\le q\le 2Q}{\sum_{r(q)}}' \int_{-\frac{1}{qM}}^{\frac{1}{qM}} \left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 d\beta \\ &\ll U\sum_{u<U}\sum_{Q\le q\le 2Q}{\sum_{r(q)}}' \int_{-\frac{1}{qM}}^{\frac{1}{qM}} \left|\sum_{\mathfrak{f}\in\mathfrak{F}}\mathcal{R}_{u,\mathfrak{f}}\left(\frac{r}{q}+\beta\right)\right|^2 d\beta \\ &\ll U\sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\mathfrak{f}'\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\sum_{m,n,m',n'(q_0)}\left(\sum_{r(q)}{}'\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,m,n)\overline{\mathcal{S}_{\mathfrak{f}'}(q_0,u_0r,m',n')}e_q(r(-b_{\mathfrak{f}}+b_{\mathfrak{f}'}))\right) \\ &\times \int_{-\frac{1}{qM}}^{\frac{1}{qM}} \lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right)\overline{\lambda_{\mathfrak{f}'}\left(X,\beta;\frac{m'}{q_0},\frac{n'}{q_0},u\right)}d\beta \end{aligned}$$

Since m,n,m',n' comes from congruence classes (mod q_0), we can choose representatives such that m,n,m',n' with absolute values bounded by $\frac{q_0}{2}$. The main contribution of $\mathcal{I}_Q$ comes from the terms $m,n,m',n' \ll \frac{uq_0}{X}$ by non-stationary phase. To see this, for the terms with any of $m,n,m',n' \gg \frac{uq_0}{X}$ (let's say $m \gg \frac{uq_0}{X}$), we use Poisson summation to rewrite $\lambda_{\mathfrak{f}}$:

$$\begin{aligned} \lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right) &= \frac{X^2}{u^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(-\frac{mX}{q_0u}x\right)e\left(-\frac{nX}{q_0u}y\right)e(\mathfrak{f}(xX,yX)\beta)dxdy \\ &+ \sum_{\substack{\xi,\zeta\in\mathbb{Z}\\(\xi,\zeta)\neq(0,0)}}\frac{X^2}{u^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(\left(\frac{\xi X}{u}-\frac{mX}{q_0u}\right)x\right)e\left(\left(\frac{\zeta X}{u}-\frac{nX}{q_0u}\right)y\right)e(\mathfrak{f}(xX,yX)\beta)dxdy \end{aligned}$$

If $\xi\neq 0$, since $\frac{mX}{q_0u}\le\frac{X}{2u}$ and $\mathfrak{f}(xX,yX)\beta\ll 1$, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(\left(\frac{\xi X}{u}-\frac{mX}{q_0u}\right)x\right)e\left(\left(\frac{\zeta X}{u}-\frac{nX}{q_0u}\right)y\right)e(\mathfrak{f}(xX,yX)\beta)dxdy \ll \left(\frac{u}{X\xi}\right)^{N_0}$$

for any $N_0 > 0$, by first applying non-stationary phase to the x variable and trivially bounding the y integral. From this, one gets

$$\lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right)=\frac{X^2}{u^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(-\frac{mX}{q_0u}\right)e\left(-\frac{nX}{q_0u}\right)e(\mathfrak{f}(xX,yX)\beta)dxdy \tag{61}$$

$$+O\left(\left(\frac{u}{X}\right)^{N_0}\right) \tag{62}$$

for any $N_0 > 0$. We use non-stationary phase again to treat the above integral, then we obtain

$$\left|\lambda_{\mathfrak{f}}\left(X,\beta,\frac{m}{q_0},\frac{n}{q_0},u\right)\right|\ll\frac{X^2}{u^2}\min\left\{\left(\frac{uq_0}{Xm}\right)^{2N_0},\left(\frac{uq_0}{Xn}\right)^{2N_0}\right\}\ll\frac{X^2}{u^2}\left(\frac{uq_0}{X}\right)^{2N_0}\frac{1}{m^{N_0}n^{N_0}}.$$

Therefore, we have

$$\int_{-\frac{1}{qM}}^{\frac{1}{qM}}\lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right)\overline{\lambda_{\mathfrak{f}'}\left(X,\beta;\frac{m'}{q_0},\frac{n'}{q_0},u\right)}d\beta\ll\frac{1}{QM}\frac{X^4}{u^4}\left(\frac{uq_0}{X}\right)^{4N_0}\frac{1}{m^{N_0}n^{N_0}m'^{N_0}n'^{N_0}}$$

Now we use (48) to bound $|\mathcal{S}|$, we thus have

$$\begin{aligned}
&U\sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\mathfrak{f}'\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\sum_{m,n,m',\text{ or }n'\gg\frac{uq_0}{X}}\left(\sideset{}{'}\sum_{r(q)}\mathcal{S}_{\mathfrak{f}}(q_0,u_0r,m,n)\mathcal{S}_{\mathfrak{f}'}(q_0,u_0r,m',n')e_q(r(-b_{\mathfrak{f}}+b_{\mathfrak{f}'}))\right)\\
&\times\int_{-\frac{1}{qM}}^{\frac{1}{qM}}\lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right)\overline{\lambda_{\mathfrak{f}'}\left(X,\beta;\frac{m'}{q_0},\frac{n'}{q_0},u\right)}d\beta\\
&\ll_\epsilon N^\epsilon UT^{2\delta}\sum_{u<U}\sum_{Q\le q\le 2Q}\frac{T^{\frac{9}{4}}u^4}{Q^{\frac{5}{4}}}\frac{1}{QM}\frac{X^4}{u^4}\left(\frac{uq_0}{X}\right)^{4N_0}\sum_{m,n,m',\text{or }n'\gg\frac{uq_0}{X}}\frac{1}{m^{N_0}n^{N_0}m'^{N_0}n'^{N_0}}
\end{aligned}\tag{63}$$

If We set $N_0 = 5$, then the above

$$\ll N^\epsilon U^{20}T^{2\delta+\frac{63}{4}}X^{\frac{7}{4}}$$

Thus we see $|\mathcal{S}|$ is indeed mainly supported on $m,n,m',n'\ll\frac{uq_0}{X}$. Now we split the terms $m,n,m',n'\ll\frac{uq_0}{X}$ into two parts according to whether $b_{\mathfrak{f}}=b_{\mathfrak{f}'}$ or not:

$$\mathcal{I}_Q\ll\mathcal{I}_Q^{(=)}+\mathcal{I}_Q^{(\neq)},$$

where

$$\begin{aligned}
\mathcal{I}_Q^{(=)}=&\sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}}\sum_{Q\le q\le 2Q}\sum_{m,n,m',n'\ll\frac{uq_0}{X}}\mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')\\
&\times\int_{-\frac{1}{qM}}^{\frac{1}{qM}}\lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right)\overline{\lambda_{\mathfrak{f}'}\left(X,\beta;\frac{m'}{q_0},\frac{n'}{q_0},u\right)}d\beta
\end{aligned}\tag{64}$$

and

$$\mathcal{I}_Q^{(\neq)} = \sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}\neq b_{\mathfrak{f}}}}\sum_{Q\leq q\leq 2Q}\sum_{m,n,m',n'\ll\frac{uq_0}{X}} \mathcal{S}(q,q_0,u_0,\xi,\zeta,\mathfrak{f},\xi',\zeta',\mathfrak{f}')$$
$$\times\int_{-\frac{1}{qM}}^{\frac{1}{qM}} \lambda_{\mathfrak{f}}\left(X,\beta;\frac{m}{q_0},\frac{n}{q_0},u\right)\overline{\lambda_{\mathfrak{f}'}\left(X,\beta;\frac{m'}{q_0},\frac{n'}{q_0},u\right)}d\beta \tag{65}$$

For λ, since the sum is supported on $x,y\asymp\frac{X}{u}$, λ has a trivial bound $\frac{X^2}{u^2}$. Therefore, for $\square\in\{=,\neq\}$, we have

$$\mathcal{I}_Q^{\square} \ll \frac{UX^4}{QM}\sum_{u<U}\frac{1}{u^4}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}\square b_{\mathfrak{f}}}}\sum_{Q\leq q\leq 2Q}\sum_{m,n,m',n'\ll\frac{uq_0}{X}}|\mathcal{S}|. \tag{66}$$

If $b_{\mathfrak{f}}\neq b_{\mathfrak{f}'}$, then we could use the bound from (48) to estimate $\mathcal{S}$. We have

$$\begin{aligned}\mathcal{I}_Q^{(\neq)} &\ll_\epsilon \frac{UX^4}{QM}\sum_{u<U}\frac{1}{u^4}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}}\sum_{Q\leq q\leq 2Q}\sum_{m,n,m',n'}(b_{\mathfrak{f}}-b_{\mathfrak{f}'},q)^{\frac14}\left(\frac{q}{q_0}\right)^2(q_0,b_{\mathfrak{f}}^2)^{\frac12}(q_0,b_{\mathfrak{f}'}^2)^{\frac12}q^{-\frac54+\epsilon}\\ &\ll_\epsilon \frac{N^\epsilon UX^4}{QM}\sum_{u<U}\frac{1}{u^4}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}}\neq b_{\mathfrak{f}'}}}\sum_{Q\leq q\leq 2Q}\left(\frac{uq_0}{X}\right)^4T^{\frac94}u^4Q^{-\frac54}\ll_\epsilon T^{2(\delta-1)}N^{1+\epsilon}(T^6X^{-\frac14}U^6)\end{aligned} \tag{67}$$

where we replaced $(b_{\mathfrak{f}}-b_{\mathfrak{f}'},q),(q_0,b_{\mathfrak{f}}^2)^{\frac12}$ and $(q_0,b_{\mathfrak{f}'}^2)^{\frac12}$ by T. Thus we have a significant power saving for $\mathcal{I}_Q^{(\neq)}$.

Next we deal with $\mathcal{I}_Q^{(=)}$, we further split $\mathcal{I}_Q^{(=)}$ into two pieces

$$\mathcal{I}_Q^{(=)} = \mathcal{I}_Q^{(=,=)} + \mathcal{I}_Q^{(=,\neq)}$$

according to whether $\mathfrak{f}(m,-n)=\mathfrak{f}'(m',-n')$ or not. For $\mathcal{I}_Q^{(=,\neq)}$, we use Lemma 3.10 to bound $|\mathcal{S}|$. We have

$$\mathcal{I}_Q^{(=,\neq)} \ll \frac{UX^4}{QM}\sum_{u<U}\sum_{Q\leq q\leq 2Q}\sum_{m,n,m',n'\ll\frac{uq_0}{X}}\sum_{\substack{\mathfrak{f},\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}}=b_{\mathfrak{f}'}\\ \mathfrak{f}(m,-n)\neq\mathfrak{f}'(m',-n')}}(q_0,b_{\mathfrak{f}}^2)q^{-\frac98+\epsilon}\left(\frac{q}{q_0}\right)^{\frac{17}{8}}\left|\mathfrak{f}(m,-n)-\mathfrak{f}'(m',-n')\right|^{\frac12} \tag{68}$$

Noticing that $(q_0,b_{\mathfrak{f}}^2)\ll T^2,\frac{q}{q_0}\ll U^2$ and $\mathfrak{f}(m,-n),\mathfrak{f}'(m',-n')\ll T(\frac{UQ}{X})^2$, we have

$$\mathcal{I}_Q^{(=,\neq)} \ll_\epsilon N^\epsilon\frac{UX^4}{QM}UQ\left(\frac{UQ}{X}\right)^4T^{2\delta}\frac{T^2}{Q^{\frac98}}U^{\frac{17}{4}}T^{\frac12}\frac{UQ}{X}\ll_\epsilon N^\epsilon U^{\frac{45}{4}}T^{2\delta+\frac{43}{8}}X^{\frac{15}{8}}, \tag{69}$$

which is again a significant power saving.

Next we deal with $\mathcal{I}_Q^{(=,=)}$. This will complete our minor arc analysis. From (48) and (66) we have

$$\mathcal{I}_Q^{(=,=)} \ll \frac{UX^4}{QM}\sum_{u<U}\frac{1}{u^4}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\sum_{m,n\ll\frac{uq_0}{X}}\frac{(b_{\mathfrak{f}}^2,q)}{q}u^4\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}}\sum_{\substack{m',n'\ll\frac{uq_0}{X}\\ \mathfrak{f}'(m',-n')=\mathfrak{f}(m,-n)}}1$$

For the inner double sum we shall prove the following lemma:

Lemma 3.15.

$$\sum_{\substack{\mathfrak{f}'\in\mathfrak{F}\\ b_{\mathfrak{f}'}=b_{\mathfrak{f}}}}\sum_{\substack{m',n'\ll\frac{uq_0}{X}\\ \mathfrak{f}'(m',-n')=\mathfrak{f}(m,-n)}}1 \ll_\epsilon N^\epsilon\left(\tilde{\mathfrak{f}}(m,-n),-8b_{\mathfrak{f}}^2\right)^{\frac{1}{2}}$$

Proof. This lemma will follow from the following three claims.

Claim 1: The number of classes of equivalent quadratic forms having discriminant $-8b_{\mathfrak{f}}^2$ and representing the integer $z=\tilde{\mathfrak{f}}(m,-n)$ is bounded by $N^\epsilon(\tilde{\mathfrak{f}}(m,-n),-8b_{\mathfrak{f}}^2)^{\frac{1}{2}}$.

Suppose $z=\tilde{\mathfrak{f}}(m,-n)$ is primitively represented by a quadratic form $\mathfrak{f}_0$ (i.e. $(m,n)=1$), then $\mathfrak{f}_0$ is equivalent to a quadratic form $zx^2+B_0xy+C_0y^2$, with $|B_0|<z$. Now since $B_0^2-4zC_0=-8b_{\mathfrak{f}}^2$, we have $B_0^2\equiv -8b_{\mathfrak{f}}^2(z)$. From the Chinese Remainder Theorem, the number of solutions of

$$B_0^2\equiv -8b_{\mathfrak{f}}^2(z) \tag{70}$$

is the the product of the numbers of solutions of

$$B_0^2\equiv -8b_{\mathfrak{f}}^2(p_i^{n_i}) \tag{71}$$

for each $p_i^{n_i}||z$.

If $\left(\frac{-2}{p_i^{n_i}}\right)=-1$, then there's no solution to (71). If $\left(\frac{-2}{p_i^{n_i}}\right)=1$, let $-8b_{\mathfrak{f}}^2\equiv kp_i^{l_i}(p_i^{n_i})$ where $0\le l_i\le n_i$ and $(k,p_i)=1$. Noticing that l_i is even, all the solutions of (71) are given by

$$\pm p^{\frac{l_i}{2}}l+p^{n_i-\frac{l_i}{2}}s,$$

where l is a solution of

$$l^2\equiv\frac{-8b_{\mathfrak{f}}^2}{p_i^{l_i}}(p^{n_i-l_i}) \tag{72}$$

and $0\le s\le p^{\frac{l_i}{2}}-1$. Thus we see there are at most $2p^{\frac{l_i}{2}}$ such solutions to (71). By multiplicativity, the number of solutions of (70) is bounded by $2^{w(\tilde{\mathfrak{f}}(m,-n))}(\tilde{\mathfrak{f}}(m,-n),-8b_{\mathfrak{f}}^2)^{\frac{1}{2}}$. Therefore, our choices for B_0 is at most $2^{w(\tilde{\mathfrak{f}}(m,-n))+1}(\tilde{\mathfrak{f}}(m,-n),-8b_{\mathfrak{f}}^2)^{\frac{1}{2}}$. If z is not primitively represented by $\mathfrak{f}_0$, then a divisor z_0 of z is primitively represented. There are at most $d(z)$ many such cases, and the bound $2^{w(\tilde{\mathfrak{f}}(m,-n))+1}(\tilde{\mathfrak{f}}(m,-n),-8b_{\mathfrak{f}}^2)^{\frac{1}{2}}$ works for each case. Thus Claim 1 follows.

Claim 2: In each equivalent class in $\mathfrak{F}$, the number of equivalent quadratic forms is bounded: Suppose $\mathfrak{f}^{'}=(A^{'},2B^{'},C^{'})$ and $\mathfrak{f}^{''}=(A^{''},2B^{''},C^{''})$ are two equivalent quadratic forms in $\mathfrak{F}$, then we can find $\begin{pmatrix} g & h \\ i & j \end{pmatrix} \in SL(2,\mathbb{Z}) \cup \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} SL(2,\mathbb{Z})$ such that

$$\begin{aligned} A^{''} &= g^2A^{'}+2giB^{'}+i^2C^{'}, \\ B^{''} &= ghA^{'}+(gi+hj)B^{'}+ijC^{'}, \\ C^{''} &= h^2A^{'}+2hjB^{'}+j^2C^{'} \end{aligned} \tag{73}$$

The first equation above can be rewritten as $A^{''}=A^{'}\left(g+i\frac{B^{'}}{A^{'}}\right)^2+i^2\frac{2b_{\mathfrak{f}}^2}{A^{'}}$. So from $i^2\frac{2b_{\mathfrak{f}}^2}{A^{'}}\leq A^{''}$, $b_{\mathfrak{f}}\asymp T$, $A^{'},A^{''}\ll T$, we know $i\ll 1$, and $A^{'},A^{''}\gg T$. Then from $A^{'}\left(g+i\frac{B^{'}}{A^{'}}\right)^2\leq A^{''}\ll T$ we also know $g\ll 1$. Similarly $h,j\ll 1$, so the number of quadratic forms in $\mathfrak{F}$ in each equivalent class is bounded. Therefore Claim 2 holds.

Claim 3: given an integer $z\ll N$ and a quadratic form $\mathfrak{f}$ of discriminant $-8b_{\mathfrak{f}}^2$, there are at most N^{ϵ} pairs of integers m,n such that $\mathfrak{f}(m,-n)=z$.

This is because $Am^2-2Bmn+Cn^2=z$ can be rewritten as

$$(Am+(B+\sqrt{-2}b_{\mathfrak{f}})n)(Am+(B-\sqrt{-2}b_{\mathfrak{f}})n)=Az$$

Since $Az\ll N^2$, the number of divisors of Az in $\mathbb{Z}[\sqrt{2}i]$ is bounded by N^{ϵ}. The pairs (m,n) can be identified with $Am+(B+\sqrt{-2}b_{\mathfrak{f}})n$, which is a divisor of Az. Therefore, Claim 3 also holds.

Our lemma then follows Claims 1, Claim 2 and Claim 3. □

We need the following final ingredient to estimate $\mathcal{I}_Q^{(=,=)}$:

Lemma 3.16. *Given a primitive quadratic form $(A,2B,C)$ of discriminant $-8b_{\mathfrak{f}}^2$, for any $d|2b_{\mathfrak{f}}^2$, and any integer $W>0$, we have*

$$\sum_{\substack{m,n\leq W \\ Am^2-2Bmn+Cn^2\equiv 0(d)}} 1\ll W^2d^{-\frac{1}{2}}+W$$

The implied constant is absolute.

Proof. First we show that $\exists\gamma=\begin{pmatrix} i & j \\ g & h \end{pmatrix}\in SL(2,\mathbb{Z})$ and $\tilde{A},\tilde{B},\tilde{C}\in\mathbb{Z}$ such that

$$Ax^2+2Bxy+Cy^2=\tilde{A}(ix+gy)^2+\tilde{B}(ij+gh)xy+\tilde{C}(jx+hy)^2$$

and

$$(\tilde{A},-2b_{\mathfrak{f}}^2)=1,\tilde{B}\equiv\tilde{C}\equiv 0(d).$$

Indeed, for each $p_i^{n_i}||d$, since $\mathfrak{f}$ is primitive, at least one of A,B,C can not be divided by p. For example, if $(A,p)=1$, then

$$Ax^2+2Bxy+Cy^2\equiv A(x+B\bar{A}y)^2+2b_{\mathfrak{f}}^2\bar{A}y^2\equiv A(x+B\bar{A}y)^2.$$

We set

$$\gamma_{p_i^{n_i}} := \begin{pmatrix} 1 & B\bar{A} \\ 0 & 1 \end{pmatrix} \in SL(2, \mathbb{Z}/p_i^{n_i}\mathbb{Z})$$

so $\gamma_{p_i^{n_i}}(A, 2B, C) = (A, 0, 0)(p_i^{n_i})$. Now from the Chinese remainder theorem, we could find $\gamma_d \in SL(2, \mathbb{Z}/d\mathbb{Z})$ such that $\gamma_d \equiv \gamma_{p_i^{n_i}}$ in $SL(2, \mathbb{Z}/p_i^{n_i})$ for each $p_i^{n_i}||d$. Since $(\tilde{A}, d) = 1$ and $\tilde{B} \equiv 0(d)$, it forces $\tilde{C} \equiv 0(d)$. Therefore,

$$\sum_{\substack{m,n \le W \\ \mathfrak{f}(m,-n) \equiv 0(d)}} 1 = \sum_{\substack{m,n \le W \\ (im+gn)^2 \equiv 0(d)}} 1$$

If $(im + gn)^2 \equiv 0(d)$, then $im + gn$ can be parametrized by sd_0, where $s \in \mathbb{Z}$ and $d_0 \ge d^{\frac{1}{2}}$. Therefore, we have

$$im + gn \equiv 0(d_0) \tag{74}$$

For the above equation to have a solution, since $(i, g) = 1$, gn should be of the form $k(i, d_0)$ where $k \in \mathbb{Z}$, so there are at most $\frac{W}{(i,d_0)} + 1$ choices for n. Fixing such an n, (74) can be reduced to

$$\frac{i}{(i, d_0)} m \equiv k \left(\mathrm{mod}\ \frac{d_0}{(i, d_0)} \right).$$

There are at most $\frac{W}{\frac{i}{d_0}} + 1$ such choices for m. Therefore,

$$\sum_{\substack{m,n \le W \\ \mathfrak{f}(m,-n) \equiv 0(d)}} 1 = \sum_{\substack{m,n \le W \\ (im+gn)^2 \equiv 0(d)}} 1 \ll \left(\frac{W}{(i, d_0)} + 1 \right) \left(\frac{W}{\frac{d_0}{(i,d_0)}} + 1 \right) \ll W^2 d^{-\frac{1}{2}} + W.$$

□

Now we can show that

Lemma 3.17.

$$\mathcal{I}_Q^{(=,=)} \ll_\eta T^{2\delta - 1} X^2 N^{-\eta}$$

Proof. Applying Lemma 3.15 and Lemma 3.16 to (68) with $W = \frac{uq_0}{X}$, we have

$$\begin{aligned}
\mathcal{I}_Q^{(=,=)} &\ll_\epsilon \frac{N^\epsilon UX^4}{QM}\sum_{u<U}\frac{1}{u^4}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\sum_{m,n\ll\frac{uq_0}{X}}\frac{(b_\mathfrak{f}^2,q)}{q}u^4(\mathfrak{f}(m,-n),-8b_\mathfrak{f}^2)^{\frac12}\\
&\ll_\epsilon \frac{N^\epsilon UX^4}{QM}\sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\frac{(b_\mathfrak{f}^2,q)}{q}\sum_{m,n\ll\frac{uq_0}{X}}(\mathfrak{f}(m,-n),-2b_\mathfrak{f}^2)^{\frac12}\\
&\ll_\epsilon \frac{N^\epsilon UX^4}{QM}\sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\frac{(b_\mathfrak{f}^2,q)}{q}\sum_{d_1|-2b_\mathfrak{f}^2}d_1^{\frac12}\sum_{\substack{m,n\ll\frac{uq_0}{X}\\ \mathfrak{f}(m,-n)\equiv 0(d_1)}}1\\
&\ll_\epsilon \frac{N^\epsilon UX^4}{QM}\sum_{u<U}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\frac{(b_\mathfrak{f}^2,q)}{q}\sum_{d_1|-2b_\mathfrak{f}^2}d_1^{\frac12}\left(\Big(\frac{uq_0}{X}\Big)^2d_1^{-\frac12}+\frac{uq_0}{X}\right)\\
&\ll_\epsilon \frac{N^\epsilon U^4X^4}{QM}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{Q\le q\le 2Q}\frac{(b_\mathfrak{f}^2,q)}{q}\cdot\frac{Tq}{X}\\
&\ll_\epsilon \frac{N^\epsilon U^4X^3T}{QM}\sum_{\mathfrak{f}\in\mathfrak{F}}\sum_{d_2|b_\mathfrak{f}^2}d_2\sum_{\substack{Q\le q\le 2Q\\ q\equiv 0(d_2)}}1\\
&\ll_\epsilon N^\epsilon U^4X^2T^\delta \ll_\epsilon N^\epsilon U^4X^2T^{2\delta-1}T^{1-\delta}
\end{aligned}\tag{75}$$

Therefore, we have a power saving here. □

From (67), (69) and (75) we obtain

Lemma 3.18.

$$\mathcal{I}_3 \ll_\eta T^{2\delta-1}X^2N^{-\eta}.$$

3.6. **Proof of Theorem 1.4.** We are now ready to give the proof of Theorem 1.4 following the strategy at the end of §3.1.

Proof of Theorem 1.4. From Lemma 3.2 we know that

$$\sum_{\frac{n}{2}<n<N}|\mathcal{R}_N(n)-\mathcal{R}_N^U(n)| \ll_\epsilon \frac{T^\delta X^{2+\epsilon}}{U} \ll_\eta T^\delta X^2N^{-\eta}.$$

From Lemma 3.8, Lemma 3.14 and Lemma 3.18 we know that

$$\sum_{\frac{n}{2}<n<N}|\mathcal{E}_N^U(n)|^2 \le \int_{-1}^1|(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2d\theta \ll_\eta T^{2\delta-1}X^2N^{-\eta}.$$

By Cauchy inequality, we then have

$$\sum_{\frac{n}{2}<n<N}|\mathcal{E}_N^U(n)| \ll_\eta T^\delta X^2N^{-\eta}.$$

From Lemma 3.6, we also have

$$\sum_{n<N}|\mathcal{M}_N(n)-\mathcal{M}_N^U(n)| \ll_\eta T^\delta X^2N^{-\eta}.$$

Since $\mathcal{M}_N = \mathcal{R}_N + \mathcal{E}_N$ and $\mathcal{M}_N^U = \mathcal{R}_N^U + \mathcal{E}_N^U$, we then have

$$\sum_{n<N} |\mathcal{E}_N(n) - \mathcal{E}_N^U(n)| \ll_\eta T^{\delta-1} X^2 N^{-\eta}.$$

As a result,

$$\sum_{n<N} |\mathcal{E}_N(n)| \ll_\eta T^{\delta} X^2 N^{-\eta}.$$

Let Z be the exceptional subset of $\{n | n \equiv \kappa_1 (\mathrm{mod}\ 8)\} \cap (\frac{N}{2}, N)$ consisting of all numbers which are not represented by our ensemble $\mathfrak{F}$. Then for $z \in Z$, we have $\mathcal{M}_N(z) \gg_\epsilon N^{-\epsilon} T^{\delta-1}$. Since $\mathcal{R}_N(z) = 0$, we have $|\mathcal{E}_N(z)| \gg_\epsilon N^{-\epsilon} T^{\delta-1}$.

Therefore,

$$|Z| T^{\delta-1} N^{-\epsilon} \ll_\epsilon \sum_{n \in Z} |\mathcal{E}_N(z)| \ll_\eta T^{\delta} X^2 N^{-\eta}.$$

So $|Z| \ll N^{1-\eta}$, and we prove the density one theorem for the C_1-orbit under Γ. There are six orbits in $\mathcal{P}$, namely $C_1, C_2, C_3, C_{1'}, C_{2'}, C_{3'}$. We can prove the same conclusion for every orbit simply by changing the order of components of $\mathbf{r}$ or $\mathbf{r}'$. Thus Theorem 1.4 follows. □

Acknowlegement This paper is essentially the content of the author's PhD thesis when he was a graduate student at Stony Brook. The author has a great many thanks to his PhD advisor, Prof. Alex Kontorovich for introducing this beautiful subject to the author and numerous enlightening discussions. The author also thanks the referee for her/his numerous corrections and helpful suggestions when the first edition of this paper was submitted. In writing up this paper, the author utilizes the codes provided by Prof. Kontorovich for several pictures. In addition, the author acknowledges support for this work from Prof. Kontorovich's NSF grants DMS-1209373, DMS-1064214, DMS-1001252 and his NSF CAREER grant DMS-1254788.

References

[1] Jean Bourgain and Elena Fuchs. A proof of the positive density conjecture for integer Apollonian circle packings. *J. Amer. Math. Soc.*, 24(4):945–967, 2011.

[2] Jean Bourgain, Alex Gamburd, and Peter Sarnak. Affine linear sieve, expanders, and sum-product. *Invent. Math.*, 179(3):559–644, 2010.

[3] Jean Bourgain, Alex Gamburd, and Peter Sarnak. Generalization of Selberg's $\frac{3}{16}$ theorem and affine sieve. *Acta Math.*, 207(2):255–290, 2011.

[4] Jean Bourgain and Alex Kontorovich. On the local-global conjecture for integral Apollonian gasket. *Invent. Math.*, July 2013.

[5] Harold Davenport. *Multiplicative number theory*, volume 1966 of *Lectures given at the University of Michigan, Winter Term*. Markham Publishing Co., Chicago, Ill., 1967.

[6] J. Elstrodt, F. Grunewald, and J. Mennicke. *Groups acting on hyperbolic space*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998. Harmonic analysis and number theory.

[7] Elena Fuchs. *Arithmetic properties of Apollonian circle packings*. ProQuest LLC, Ann Arbor, MI, 2010. Thesis (Ph.D.)–Princeton University.

[8] A. Salehi Golsefidy and Péter P. Varjú. Expansion in perfect groups. *Geom. Funct. Anal.*, 22(6):1832–1891, 2012.

[9] Ronald L. Graham, Jeffrey C. Lagarias, Colin L. Mallows, Allan R. Wilks, and Catherine H. Yan. Apollonian circle packings: number theory. *J. Number Theory*, 100(1):1–45, 2003.

[10] Ronald L. Graham, Jeffrey C. Lagarias, Colin L. Mallows, Allan R. Wilks, and Catherine H. Yan. Apollonian circle packings: geometry and group theory. I. The Apollonian group. *Discrete Comput. Geom.*, 34(4):547–585, 2005.
[11] Gerhard Guettler and Colin Mallows. A generalization of Apollonian packing of circles. *J. Comb.*, 1(1, [ISSN 1097-959X on cover]):1–27, 2010.
[12] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
[13] Inkang Kim. Counting, mixing and equidistribution of horospheres in geometrically finite rank one locally symmetric manifolds. 03 2011.
[14] H. D. Kloosterman. On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$. *Acta Math.*, 49(3-4):407–464, 1927.
[15] Alex Kontorovich and Hee Oh. Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds. *J. Amer. Math. Soc.*, 24(3):603–648, 2011. With an appendix by Oh and Nimish Shah.
[16] D. G. Larman. On the Besicovitch dimension of the residual set of arbitrarily packed disks in the plane. *J. London Math. Soc.*, 42:292–302, 1967.
[17] Peter D. Lax and Ralph S. Phillips. The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces. *J. Funct. Anal.*, 46(3):280–350, 1982.
[18] C. R. Matthews, L. N. Vaserstein, and B. Weisfeiler. Congruence properties of Zariski-dense subgroups. I. *Proc. London Math. Soc. (3)*, 48(3):514–532, 1984.
[19] S. J. Patterson. The limit set of a Fuchsian group. *Acta Math.*, 136(3-4):241–273, 1976.
[20] Peter Sarnak. Letter to J. Lagarias about integral Apollonian packings, June 2007.
[21] Soddy. The bowl of integers and hexlet. *Nature*, 139(77-79), 1937.
[22] Dennis Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.*, 153(3-4):259–277, 1984.
[23] Ilya Vinogradov. *Effective bisector estimate with application to Apollonian circle packings*. ProQuest LLC, Ann Arbor, MI, 2012. Thesis (Ph.D.)–Princeton University.
[24] Boris Weisfeiler. Strong approximation for Zariski-dense subgroups of semisimple algebraic groups. *Ann. of Math. (2)*, 120(2):271–315, 1984.