Local-global principles in circle packings

Elena Fuchs, Katherine E. Stange and Xin Zhang

Abstract

We generalize work of Bourgain-Kontorovich [6] and Zhang [37], proving an almost local-to-global property for the curvatures of certain circle packings, to a large class of Kleinian groups. Specifically, we associate in a natural way an infinite family of integral packings of circles to any Kleinian group $\mathcal{A} \leqslant \mathrm{PSL}_2(K)$ satisfying certain conditions, where K is an imaginary quadratic field, and show that the curvatures of the circles in any such packing satisfy an almost local-to-global principle. A key ingredient in the proof of this is that $\mathcal{A}$ possesses a spectral gap property, which we prove for any infinite-covolume, geometrically finite, Zariski dense Kleinian group in $\mathrm{PSL}_2(\mathcal{O}_K)$ containing a Zariski dense subgroup of $\mathrm{PSL}_2(\mathbb{Z})$.

1. Introduction

Local-to-global questions have been studied throughout the history of number theory. Here, we consider the set of curvatures appearing in circle packings which are orbits of thin Kleinian groups: when is the set of curvatures essentially characterised by congruence conditions alone? In this context, a *thin Kleinian group* is one commensurable to an infinite index subgroup of a Bianchi group $\mathrm{PSL}_2(\mathcal{O}_K)$, but simultaneously Zariski dense in PGL_2.

This question was first considered in 2003 in a groundwork paper by Graham, Lagarias, Mallows, Wilks and Yan [14]. They observed that for several primitive integral Apollonian packings there appears to be a set of congruence classed modulo 24 or 48 such that any large enough integer having such a residue is indeed a curvature in that packing. They conjectured that this is the case for all packings. In 2011, the first-named author of the present paper made a detailed study of congruence conditions for Apollonian packings [12]. Together with Sanden, this author performed extensive numerical experiments and conjectured that in fact all primitive integral Apollonian packings can be described in terms of conditions modulo 24 [13].

The first step towards trying to prove this conjecture is in [14], where it is shown that at least $cx^{1/2}$ integers less than x appear as curvatures in a given integral Apollonian packing, where c is a constant depending on the packing. Sarnak then made an observation in [27] which became the basis for all future developments on this question. In that letter, Sarnak showed that in any primitive Apollonian packing there are, up to a constant, at least $\frac{x}{\sqrt{\log x}}$ integers less than x which appear as curvatures in the packing. His approach was to observe that if one fixes a circle in the packing and considers only those circles tangent to it, their curvatures, without multiplicity, are exactly the set of numbers that are primitively represented by a shifted binary quadratic form $f(x, y) - a$ whose coefficients depend on the circle that is fixed. Sarnak's idea was then

2010 Mathematics Subject Classification Primary: 52C26, 30F40, 11D85 Secondary: 20H10, 22E40
Keywords: local-to-global, Kleinian group, circle method, Apollonian circle packing
Fuchs has been supported by NSF DMS-1501970, the Sloan Foundation, and the BSF. Stange has been supported by NSF EAGER DMS-1643552 and NSF CAREER CNS-1652238.

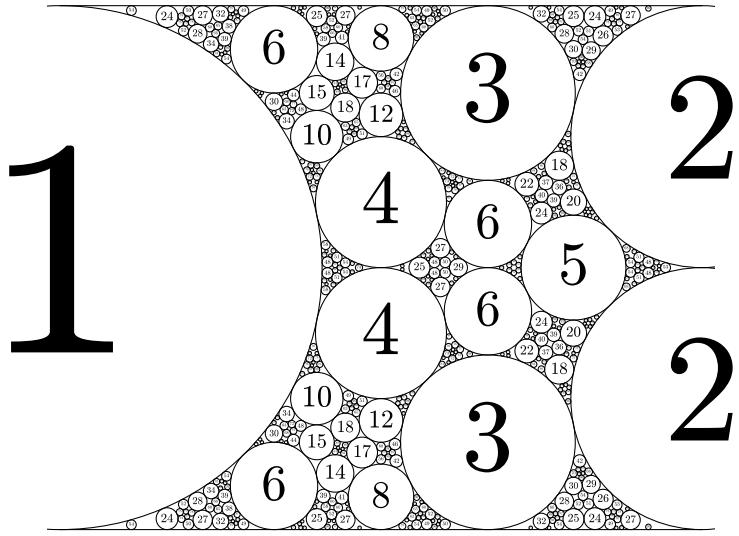

Figure 1. The limit set of an example packing to which Theorem 1.6 applies (approximation to portion with $0 \leqslant x \leqslant 3$), with curvatures shown (scaled by $3/\sqrt{6}$ to give a primitive integral packing). See Section 9.1.

expanded by Bourgain and Fuchs to prove that in fact a positive fraction of all integers appear in any primitive integral Apollonian packing [3]. The methods of [3] were then taken several steps further by Bourgain and Kontorovich in [6] to prove an asymptotic local-to-global principal for Apollonian packings: they showed that, if A is the set of positive integers that are admissible as curvatures in a given primitive integral Apollonian packing according to their residue modulo 24, the subset of A of integers which do *not* appear as curvatures in the packing make up a zero density subset of all integers.

How far can one take the method in [6] to prove asymptotic local-to-global principles in the thin setting? For example, the third-named author of this paper successfully used the tools of [6] to prove an asymptotic local-to-global principle in so-called *integral Apollonian 3-packings* [37]. In this paper, we identify the key necessary conditions for these methods to work, which, when satisfied, guarantee an asymptotic local-to-global principle for an integral circle packing or, viewed differently, an orbit of a thin subgroup of $\mathrm{PSL}_2(\mathbb{C})$. As a consequence, we immediately have that an asymptotic local-to-global principle holds for the *K-Apollonian packings* described by the second-named author [32] and for *superintegral polyhedral packings* described by Kontorovich-Nakamura [18]. We provide a concrete example of such a packing and give more details on the packings of Stange and Kontorovich-Nakamura in Section 9. See Figures 1 and 2.

In the work on Apollonian packings by Bourgain, Fuchs, Kontorovich, and Zhang, the curvatures in the packings were represented as coordinates of points in an orbit of a thin subgroup of $\mathrm{O}_Q(\mathbb{Z})$, where Q is a signature $(3,1)$ quadratic form which is simply the Descartes form in the Apollonian case, and an analogue thereof in the 3-packing case. In both of these cases, one can view the curvatures as curvatures of circles obtained by considering the orbit via Möbius transformations of a fixed circle (or line) in the complex plane under the action of a thin (Kleinian)

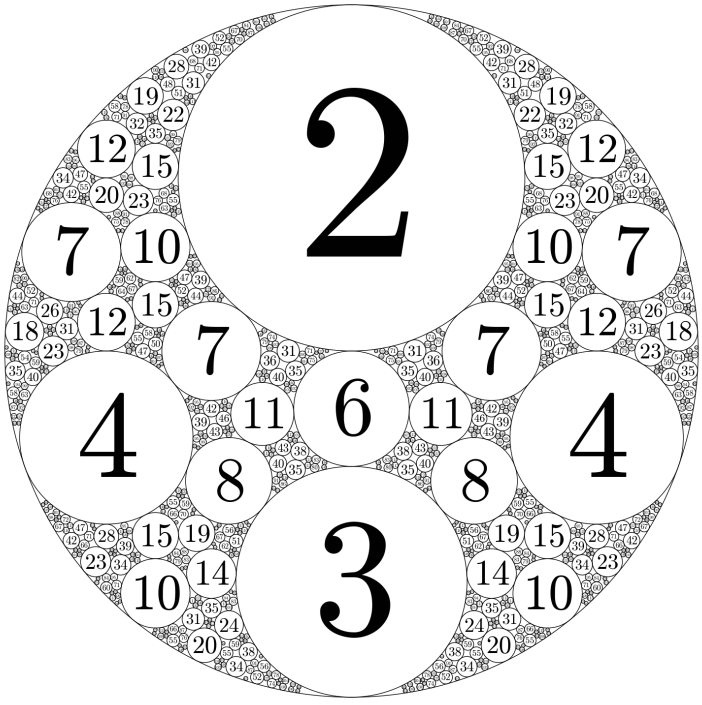

Figure 2. The limit set of an example packing to which Theorem 1.6 applies, with curvatures shown (scaled by $1/\sqrt{2}$ to give a primitive integral packing). This is an example of a K-Apollonian packing for $K = \mathbb{Q}(\sqrt{-2})$. See Section 9.2.

subgroup $\mathcal{A}$ of $\mathrm{PSL}_2(\mathcal{O}_K)$ where K is an imaginary quadratic field. In the original Apollonian case, $K = \mathbb{Q}(i)$, and in the 3-packing case $K = \mathbb{Q}(\sqrt{-2})$. One can pass between these two interpretations of the set of curvatures via the spin homomorphism $\rho : \mathrm{PSL}_2(\mathbb{C}) \to \mathrm{O}_{\mathbb{R}}(3,1)$, but the PSL_2 setup is more convenient for several reasons: for example, there are numerous choices for the analogue of the Descartes form if one chooses to work in $\mathrm{O}_{\mathbb{R}}(3,1)$; also, SL_2 is simply connected, while the orthogonal group is not.

Definition 1.1. *Let $\mathcal{A}$ be a Kleinian group, and let $C_1, \ldots, C_n$ be circles in the extended complex plane. Write $\mathcal{A}C_i$ for the orbit of C_i under $\mathcal{A}$, as a subset of the plane (a union of circles). Then*

$$\bigcup_{i=1}^{n} \mathcal{A}C_i$$

is called a Kleinian circle packing. *Such a packing is called* integral *if, after a universal scaling factor is applied, the set of curvatures can be taken to be a subset of $\mathbb{Z}$.*

We define the curvature of a circle $N(\widehat{\mathbb{R}})$, where $N = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, to be $2\Im(\overline{C}D)$; then the radius is $1/|2\Im(\overline{C}D)|$, but the curvature contains some further information in the form of the sign, which can be interpreted as orientation. In general, the circles in a Kleinian circle packing may overlap, although they do not in the most famous cases, such as the Apollonian circle packing.

Although one might conjecture a local-global principle for a larger class of integral Kleinian circle packings, our methods require that the packing contain 'congruence families' of circles, which give rise to integral binary quadratic forms as in the Apollonian case. Therefore we define a restricted class of groups.

Definition 1.2. *A Kleinian group $\mathcal{A}$ is called* familial *if:*

(i) $\mathrm{PSL}_2(\mathbb{Z}) \cap \mathcal{A}$ *contains a principal congruence subgroup, and*

(ii) *the entries of $\mathcal{A}$ are contained in some fractional ideal $\mathfrak{a}$ of an imaginary quadratic field $K = \mathbb{Q}(\sqrt{-d})$, $d > 0$.*

Furthermore, the methods require that the group $\mathcal{A}$ has a spectral gap property: i.e. that the family of graphs $\{\mathrm{Cay}(\mathcal{A}/\mathcal{A}(q), \overline{S})\}_q$ is an expander family. Here $\mathcal{A}/\mathcal{A}(q)$ denotes $\mathcal{A}$ reduced modulo q, the set S is a finite generating set of $\mathcal{A}$, $\overline{S}$ denotes its image under reduction, and q ranges over all positive integers. In Section 8, we show that this is the case for a class of groups including those we intend to consider, i.e., we show the following.

Theorem 1.3. *Any infinite-covolume, geometrically finite, Zariski dense Kleinian group contained in $PSL_2(\mathcal{O}_K)$, containing a Zariski dense subgroup of $PSL_2(\mathbb{Z})$ has a combinatorial spectral gap.*

In our setting, the group $\mathcal{A}$ has critical exponent greater than 1. It then follows from the arguments in [17, Theorem 11.2, Step II], which is an immediate generalization of [4, Theorem 1.2], that $\mathcal{A}$ also enjoys a *geometric spectral gap* (see [4, p. 39] for a definition).

Previously, Sarnak established the spectral gap property for finite-index subgroups of $\mathrm{PSL}_2(\mathcal{O}_K)$ [25], and Magee established this property for geometrically finite subgroups of $\mathrm{PSL}_2(\mathcal{O}_K)$ with critical exponents greater than $\frac{11}{6}$ [21]. Salehi-Golsefidy and Zhang, the third named author, generalize Theorem 1.3 even further in an upcoming preprint [24]. The existence of a geometric

spectral gap is a crucial ingredient both in [6] and in [37], and indeed in almost all works that have investigated arithmetic aspects of thin groups. We indicate exactly how this spectral gap is relevant in Section 5. Note that, as explained below, every familial group $\mathcal{A}$ does satisfy the Hausdorff dimension hypothesis in Theorem 1.3.

We now state the general setup of the paper. Let K be an imaginary quadratic field. Henceforth, we will assume that $\mathcal{A}$ denotes an infinite-covolume, geometrically finite, Zariski-dense, familial Kleinian group in $\mathrm{PSL}_2(K)$. We will consider an associated packing $\mathcal{P} := M\mathcal{A}C$, where C is any circle tangent to the real line and having the form $C = N(\widehat{\mathbb{R}})$, where $N, M \in \mathrm{PSL}_2(K)$.

This last condition, on the tangency of C to the real line, is crucial to the methods of the paper, as it guarantees, together with the congruence subgroup condition of Definition 1.2, that a collection of integral binary quadratic forms govern the curvatures of the packing.

Under these conditions, the packing $\mathcal{P}$ is necessarily integral as in Definition 1.2 (see Section 3). We let $\mathcal{K} \subset \mathbb{Z}$ be the set of curvatures, after some universal scaling factor is applied as in the definition of integrality.

Let $\mathcal{K}_a$ be the set of integers passing all the local obstructions by $\mathcal{K}$. In other words,

$$\mathcal{K}_a = \{n \in \mathbb{Z} \mid \forall q \in \mathbb{Z}, \exists k \in \mathcal{K}, \text{such that } n \equiv k (\mathrm{mod}\ q)\} \tag{1.1}$$

We call the integers in $\mathcal{K}_a$ *admissible*.

An immediate corollary of the spectral gap statement in Theorem 1.3 is the following.

Corollary 1.4. *There exists a positive integer L_0 such that $\mathcal{K}_a$ is the union of some congruence classes mod L_0.*

Of course, this also follows by strong approximation (see [23]) for SL_2. However, the proof of our Theorem 1.3 not only gives the existence of L_0 but also gives an algorithm to quickly determine its exact value: in particular, the prime factors of L_0 will come from the level of the congruence subgroup contained in $\mathcal{A}$, any failure of primitivity of the packing $\mathcal{P}$, and the primes 2 and 3, as well as the matrix M if M is fractional. See Theorem 8.1 and (6.23), for details.

Now let $\mathcal{K}_a(N) = \mathcal{K}_a \cap [0, N]$ be the set of admissible integers up to N, and similarly denote $\mathcal{K}(N) = \mathcal{K} \cap [0, N]$. Then Corollary 1.4 directly implies that

$$\#\mathcal{K}_a(N) = c_{M,\mathcal{A},C} N + O(1), \tag{1.2}$$

where $c_{M,\mathcal{A},C}$ is the proportion of admissible congruence classes. We predict that all sufficiently large admissible integers are actually curvatures, or in other words,

Conjecture 1.5.

$$\#\mathcal{K}(N) = c_{M,\mathcal{A},C} N + O(1). \tag{1.3}$$

In place of the full conjecture, we prove the following theorem:

Theorem 1.6. *Let $\mathcal{A}$ and $\mathcal{P} = M\mathcal{A}C$ be as above. There exists a positive number η, depending only on M, $\mathcal{A}$ and C, such that*

$$\#\mathcal{K}(N) = c_{M,\mathcal{A},C} N + O(N^{1-\eta}) \tag{1.4}$$

We feel that it is unlikely that our method can prove Conjecture 1.5 without significant new ideas.

We mention a remark of Chris Leininger: in fact our geometric finiteness assumption can be relaxed to be *finitely generated*. It is a corollary of the Tameness Theorem [1] that any finitely-generated Zariski dense subgroup of the Bianchi group containing a congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ must be geometrically finite.

Kontorovich and Nakamura define a family of dense circle packings of the plane defined by hyperbolic reflection groups built from uniform polyhedra and their growths [18]. For infinitely many of their examples, Kontorovich and Nakamura verify in their paper that such packings satisfy the hypotheses of Theorem 1.6, and hence have a local-to-global principle.

Of course, it is possible to construct examples of integral Kleinian packings which *fail* to satisfy the hypotheses of Theorem 1.6. For example, one may take a non-congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$, and adjoin another element to obtain a non-Fuchsian group; then consider the orbit of a K-rational circle tangent to $\widehat{\mathbb{R}}$. In such a case, one cannot guarantee the existence of a suitable family of binary quadratic forms: one only obtains quadratic forms in four related variables. It is therefore an interesting open question to develop methods which will prove an analogue to Theorem 1.6 for such packings.

In Section 9, in order the demonstrate the variety of examples to which our work applies, we verify that the hypotheses of Theorem 1.6 hold for the K-Apollonian packings of the second-named author [32], and also for an explicit example of a cuboctahedral packing (which also arises in the work of Kontorovich and Nakamura; Figure 1).

The main method in the proof of this theorem is the Hardy-Littlewood circle method. In the major arc analysis of the circle method, the main ingredient is an effective counting of group elements for $\mathcal{A}$ and its congruence subgroups originally achieved by Vinogradov [34]. In doing this, we require a geometric spectral gap for $\mathcal{A}$ in order to have a uniform control over the error terms. In Section 8 we establish a combinatorial spectral gap for $\mathcal{A}$, In Section 8 we establish a combinatorial spectral gap for $\mathcal{A}$, which in turn implies a geometric spectral gap for $\mathcal{A}$ by the methods in [4], proving Theorem 1.3. Moreover, we require that $\delta = \delta(\mathcal{A})$, the critical exponent of $\mathcal{A}$, which is also the Hausdorff dimension of the limit set of $\mathcal{A}$, is strictly greater than 1, which is guaranteed by our assumption that $\mathcal{A}$ is familial, and a limit set classification theorem of Bishop-Jones [2, Corollary 1.8].

Besides the existence of the spectral gap, which is crucial for minor arcs as well as major arcs, the main ingredient in the minor arc analysis is the quadratic form structure, which allows us to do abelian harmonic analysis of two free variables. Certain Kloosterman-type sums naturally appear here, where we apply standard methods to gain power savings. In fact the power saving here, as well as in [6] and [37], is so significant that one does not need further restriction on the critical exponent δ (besides $\delta > 1$), in contrast to the works [5], [7], and [35], which require the critical exponent to be very big in order to get enough cancellation in the minor arc analysis.

Note that our methods, while similar to that in [6] and [37], require several new ingredients and careful generalizations to work. One crucial such ingredient is the spectral gap of Theorem 1.3. This theorem applies to a much wider class of groups than our local-to-global analysis, and generalizes the case of the Apollonian group, proven by P. Varjú in the appendix of [6]. In proving this theorem, we do not, for instance, have any concrete information about the generators of the group we work with, or exactly at which primes and to what level there are local obstructions for the group. Indeed, in the proof of Theorem 1.3, we are able to derive, in the case of the groups considered within this paper, exactly what the local obstructions should be: something that was done explicitly for the Apollonian group in [12].

Secondly, the fact that we work with an arbitrary imaginary quadratic field K (as opposed to $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$ as in [6] and [37], respectively), and an abstract subgroup of $\mathrm{PSL}_2(\mathcal{O}_K)$ makes the local analysis in the major arcs section (Section 6) much less straightforward: where the authors of [6] and [37] could depend on concrete local information about the groups they work with, we derive this without relying on explicit information about the local obstructions.

Thirdly, in both [6] and [37], the level of the congruence subgroup contained in the Apollonian group in question is 2, which means that the curvatures of the circles are exactly the set of integers represented by a corresponding class of shifted binary quadratic forms. In our paper this is no longer the case and it is possible that the curvatures we consider (after appropriate scaling to make them integral) comprise a subset of values of the corresponding class of shifted forms. In fact, while the methods here deal with this nicely, this would make executing the positive density proof in [3] significantly more cumbersome in our setting than in the original setup of the classical Apollonian group.

We have made a special effort to make our exposition of these methods particularly accessible, in the hope that it may benefit students and experts alike.

Notation: Sections 2 through 7 are notation-heavy. For ease of reading, we include a table of the major notation used in those sections in Table 1 of Section 10. We also note that whenever the constant η appears, it is assumed to satisfy not only the current claim, but also all claims in previous contexts.

Acknowledgements: We would like to thank Hee Oh for raising the question of how general the methods in [3] and [6] are, which is what motivated this paper. We also thank Nathan Dunfield, Alireza Salehi-Golsefidy, Alex Kontorovich, Chris Leininger, Kei Nakamura and Hee Oh for helpful conversations. We would also like to thank the referee for many very helpful and thoughtful comments.

Figures: Figures were produced with Sage Mathematics Software [11].

2. Integrality of $\mathcal{A}$

For the purpose of our methods, we intend to replace $\mathcal{A}$ with $\tilde{\mathcal{A}} := \mathcal{A} \cap \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$, where d is as in Definition 1.2, since we would like to work with an integral group. The next lemma asserts that $\tilde{\mathcal{A}}$ is finite index in $\mathcal{A}$.

Without loss of generality, we can replace $\mathcal{A}$ with any finite-index subgroup for the purposes of Theorem 1.6. This is because a finite number of orbits of the subgroup comprise the full orbit of $\mathcal{A}$, and the congruence obstructions from these orbits can be combined to give the obstruction for the union.

For this reason, we are free to assume throughout the paper that $\mathcal{A}$ is torsion-free, by Selberg's theorem, saying that any matrix group contains a finite-index torsion-free subgroup [28], and, by the following lemma, that it is a subgroup of $\mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$.

Lemma 2.1. *Let $\mathcal{A}$ be as defined in the introduction, and let $\tilde{\mathcal{A}} = \mathcal{A} \cap PSL_2(\mathbb{Z}[\sqrt{-d}])$. Then $[\mathcal{A} : \tilde{\mathcal{A}}]$ is finite.*

Proof. Recall that $K = \mathbb{Q}(\sqrt{-d})$. If $\mathcal{A} \subset \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$, then the statement is trivial. Hence, suppose $\mathcal{A} \not\subset \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$, such that the denominators featured in its elements are bounded above, as assumed in the previous section. Let $q = p_1^{e_1} \cdots p_k^{e_k}$, where $p_1, \ldots, p_k$ are distinct primes,

be the least common multiple of all denominators featured among entries of elements of $\mathcal{A}$.

Let $H_1 = \mathrm{PSL}_2(\frac{1}{q}\mathbb{Z}[\sqrt{-d}])$, let $H_2 = \mathcal{A}$, and let $H_3 = \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$. Note that H_1 is not a group, but contains both H_2 and H_3. Furthermore, it is covered by some union of cosets of H_3 in $\mathrm{PSL}_2(K)$. If H_1 is covered by a finite union of cosets of H_3, then $H_2 = H_2 \cap H_1$ is covered by a finite union of cosets of $H_2 \cap H_3$, i.e. $[\mathcal{A} : \tilde{\mathcal{A}}]$ is finite.

Therefore, we will cover H_1 by a finite union of cosets of H_3. To show this, note that if the p_i-adic expansions of $\gamma_1, \gamma_2 \in \mathrm{PSL}_2(\frac{1}{q}\mathbb{Z}[\sqrt{-d}])$ agree in the $p_i^{-e_i}, p_i^{-e_i+1}, \dots, p_i^{e_i}$ terms for all $1 \leqslant i \leqslant k$, then the "coefficients" of the entries of $\gamma_1\gamma_2^{-1}$ are p_i-adic integers for all i. Here, what we mean by p_i-adic expansions of γ is what one gets when one considers for each entry of γ of the form $a + b\sqrt{-d}$ the p_i-adic expansion of a and b. By "coefficients" of an entry $a + b\sqrt{-d}$ of γ we mean precisely a and b. Since $\gamma_1\gamma_2^{-1} \in \mathrm{PSL}_2(\frac{1}{q^2}\mathbb{Z}[\sqrt{-d}])$, this in fact implies that $\gamma_1\gamma_2^{-1} \in \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$. Since there are only finitely many possibilities for the $p_i^{-e_i}, p_i^{-e_i+1}, \dots, p_i^{e_i}$ terms in the p_i-adic expansion of any number, where i ranges over finitely many indices, we have that there are in fact finitely many cosets of $\mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$ in $\mathrm{PSL}_2(\frac{1}{q}\mathbb{Z}[\sqrt{-d}])$, as desired. □

We remark that a converse also holds: if $\mathcal{A}$ has its intersection with the Bianchi group as a subgroup of finite index, then $\mathcal{A}$ has bounded denominators.

Therefore, from this point on we assume $\mathcal{A}$ is a torsion-free subgroup of $\mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$.

3. Families of quadratic forms

We now describe the set of curvatures $\mathcal{K}$ as a union of values of a family of quadratic forms. Write Δ for the discriminant of $\mathcal{O}_K$. If $d \equiv 1, 2 (\mathrm{mod}\ 4)$, then $\Delta = -4d$, and if $d \equiv 3 (\mathrm{mod}\ 4)$, then $\Delta = -d$. Letting $\gamma = \begin{pmatrix} A_\gamma & B_\gamma \\ C_\gamma & D_\gamma \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{C})$, direct computation shows that γ sends the horizontal line $\widehat{\mathbb{R}}$ to a circle of curvature

$$\kappa(\gamma(\widehat{\mathbb{R}})) = 2\Im(\overline{C_\gamma} D_\gamma) \in \mathbb{R}. \tag{3.1}$$

If $\gamma \in \mathrm{PSL}_2(\mathcal{O}_K)$, then $\kappa(\gamma(\widehat{\mathbb{R}})) \in \sqrt{-\Delta}\mathbb{Z}$.

We may assume without loss of generality that $N(\widehat{\mathbb{R}}) = \widehat{\mathbb{R}} + \sqrt{\Delta}/2$. For, $\mathrm{PSL}_2(\mathbb{Q})$ is transitive on circles of $\mathrm{PSL}_2(K)\widehat{\mathbb{R}}$ tangent to $\widehat{\mathbb{R}}$. Therefore we may choose N_0 satisfying $N_0(\widehat{\mathbb{R}}) = \widehat{\mathbb{R}} + \sqrt{\Delta}/2$, and $NN_0^{-1} \in \mathrm{PSL}_2(\mathbb{Q})$. Then we have

$$M\mathcal{A}N = (MNN_0^{-1})(N_0N^{-1}\mathcal{A}NN_0^{-1})N_0.$$

But $M' = MNN_0^{-1} \in PSL_2(K)$, $N_0 \in PSL_2(\mathcal{O}_K)$, and $\mathcal{A}' = N_0N^{-1}\mathcal{A}NN_0^{-1}$ is again Zariski dense, infinite covolume, geometrically finite and familial. Therefore let us assume $N(\widehat{\mathbb{R}}) = \widehat{\mathbb{R}} + \sqrt{\Delta}/2$. By Lemma 2.1, we may again pass to a finite index subgroup $\mathcal{A}$ of $\mathcal{A}'$ and work with this group in order to prove Theorem 1.6.

With this choice of N, for any $\gamma \in \mathcal{A}$, and M as above, the curvatures of the orbit $M\gamma\,\mathrm{PSL}_2(\mathbb{Z})(\widehat{\mathbb{R}} + \frac{\sqrt{\Delta}}{2})$ are given by the shifted quadratic form

$$\widehat{\mathfrak{f}}_{M\gamma}(a, c) = \sqrt{-\Delta}\,|C_{M\cdot\gamma}a + D_{M\gamma}c|^2 + 2\Im(\overline{C_{M\gamma}} D_{M\gamma}) \tag{3.2}$$

in terms of the entries a and c of $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z})$. Therefore the packing $M\mathcal{A}(\widehat{\mathbb{R}} + \frac{\sqrt{\Delta}}{2})$

contains the curvatures of

$$\left\{\widehat{\mathfrak{f}}_{M\gamma}(Lx+1,Ly) : \gcd(x,y)=1\right\},$$

where L is the level of the congruence subgroup contained in $\mathcal{A}$. Write

$$\mathfrak{f}_{M\gamma}(a,c) = \frac{1}{\sqrt{-\Delta}}\widehat{\mathfrak{f}}_{M\gamma}(a,c).$$

Then $\mathfrak{f}_{M\gamma}(a,c)$ is a shifted binary rational quadratic form, i.e.

$$\mathfrak{f}_{M\gamma}(a,c) = \widetilde{\mathfrak{f}}_{M\gamma}(a,c) + \mathfrak{d}_\gamma \tag{3.3}$$

where

$$\widetilde{\mathfrak{f}}_{M\gamma}(a,c) = |C_{M\gamma}a + D_{M\gamma}c|^2, \quad \text{and} \quad \mathfrak{d}_\gamma = 2\frac{\Im(\overline{C_{M\gamma}}D_{M\gamma})}{\sqrt{-\Delta}}.$$

In particular, $\widetilde{f}_{M\gamma}$ has discriminant $\Delta\mathfrak{d}_\gamma^2 = -4\left(\Im(\overline{C_{M\gamma}}D_{M\gamma})\right)^2 < 0$.

Unlike in the Apollonian case, it is possible that not all of these forms are primitive integral binary quadratic forms. However, their deviation from such forms, which is a function of the denominators introduced by M, is uniformly bounded.

Lemma 3.1. *Let $M \in \mathrm{PSL}_2(K)$ and let d_1 be such that $d_1 M \in \mathrm{PGL}_2(\mathcal{O}_K)$. Up to multiplying and/or dividing by integers dividing d_1^4, the form $\widetilde{f}_{M\gamma}$ becomes a primitive integral binary quadratic form.*

Proof. We have that $M\gamma \in \mathrm{PSL}_2(K)$. In particular, we have

$$A_{M\gamma}D_{M\gamma} - B_{M\gamma}C_{M\gamma} = 1. \tag{3.4}$$

By assumption, $C_{M\gamma}, D_{M\gamma} \in \frac{1}{d_1}\mathcal{O}_K$. Write

$$C_{M\gamma} = \frac{C'_{M\gamma}}{d_1}, \quad D_{M\gamma} = \frac{D'_{M\gamma}}{d_1}.$$

Where $C'_{M\gamma}, D'_{M\gamma} \in \mathcal{O}_K$. In particular, the ideal generated by $C'_{M\gamma}$ and $D'_{M\gamma}$ has norm at most d_1^4 by (3.4).

For any $C, D \in \mathcal{O}_K$, if the integral form

$$|Cx+Dy|^2 = C\overline{C}x^2 + (C\overline{D} + \overline{C}D)xy + D\overline{D}y^2$$

is imprimitive by a factor of, say, e dividing all its coefficients, then $e \mid N(C,D)$ (the norm of the ideal). To see this, suppose p is prime and $p^k \mid |Cx+Dy|^2$ for all (x,y). If p is inert, then this implies $(C,D) \subset p^{\lceil k/2\rceil}\mathcal{O}_K$, so $p^k \mid N(C,D)$. If $p = \mathfrak{p}\overline{\mathfrak{p}}$ is split, then C, D and $C+D$ are each contained in some ideal $\mathfrak{p}^s\overline{\mathfrak{p}}^t$, where $s+t=k$. Call these three pairs $(s,t) = (s_1,t_1),(s_2,t_2),(s_3,t_3)$, ordered so that $s_1 < s_2 < s_3$. Then

$$(C,D) \subset \mathfrak{p}^{s_2}, \overline{\mathfrak{p}}^{t_2},$$

since any two of $C, D, C+D$ generate (C,D). Hence, $p^k \mid N(C,D)$.

Therefore $|C'_{M\gamma}x + D'_{M\gamma}y|^2$ is imprimitive by a factor dividing d_1^4.

This shows that the integrality and/or primitivity of $\widetilde{f}_{M\gamma}$ is achieved by multiplication and/or division by a factor of at most d_1^4, where d_1 is independent of γ. □

Finally, the integral curvatures we seek to study are given by the union of the integers represented by these shifted forms, i.e.

$$\mathcal{K}_{M\mathcal{A}(\widehat{\mathbb{R}}+\frac{\sqrt{\Delta}}{2})} = \bigcup_{\gamma\in\mathcal{A}} \{\mathfrak{f}_{M\gamma}(Lx+1, Ly) : \gcd(x,y)=1\}. \tag{3.5}$$

4. setup of the circle method

Throughout the circle method, there are the following growing parameters:

$$T, X, N, T_1, T_2, J, Q_0, K_0, U, H.$$

Their precise relationships, used to tune the result, are boxed throughout the paper, and these are: (4.1), (4.2), (4.11), (5.1) and (7.53). We collect these equations here for reference:

$$\boxed{\begin{array}{ll} N = T^2X^2, & Q_0 = T^{\frac{2\delta-2\Theta}{80}} \\ T = N^{1/200}, & K_0 = Q_0^3, \\ T = T_1T_2, & H = Q_0^{\frac{\eta_0}{4}}, \\ T_2 = T_1^\nu, & \\ J = T^2X, & U = H^{\frac{\eta_0}{20}}. \end{array}}$$

Each element $\gamma \in \mathcal{A}$ corresponds to a shifted quadratic form of two variables that represents curvatures of circles tangent to $M\gamma(\mathbb{R} + \frac{\sqrt{\Delta}}{2})$, given in (3.3). Note that M is fixed throughout the paper.

Our goal is to show that almost all admissible integers are represented by some such shifted form. To do this, we consider this problem applied to growing subsets of $\mathcal{A}$, and the shifted binary forms corresponding to the elements in these subsets.

We now define these growing subsets. We choose three growing parameters N, T, and X such that

$$\boxed{N = T^2X^2, \quad T = N^{\frac{1}{200}}.} \tag{4.1}$$

Since T is a small power of N, we have that X is almost of the scale of $N^{\frac{1}{2}}$. We further write

$$\boxed{T = T_1T_2, T_2 = T_1^\nu} \tag{4.2}$$

where $\nu > 0$ is a large number depending only on the spectral gap of $\mathcal{A}$, and we define the following set (counting with multiplicity):

$$\mathfrak{F} = \mathfrak{F}_T = \left\{ \gamma = \gamma_1\gamma_2 : \begin{array}{c} \gamma_1, \gamma_2 \in \mathcal{A} \\ T_1/2 \leqslant \|M\gamma_1\| \leqslant T_1 \\ T_2/2 \leqslant \|\gamma_2\| \leqslant T_2 \\ \Im(\overline{C_{M\gamma}}D_{M\gamma}) \geqslant T/100 \end{array} \right\} \tag{4.3}$$

Here $\|\cdot\|$ stands for the Frobenius norm.

The reason that we define $\mathfrak{F}_T$ using two parameters T_1 and T_2 is that this is the necessary setup for Lemma 5.1 which is a result of Bourgain-Kontorovich from [6], and one that we will be using in the minor arcs analysis in this paper. Lemma 5.2 and Lemma 5.3 are also stated within

this setup, although for these two results one can set up the problem with just a growing ball of radius T.

Finally, we wish to let two integers x and y range over two sets of integers that are $\asymp X$. For this reason we introduce a smooth function ψ supported on $[1,2]$, such that $\psi \geqslant 0$ and $\int_{\mathbb{R}} \psi(x)dx = 1$. If $(Lx+1, Ly) = 1$, then $\mathfrak{f}_{M\cdot\gamma}(Lx+1, Ly)$ is a curvature.

We then define

$$\mathcal{R}_N(n) = \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{x,y\in\mathbb{Z}\\(Lx+1,Ly)=1}} \psi\left(\frac{Lx+1}{X}\right)\psi\left(\frac{Ly}{X}\right)\mathbf{1}\{\mathfrak{f}_{M\gamma}(Lx+1,Ly)=n\} \tag{4.4}$$

If $\mathcal{R}_N(n) > 0$ then n is a curvature. Our goal is to show that $\mathcal{R}_N(n) > 0$ for almost all n, in the sense described in Theorem 1.6.

From Theorem 2.2 in [34], the size of $\mathfrak{F}_T$ is $\asymp T^{2\delta}$ (recall that δ is the critical exponent of $\mathcal{A}$). Given the definition of $\mathfrak{F}_T$, the function $\mathcal{R}_N$ is supported on $n \asymp N$. We obtain

$$\|\mathcal{R}_N\|_{l_1} = \sum_{n\asymp N} \mathcal{R}_N(n) \asymp T^{2\delta}X^2.$$

It is expected that this is roughly equidistributed on the set of admissible integers, so that

$$\mathcal{R}_N(n) \gg \frac{T^{2\delta}X^2}{N} = T^{2\delta-2} \tag{4.5}$$

for every admissible n.

It would be ideal to show (4.5). However, current technology does not enable us to prove this. Instead, we will show that $\mathcal{R}_N(n) \gg T^{2\delta-2}$ for every admissible integer in $[N/2, N]$ outside of an exceptional set of size $O(N^{1-\eta})$ for some $\eta > 0$.

Notice that in the definition of $\mathcal{R}_N$ in (4.4), the second sum is over pairs of integers $(Lx+1, Ly)$ which satisfy a coprimality condition that is hard to track directly in computations. We hence rewrite this sum as one over all pairs $(Lx+1, Ly)$ using Möbius orthogonality:

$$\sum_{d|n}\mu(d) = \begin{cases} 1 & \text{if } n = 1,\\ 0 & \text{if } n > 1.\end{cases} \tag{4.6}$$

Then

$$\mathcal{R}_N(n) = \sum_{\gamma\in\mathfrak{F}_T}\sum_{x,y\in\mathbb{Z}}\sum_{u|(Lx+1,Ly)} \mu(u)\psi\left(\frac{Lx+1}{X}\right)\psi\left(\frac{Ly}{X}\right)\mathbf{1}\{\mathfrak{f}_{M\gamma}(Lx+1,Ly)=n\}.$$

Notice that if $u|(Lx+1, Ly)$, then $(u, L) = 1$, so $y \equiv 0 \pmod u$ and $Lx \equiv -1 \pmod u$. Let u^* be the integer from $[1, L-1]$ such that $uu^* \equiv 1 \pmod L$. Then we can write

$$\mathcal{R}_N(n) = \sum_{(u,L)=1}\mu(u)\sum_{\gamma\in\mathfrak{F}_T}\sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right)\mathbf{1}\{\mathfrak{f}_{M\gamma}(Lux+uu^*,Luy)=n\}. \tag{4.7}$$

With this manipulation, the innermost sum becomes one over free variables x, y, allowing us to use abelian harmonic analysis to analyze it.

To facilitate our analysis we will study a relative of $\mathcal{R}_N$ which we denote by $\mathcal{R}_N^U$, where U is a small power of N, and determined at (7.53). We restrict the u-sum in (4.7) to $u < U$ and

define

$$\mathcal{R}_N^U(n) = \sum_{\substack{u<U \\ (u,L)=1}} \mu(u) \sum_{\gamma\in\mathfrak{F}_T} \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right)\mathbf{1}\{\mathfrak{f}_{M\gamma}(Lux+uu^*,Luy)=n\} \quad (4.8)$$

The following lemma shows that the difference between $\mathcal{R}_N$ and $\mathcal{R}_N^U$ is small in l_1:

Lemma 4.1.

$$||\mathcal{R}_N - \mathcal{R}_N^U||_{l_1} \ll \frac{T^{2\delta}X^2}{U}.$$

Proof. From (4.7) and (4.8),

$$\begin{aligned}
&\sum_{n\in\mathbb{Z}} |\mathcal{R}_N(n) - \mathcal{R}_N^U(n)| \\
&\leqslant \sum_{n\in\mathbb{Z}} \left| \sum_{u\geqslant U} \mu(u) \sum_{\gamma\in\mathfrak{F}_T} \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right)\mathbf{1}\{\mathfrak{f}_{M\gamma}(Lux+uu^*,Luy)=n\}\right| \\
&\leqslant \sum_{n\in\mathbb{Z}}\sum_{u\geqslant U}\sum_{\gamma\in\mathfrak{F}_T}\sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right)\mathbf{1}\{\mathfrak{f}_{M\gamma}(Lux+uu^*,Luy)=n\} \\
&\leqslant \sum_{u\geqslant U}\sum_{\gamma\in\mathfrak{F}_T}\sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right) \\
&\ll \sum_{u\geqslant U} T^{2\delta}\frac{X^2}{u^2} \ll \frac{T^{2\delta}X^2}{U}.
\end{aligned}$$

□

We will study $\mathcal{R}_N$ (and $\mathcal{R}_N^U$) via its Fourier transform:

$$\widehat{\mathcal{R}}_N(\theta) = \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{x,y\in\mathbb{Z} \\ (Lx+1,Ly)=1}} \psi\left(\frac{Lx+1}{X}\right)\psi\left(\frac{Ly}{X}\right) e\left(\mathfrak{f}_{M\gamma}(Lx+1,Ly)\theta\right) \quad (4.9)$$

using the fact that we can recover $\mathcal{R}_N$ from $\widehat{\mathcal{R}}_N$ via the Fourier inversion formula:

$$\mathcal{R}_N(n) = \int_0^1 \widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta. \quad (4.10)$$

It is in evaluating this integral in (4.10) that the circle method will be applied.

By Dirichlet's approximation theorem, given any positive integer J, for every real number $\theta\in[0,1)$, there exist integers r,q such that $1\leqslant q\leqslant J$ and $\left|\theta-\frac{r}{q}\right| < \frac{1}{q\cdot J}$. The integer J is called the depth of approximation, and we will take

$$\boxed{J = T^2X.} \quad (4.11)$$

The general philosophy of the circle method is that most of the contribution to the integral (4.10) should come from neighborhoods of rationals with small denominator. Such neighborhoods are called *major arcs*. One shows that (4.10) is bounded below, by bounding the major arcs below, and then bounding the minor arcs, considered an error term, above.

In our case the major arcs are comprised of $\theta \in [0,1]$ such that $|\theta - \frac{r}{q}| \leqslant \frac{K_0}{N}$, where $q \leqslant Q_0$. Here Q_0 and K_0 are small powers of N which depend on the spectral gap and are given in (5.1). Write $\beta = \theta - \frac{r}{q}$.

To define what we call the major arc contribution, we first introduce the hat function

$$\mathfrak{t}(x) := \max\{0, 1 - |x|\},$$

whose Fourier transform is

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

In particular, $\hat{\mathfrak{t}}$ is nonnegative (we take $\hat{\mathfrak{t}}(0) = 1$).

From $\mathfrak{t}$, we construct a spike function $\mathfrak{T}$, with period 1 on $\mathbb{R}$, to capture the major arcs:

$$\mathfrak{T}(\theta) := \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{m\in\mathbb{Z}} \mathfrak{t}\left(\frac{N}{K_0}\left(\theta + m - \frac{r}{q}\right)\right). \tag{4.12}$$

Our main term is then

$$\mathcal{M}_N(n) := \int_0^1 \mathfrak{T}(\theta)\widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta \tag{4.13}$$

and the error term is

$$\mathcal{E}_N(n) := \int_0^1 (1 - \mathfrak{T}(\theta))\widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta. \tag{4.14}$$

Similarly, we define

$$\mathcal{M}_N^U(n) := \int_0^1 \mathfrak{T}(\theta)\widehat{\mathcal{R}}_N^U(\theta)e(-n\theta)d\theta \tag{4.15}$$

and

$$\mathcal{E}_N^U(n) := \int_0^1 (1 - \mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)e(-n\theta)d\theta. \tag{4.16}$$

In Lemma 4.1 we have shown that $\|\mathcal{R}_N - \mathcal{R}_N^U\|_{l_1}$ is small. Running the same argument as in the proof of Lemma 4.1, one can bound the difference between $\mathcal{M}_N$ and $\mathcal{M}_N^U$ in l_1 norm. Then, one obtains

Lemma 4.2.

$$\|\mathcal{M}_N - \mathcal{M}_N^U\|_{l_1} \leqslant \frac{T^{2\delta}X^2}{U}.$$

Together, Lemma 4.1 and Lemma 4.2 then imply that

Lemma 4.3.

$$\|\mathcal{E}_N - \mathcal{E}_N^U\|_{l_1} \leqslant \frac{T^{2\delta}X^2}{U}.$$

Appropriate lower bounds on $\mathcal{M}_N(n)$ and average upper bounds on $\mathcal{E}_N^U$ are then combined to prove the main theorem. Specifically, in Section 6 we show that

Theorem 4.4. *For any $n \in [N/2, N] \cap \mathcal{K}_a$, we have*

$$\mathcal{M}_N(n) \gg T^{2\delta-2}.$$

In Section 7.2, Section 7.3, and Section 7.4 we work towards giving an l_2 bound for $\mathcal{E}_N^U$:

Theorem 4.5.

$$\|\mathcal{E}_N^U\|_{l_2}^2 \ll T^{4\delta-4}N^{1-\eta}.$$

The value of η will be described in the course of the proof.

Let $\mathbf{1}_{[N/2,N]}$ be the indicator function on $\mathbb{Z}$ defined by

$$\mathbf{1}_{[N/2,N]}(n) = \begin{cases} 1 & \text{if } n \in [N/2, N], \\ 0 & \text{otherwise.} \end{cases}$$

Then using Cauchy-Schwarz, we have

$$\|\mathcal{E}_N^U \cdot \mathbf{1}_{[N/2,N]}\|_{l_1} \leqslant \|\mathcal{E}_N^U\|_{l_2}\|\mathbf{1}_{[N/2,N]}\|_{l_2} \ll T^{2\delta-2}N^{1-\eta}, \tag{4.17}$$

replacing $\eta/2$ by η.

Using (4.17) together with Lemma 4.3, we have

$$\|\mathcal{E}_N \cdot \mathbf{1}_{[N/2,N]}\|_{l_1} \ll T^{2\delta-2}N^{1-\eta}. \tag{4.18}$$

We are now able to prove Theorem 1.6 assuming Theorem 4.4 and (4.18).

Proof of Theorem 1.6:. Let $\mathfrak{E}(N)$ be the set of exceptional numbers $[N/2, N]$ (those admissible but not occurring as curvatures). Then, by (4.18),

$$\sum_{n\in\mathfrak{E}(N)} |\mathcal{E}_N(n)| \leqslant \|\mathcal{E}_N\|_{l_1} \ll T^{2\delta-2}N^{1-\eta}.$$

For $n \in \mathfrak{E}(N)$, $\mathcal{R}_N(n) = 0$ and, by Theorem 4.4, $\mathcal{M}_N(n) \gg T^{2\delta-2}$. Therefore

$$|\mathcal{E}_N(n)| = |\mathcal{R}_N(n) - \mathcal{M}_N(n)| \gg T^{2\delta-2}.$$

Thus

$$\#\mathfrak{E}(N) \cdot T^{2\delta-2} \ll \sum_{n\in\mathfrak{E}(N)} |\mathcal{E}_N(n)| \ll T^{2\delta-2}N^{1-\eta},$$

so that

$$\#\mathfrak{E}(N) \ll N^{1-\eta}. \tag{4.19}$$

This is the desired result for the interval $[N/2, N]$, and we extend it to the full interval $[0, N]$ as follows. Divide $[0, N]$ into a union of subintervals dyadically: $[0, N] = [N/2, N] \cup [N/4, N/2] \cup [N/8, N/4] \cup \cdots$. Applying (4.19) to each subinterval (replacing N by $N/2^m$ for $0 \leqslant m < \log_2(N)$) and collecting the error terms, we obtain Theorem 1.6 as desired. □

5. preliminary lemmata

In this section we introduce several lemmata due to Bourgain-Kontorovich which will be used in later sections. Note that they are not stated exactly as the lemmata which we cite from [6], which are stated in the framework of counting in orbits of the Apollonian group in $\mathrm{O}_{\mathbb{R}}(3, 1)$ acting on Descartes quadruples in $\mathbb{Z}^4$, while we use the lemmata in the context of subgroups of $\mathrm{PSL}_2(\mathcal{O}_K)$ acting on a circle. However, the proofs of these lemmata in [6] are very general, and apply almost verbatim to the context in which we phrase them below, with their group Γ replaced by $\mathcal{A}$ in our case, and the set of first coordinates of points in the orbit of Γ acting on a vector replaced by

the curvatures of the circles one gets as in the orbit of $\mathcal{A}$ that we consider. We note also that in Lemma 5.2 we sum over cosets of $\mathcal{A}(q)$ while Bourgain-Kontorovich sum over cosets of a larger subgroup. However, this is not necessary to execute the circle method as we do here. Finally, as stated below, Bourgain-Kontorovich's bounds involving T^{δ} and T^{Θ} are adjusted to involve $T^{2\delta}$ and $T^{2\Theta}$, respectively. This is because we work in PSL_2 and not in $\mathrm{SO}_{\mathbb{R}}(3,1)$ as is the case in [6], and the spin homomorphism from PSL_2 to $\mathrm{SO}_{\mathbb{R}}(3,1)$ is quadratic in the entries of the matrices of PSL_2.

These results are the point at which the spectral gap for $\mathcal{A}$ feeds into our analysis. The first two of these are statements about equidistribution modulo q. The first says that the curvatures cannot have too strong a preference for a given congruence class modulo q, as γ varies. It is used in the minor arc analysis.

Lemma 5.1 Bourgain-Kontorovich [6], Lemma 5.2. *There exists a positive constant ν and some $\eta_0 > 0$ which only depend on the spectral gap of $\mathcal{A}$, such that for any $1 \leqslant q < N$ and any $r (\mathrm{mod}\ q)$,*

$$\sum_{\gamma \in \mathfrak{F}_T} \mathbf{1}\left\{ \frac{2\Im(C_{M\gamma}\overline{D_{M\cdot\gamma}})}{\sqrt{-\Delta}} \equiv r (\mathrm{mod}\ q) \right\} \ll \frac{T^{2\delta}}{q^{\eta_0}},$$

where $T = T_1 T_2$ (for notations see the definition of $\mathfrak{F}$ in (4.3)). The implied constant is independent of r.

The second lemma states that the behaviour of the form $\mathfrak{f}_{M\gamma}$ on γ from any given congruence class is independent of the congruence class, in the sense that each class contributes equally to an exponential sum. It is used for the setup of the major arc analysis, to separate the non-archimedean and archimedean contributions. Write $\mathcal{A}(q)$ for the kernel of reduction modulo q.

Lemma 5.2 Bourgain-Kontorovich [6], Lemma 5.3. *Let $1 < K < T_2^{\frac{1}{10}}$, fix $|\beta| < \frac{K}{N}$, and fix $x, y \asymp X$. Then for any $\gamma_0 \in \mathcal{A}$, any $q \geqslant 1$, we have*

$$\sum_{\gamma \in \mathfrak{F}_T \cap \gamma_0 \mathcal{A}(q)} e(\beta \mathfrak{f}_{M\gamma}(Lx+1, Ly)) = \frac{1}{[\mathcal{A} : \mathcal{A}(q)]} \sum_{\gamma \in \mathfrak{F}_T} e(\beta \mathfrak{f}_{M\gamma}(Lx+1, Ly)) + O(T^{2\Theta_1} K),$$

where $\Theta_1 < \delta$ depends only on the spectral gap for $\mathcal{A}$, and the implied constant does not depend on q, γ_0, x or y.

The last lemma is used to bound the Archimedean piece of the major arc analysis. It uses the spectral gap to control the error in counting $\gamma \in \mathfrak{F}_T$ where $\mathfrak{f}_{M\gamma}$ takes certain values.

Lemma 5.3 Bourgain-Kontorovich [6], Lemma 5.4. *Fix $N/2 \leqslant n \leqslant N, 1 < K \leqslant T_2^{\frac{1}{10}}$, and $x, y \asymp X$. Then*

$$\sum_{\gamma \in \mathfrak{F}_T} \mathbf{1}\left\{ |\mathfrak{f}_{M\cdot\gamma}(Lx+1, Ly) - n| < \frac{N}{K} \right\} \gg \frac{T^{2\delta}}{K} + T^{2\Theta_2},$$

where $\Theta_2 < \delta$ depends only on the spectral gap for $\mathcal{A}$. The implied constant is independent of x, y and n.

Let Θ be the larger of the Θ's that satisfy Lemma 5.2 and Lemma 5.3 respectively, then we set the two parameters Q_0, K_0 as

$$\boxed{Q_0 = T^{\frac{2\delta - 2\Theta}{80}}, K_0 = Q_0^3.} \tag{5.1}$$

6. Major arc analysis

In this section we prove Theorem 4.4 bounding $\mathcal{M}_N(n)$ below. We give a brief overview of the argument, before treating all the details. First, we will write

$$\mathcal{M}_N(n) = \sum_{x,y \text{ in a region}} \mathfrak{S}_{Q_0}(n)\mathfrak{M}(n) + \text{error},$$

where $\mathfrak{M}(n)$ is the *Archimedean part* and $\mathfrak{S}_{Q_0}(n)$ is the *non-Archimedean part* (depending on a parameter Q_0 controlling the size of the major arcs); both depend on x, y. We need Lemma 5.2 (dependent on the spectral gap) in order to accomplish this separation of Archimedean from non-Archimedean.

We can use Lemma 5.3 to give a lower bound on the Archimedean part, and most of the attention of this section is given to bounding $\mathfrak{S}_{Q_0}(n)$ below. This requires a careful local analysis that is one of the novelties of our treatment as compared with previous works [6, 37].

The limit $\mathfrak{S}(n) = \lim_{Q_0\to\infty} \mathfrak{S}_{Q_0}(n)$ is the *singular series*, whose purpose is to be supported only on the admissible values of n, and bounded below where it is supported. We break it down as

$$\mathfrak{S}(n) = \sum_{q=1}^{\infty} B_q(n) = \prod_p (1 + B_p(n) + B_{p^2}(n) + \cdots).$$

In turn,

$$B_q(n) = \sum_{r(q)} \tau_q(r) c_q(r-n),$$

where $\tau_q(r)$ is the probability of the quadratic form $\mathfrak{f}_{M\gamma}$ taking on the value r modulo q, as γ ranges among cosets of $\mathcal{A}$ modulo q, and c_q is a Ramanujan sum, which is multiplicative with respect to q. For a prime p, one should think of $B_p(n)$ as measuring some deviation from the equidistribution of the probabilities $\tau_p(n)$ modulo p; $B_{p^k}(n)$ for larger k gives finer information about the behaviour of these probabilities as we lift to powers of p. This is captured by the relationship

$$1 + B_p(n) + B_{p^2}(n) + \cdots + B_{p^k}(n) = p^k \tau_{p^k}(n).$$

This factor is non-zero if and only if n is represented as a curvature modulo p^k.

The goal, then, is to understand $B_{p^k}(n)$. First, we use strong approximation for $\mathcal{A}$ to show that the $B_{p^k}(n)$ eventually vanish as k increases. In particular, $B_{p^k}(n) = 0$ once we have uniform lifting in the sense of strong approximation (Lemmas 6.3 and 6.4). We find that for all but finitely many primes, $B_{p^k}(n) = 0$ for $k \geqslant 2$. Therefore $\mathfrak{S}(n)$ is controlled by the product over good primes $\prod_p'(1 + B_p(n))$.

The final step is to control $B_p(n)$: we show that for $p \nmid n$, $B_p(n) = O(1/p)$, while for $p \mid n$, $B_p(n) = O(1/p^2)$. This requires a direct counting argument, finding all solutions modulo p to the requirement that the curvature be equal to n; at its core is an argument using Gauss sums. In other words, we show that equidistribution of curvatures modulo p does not fail too badly.

Now we begin. From (4.13), (4.12) and (4.9), we have

$$\begin{aligned}\mathcal{M}_N(n) &= \int_0^1 \mathfrak{T}(\theta)\widehat{\mathcal{R}}_N(\theta)e(-n\theta)d\theta \\ &= \int_{-\infty}^{\infty} \sum_{q<Q_0}\sum_{r(q)}{}'\mathfrak{t}\left(\frac{N}{K_0}\beta\right)\widehat{\mathcal{R}}_N\left(\beta+\frac{r}{q}\right)e\left(-n\left(\beta+\frac{r}{q}\right)\right)d\beta \\ &= \sum_{\substack{x,y\in\mathbb{Z}\\(Lx+1,Ly)=1}} \psi\left(\frac{Lx+1}{X}\right)\psi\left(\frac{Ly}{X}\right)\sum_{q<Q_0}\sum_{r(q)}{}'\sum_{\gamma\in\mathfrak{F}_T} e\left(\frac{r}{q}(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n)\right) \\ &\quad\cdot\int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)e(\beta(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n)d\beta \qquad (6.1)\end{aligned}$$

Now we decompose the set $\mathfrak{F}$ as the intersection of $\mathfrak{F}$ and the left cosets of

$$\mathcal{A}(q)=\left\{\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathcal{A}\Big|\begin{pmatrix}a&b\\c&d\end{pmatrix}\equiv I\pmod q\right\}$$

and apply Lemma 5.2 with $K=K_0$ to obtain

$$\begin{aligned}&\sum_{\gamma\in\mathfrak{F}} e\left(\frac{r}{q}(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n)\right)\cdot\int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)e(\beta(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n))d\beta \\ =&\sum_{\gamma_0\in\mathcal{A}/\mathcal{A}(q)} e\left(\frac{r}{q}(\mathfrak{f}_{M\gamma_0}(Lx+1,Ly)-n)\right)\int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)\sum_{\gamma\equiv\gamma_0(q)} e(\beta(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n))d\beta \\ =&\frac{1}{[\mathcal{A}:\mathcal{A}(q)]}\sum_{\gamma_0\in\mathcal{A}/\mathcal{A}(q)} e\left(\frac{r}{q}(\mathfrak{f}_{M\gamma_0}(Lx+1,Ly)-n)\right)\int_{-\infty}^{\infty}\mathfrak{t}\left(\frac{N}{K_0}\beta\right)\sum_{\gamma\in\mathfrak{F}_T} e(\beta(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n))d\beta \\ &+O\left(\frac{T^{2\Theta}K_0^2Q_0^6}{N}\right) \\ =&\frac{K_0}{N}\cdot\frac{1}{[\mathcal{A}:\mathcal{A}(q)]}\sum_{\gamma_0\in\mathcal{A}/\mathcal{A}(q)} e\left(\frac{r}{q}(\mathfrak{f}_{M\gamma_0}(Lx+1,Ly)-n)\right)\cdot\sum_{\gamma\in\mathfrak{F}_T}\hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n)\right) \\ &+O\left(\frac{T^{2\Theta}K_0^2Q_0^6}{N}\right), \qquad (6.2)\end{aligned}$$

where Lemma 5.2 is applied to obtain the third line above. Inserting (6.2) into (6.1), we get

$$\mathcal{M}_N(n)=\sum_{\substack{x,y\in\mathbb{Z}\\(Lx+1,Ly)=1}}\psi\left(\frac{Lx+1}{X}\right)\psi\left(\frac{Ly}{X}\right)\mathfrak{S}_{Q_0}(n)\mathfrak{M}(n)+O\left(\frac{T^{2\Theta}X^2K_0^2Q_0^8}{N}\right) \qquad (6.3)$$

where

$$\mathfrak{S}_{Q_0}(n)=\mathfrak{S}_{Q_0;x,y}(n):=\sum_{q<Q_0}\frac{1}{[\mathcal{A}:\mathcal{A}(q)]}\sum_{\gamma_0\in\mathcal{A}/\mathcal{A}(q)} c_q\left(\mathfrak{f}_{M\cdot\gamma_0}(Lx+1,Ly)-n\right) \qquad (6.4)$$

and

$$\mathfrak{M}(n)=\mathfrak{M}_{x,y}(n):=\frac{K_0}{N}\sum_{\gamma\in\mathfrak{F}}\hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}_{M\gamma}(Lx+1,Ly)-n)\right). \qquad (6.5)$$

Here c_q is the Ramanujan sum defined by

$$c_q(n) = \sum_{r(q)}{}' e\left(\frac{rn}{q}\right). \tag{6.6}$$

Fixing n, we have that $c_q(n)$ is multiplicative with respect to q, and locally,

$$c_{p^k}(n) = \begin{cases} 0 & \text{if } p^m \| n, m \leqslant k-2, \\ -p^{k-1} & \text{if } p^{k-1} \| n, \\ p^{k-1}(p-1) & \text{if } p^k | n. \end{cases} \tag{6.7}$$

The error term in (6.3) is $O(T^{2\delta-2-\epsilon})$ by our choice of K_0 (see (4.1) and (5.1)), where ϵ is any small positive number at most $\frac{33}{20}(\delta-\Theta) > 0$. Applying Lemma 5.3 with $K = K_0$, we can give a lower bound for the Archimedean piece $\mathfrak{M}(n)$ for any $N/2 \leqslant n \leqslant N$:

$$\mathfrak{M}(n) \gg \frac{T^{2\delta}}{N}. \tag{6.8}$$

Therefore, Theorem 4.4 is proved once we show that $\mathfrak{S}_{Q_0}(n) \gg 1$ for every n admissible (or, what actually suffices, once we have proven it up to log factors, since our aim is to get a power saving, which absorbs all log powers). The rest of this section is devoted to proving the following.

Proposition 6.1. *We have $\mathfrak{S}_{Q_0}(n) \gg \frac{1}{\log n}$ if n is admissible, and $\mathfrak{S}_{Q_0}(n) \ll \frac{\log n}{Q_0}$ if n is not admissible.*

To understand $\mathfrak{S}_{Q_0}(n)$, we first push Q_0 to ∞. We define a formal singular series

$$\mathfrak{S}(n) = \sum_{q=1}^{\infty} B_q(n), \tag{6.9}$$

where

$$B_q(n) = \frac{1}{[\mathcal{A} : \mathcal{A}(q)]} \sum_{\gamma_0 \in \mathcal{A}/\mathcal{A}(q)} c_q\left(\mathfrak{f}_{M\gamma_0}(Lx+1, Ly) - n\right). \tag{6.10}$$

So to understand $\mathfrak{S}_{Q_0}(n)$ or $\mathfrak{S}(n)$ one must understand $B_q(n)$ for each q.

We rewrite $B_q(n)$ as

$$B_q(n) = \sum_{r(q)} \tau_q(r) c_q\left(r - n\right), \tag{6.11}$$

where

$$\tau_q(r) = \frac{1}{[\mathcal{A} : \mathcal{A}(q)]} \sum_{\gamma_0 \in \mathcal{A}/\mathcal{A}(q)} \mathbf{1}\{\mathfrak{f}_{M\gamma_0}(Lx+1, Ly) \equiv r \pmod q\} \tag{6.12}$$

$$= \frac{1}{[\mathcal{A} : \mathcal{A}(q)]} \sum_{\gamma_0 \in \mathcal{A}/\mathcal{A}(q)} \mathbf{1}\left\{\kappa\left(M\gamma_0\left(\mathbb{R} + \frac{\sqrt{\Delta}}{2}\right)\right) \equiv r \pmod q\right\}. \tag{6.13}$$

The term $\tau_q(r)$ can be viewed as the probability that a curvature is congruent to r mod q, as γ ranges over $\mathcal{A}$. To get from (6.12) to (6.13) we used the fact that $\mathfrak{f}_{M\gamma}(Lx+1, Ly) = \kappa\left(M\gamma w_{x,y}\left(\mathbb{R} + \frac{\sqrt{\Delta}}{2}\right)\right)$ for some $w_{x,y} \in \Gamma(L)$ with left column $(Lx+1, Ly)^T$.

First we need the multiplicativity of $\mathcal{A}$ which will lead to the multiplicativity of $B_q(n)$:

LEMMA 6.2. *Write $q = \prod_i p_i^{n_i}$, then*

$$\mathcal{A}(q) \cong \prod_i \mathcal{A}(p_i^{n_i}).$$

Lemma 6.2 will lead immediately to the multiplicativity of $B_q(n)$ with respect to q. A priori Lemma 6.2 is not true for a general group $\mathcal{A}$. If this is the case, we replace $\mathcal{A}$ by some congruence subgroup of $\mathcal{A}$ which satisfies the multiplicative property (such a subgroup exists by strong approximation in SL_2). As noted in Section 2, we may move to a finite index subgroup without loss of generality.

Given this multiplicativity, we split (6.9) into an Euler product

$$\mathfrak{S}(n) = \prod_p (1 + B_p(n) + B_{p^2}(n) + \cdots). \tag{6.14}$$

The arithmetic meaning of each factor of the Euler product is illustrated by the following formula:

$$1 + B_p(n) + B_{p^2}(n) + \cdots + B_{p^k}(n) = p^k \tau_{p^k}(n). \tag{6.15}$$

To see this, let s_γ be such that $p^{s_\gamma} \mid\mid \mathfrak{f}_{M\gamma}(Lx+1, Ly) - n$. Then,

$$\begin{aligned}
&1 + \sum_{m=1}^{k} B_{p^m}(n) \\
&= \frac{1}{[\mathcal{A} : \mathcal{A}(p^k)]} \sum_{\gamma \in \mathcal{A}/\mathcal{A}(p^k)} \left(1 + \sum_{m=1}^{k} c_{p^m}(\mathfrak{f}_{M\gamma}(Lx+1, Ly) - n) \right) \\
&= \frac{1}{[\mathcal{A} : \mathcal{A}(p^k)]} \sum_{\gamma \in \mathcal{A}/\mathcal{A}(p^k)} \begin{cases} 0 & s_\gamma < k \\ p^k & s_\gamma \geqslant k \end{cases}.
\end{aligned}$$

Therefore, $1 + B_p(n) + B_{p^2}(n) + \cdots + B_{p^k}(n)$ is non-zero if and only if n is represented (mod p^k).

Our goal for the rest of the section is to access $\mathfrak{S}(n)$ (and prove Proposition 6.1) by analysing the values of $B_{p^k}(n)$. First, we will show that

LEMMA 6.3. *There is an integer $P_{bad} \geqslant 1$ such that*

(i) *For any $p \nmid P_{bad}$ and $k \geqslant 2$, $B_{p^k}(n) = 0$.*

(ii) *For each of the finitely many primes $p \mid P_{bad}$, $\exists k'_p$ such that $B_{p^k}(n) = 0$ for any $k \geqslant k'_p$.*

Indeed, Lemma 6.3 follows from the following fact for $\mathcal{A}(q)$ given by strong approximation in SL_2:

LEMMA 6.4. *There is an integer $L_1 \geqslant 1$ such that*

(i) *For any $p \nmid L_1$ and $k \geqslant 1$,*

$$\mathcal{A}(p^{k-1})/\mathcal{A}(p^k) = SL_2(\mathcal{O}_K)(p^{k-1})/SL_2(\mathcal{O}_K)(p^k) \tag{6.16}$$

(ii) *For each of the finitely many primes $p \mid L_1$, $\exists k_p$ such that (6.16) holds for any $k \geqslant k_p$.*

We refer to primes that divide P_{bad} as *bad primes*, and those that do not divide P_{bad} as *good primes*. We given an explicit form of Lemma 6.4 in Theorem 8.1, which allows the computation of a valid L_1. We use a Hensel lifting argument to deduce Lemma 6.3 from Lemma 6.4.

Proof of Lemma 6.3. First we rewrite $B_{p^k}(n)$:

$$\begin{aligned} B_{p^k}(n) =& \frac{1}{[\mathcal{A} : \mathcal{A}(p^k)]} \sum_{\gamma \in \mathcal{A}/\mathcal{A}(p^k)} c_{p^k}(F_1(\gamma) - n) \\ =& \frac{1}{[\mathcal{A} : \mathcal{A}(p^k)]} \sum_{\gamma_0 \in \mathcal{A}/\mathcal{A}(p^{k-1})} \sum_{\substack{\gamma \in \mathcal{A}/\mathcal{A}(p^k) \\ \gamma \equiv \gamma_0 (p^{k-1})}} c_{p^k}(F_1(\gamma) - n) \end{aligned} \tag{6.17}$$

where $F_1(\gamma) = \kappa(M\gamma(\widehat{\mathbb{R}} + \frac{\sqrt{\Delta}}{2}))$ and we view F_1 as an algebraic function over the real and imaginary parts of the entries of γ. As we have assumed $\mathcal{A} \subset \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$ in Section 2, we may write

$$\gamma = \begin{pmatrix} a_1 + a_2\sqrt{d}\mathbf{i} & b_1 + b_2\sqrt{d}\mathbf{i} \\ c_1 + c_2\sqrt{d}\mathbf{i} & d_1 + d_2\sqrt{d}\mathbf{i} \end{pmatrix}.$$

We assume for the moment that M is also in $\mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$, and write

$$M = \begin{pmatrix} M_{11} + M_{12}\sqrt{d}\mathbf{i} & M_{21} + M_{22}\sqrt{d}\mathbf{i} \\ M_{31} + M_{32}\sqrt{d}\mathbf{i} & M_{41} + M_{42}\sqrt{d}\mathbf{i} \end{pmatrix}.$$

Then

$$F_1(\gamma) = F_1(a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2) \tag{6.18}$$

$$= (M_{31}a_1 + M_{41}c_1)^2 + d(M_{32}a_2 + M_{42}c_2)^2 + (M_{31}a_1 + M_{41}c_1)(M_{32}b_2 + M_{42}d_2) \tag{6.19}$$

$$- (M_{31}b_1 + M_{41}d_1)(M_{32}a_2 + M_{42}c_2). \tag{6.20}$$

As $\gamma \in \mathrm{PSL}_2(\mathcal{O}_K)$, these variables are also subject to the following conditions:

$$\begin{aligned} F_2(a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2) &:= a_1d_1 - a_2d_2d - b_1c_1 + b_2c_2d - 1 = 0 \\ F_3(a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2) &:= a_1d_2 + a_2d_1 - b_1c_2 - b_2c_1 = 0 \end{aligned} \tag{6.21}$$

If M is integral, one can check that the Jacobian matrix

$$J = \frac{\partial(F_1, F_2, F_3)}{\partial(a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2)}$$

maps $\mathbb{Z}_p^8$ onto $\mathbb{Z}_p^3$ as a linear transformation, at each point of the affine variety $\mathbb{V}[\mathbb{Q}_p]$ defined by the following equations:

$$\begin{cases} F_1 = n \\ F_2 = 0 \\ F_3 = 0 \end{cases}.$$

For any $k \geqslant k_p$ (k_p can be taken to be 2 if p is good), the lifting $\mathbb{V}[\mathbb{Z}/p^{k-1}\mathbb{Z}] \to \mathbb{V}[\mathbb{Z}/p^k\mathbb{Z}]$ becomes regular by Lemma 6.4, and for γ_0 such that $F_1(\gamma_0) \equiv n \pmod{p^{k-1}}$, this gives

$$\mathrm{Prob}\left(F_1(\gamma) \equiv n \pmod{p^k} \;\middle|\; \gamma \in \mathcal{A}/\mathcal{A}(p^k), \gamma \equiv \gamma_0 \pmod{p^{k-1}}\right) = \frac{1}{p}. \tag{6.22}$$

Returning to (6.17), from (6.7) the innermost sum of (6.17) is zero unless $F_1(\gamma_0) \equiv n \pmod{p^{k-1}}$. However, in this case, if $k \geqslant k_p$, by (6.22) (at k and $k+1$) and (6.7), each innermost sum of (6.17) is still zero. Therefore $B_{p^k}(n) = 0$, for $k \geqslant k_p$.

Above, we assumed $M \in \mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$, and we found that the k_p' of Lemma 6.3 agree with the k_p of Lemma 6.4. If instead M is fractional in the sense that the M_{ij} have denominator

q_0, then we need to multiply F_1 by q_0^2 to make it integral. For any $p^{n_p}||q_0$, one can check that $\mathbb{Z}_p^3 \subset \frac{1}{p^{2n_p}} J(\mathbb{Z}_p^8)$. In this case, $B_{p^k}(n) = 0$ for $k \geqslant k_p + 2n_p$. □

The obstruction number L_0 in the statement of Corollary 1.4 is thus given by

$$L_0 = \prod_p p^{k_p+2n_p}. \tag{6.23}$$

The computation of upper bounds on k_p is given by Theorem 8.1, and some examples are given in Section 9.

At this point, in order to prove Proposition 6.1, as there are only finitely many bad primes, we have shown that it suffices to analyze the contribution of $1 + B_p(n)$ for good odd primes p.

LEMMA 6.5. *Suppose n is admissible. Let p be an odd prime not dividing P_{bad}. Then $B_p(n) = O(1/p^2)$ if $p \nmid n$ and $B_p(n) = O(1/p)$ if $p \mid n$, where the implied constants are independent of n.*

To prove this, we first prove the following.

LEMMA 6.6. *Let p be an odd prime not dividing P_{bad}.*

$$\tau_p(n) = \begin{cases} \frac{1}{p} + O(\frac{1}{p^3}) & \text{if } n \not\equiv 0 \pmod p \\ \frac{1}{p} + O(\frac{1}{p^2}) & \text{if } n \equiv 0 \pmod p \end{cases}, \tag{6.24}$$

where the implied constants are independent of n.

There are at least two proofs of this fact. One is the proof we give below, which works directly in the group $\mathrm{SL}_2(\mathcal{O}_K)$. Another approach is to consider the image under the spin homomorphism ρ of $\mathcal{A}$ in $\mathrm{O}_Q(\mathbb{Z})$ where $Q(x_1, x_2, x_3, x_4) = x_2^2 + dx_3^2 + x_1x_4$, and note that the set of curvatures we are interested in is, up to a factor of d, exactly the set of fourth coordinates of points in the orbit $\rho(\mathcal{A})v^T$, where $v = (-d, 0, 1, 0)$. By strong approximation, modulo p the orbit $\rho(\mathcal{A})v^T$ is simply the set of all solutions to $Q(x_1, x_2, x_3, x_4) \equiv d \pmod p$, and $\tau_p(n)$ is easily computed by counting representations modulo p of d by specific quadratic forms. This passing between $\mathrm{SL}_2(\mathbb{C})$ and $\mathrm{O}_{\mathbb{R}}(3, 1)$ is a nod to the description of Apollonian circle packings in [6], [14] and [37], where curvatures can be seen by looking at orbits of certain thin subgroups of $\mathrm{O}_F(\mathbb{Z})$ as described in the introduction. Since we describe Apollonian packings somewhat more geometrically, we present the proof from that point of view.

Proof. Let p be an odd prime not dividing P_{bad}. Let $\gamma \in \mathcal{A}$. Write γ_p for the reduction of γ in $\mathcal{A}/\mathcal{A}(p)$. By Lemma 6.4, we have that γ_p ranges over all of

$$\mathrm{SL}_2(\mathbb{Z}[\sqrt{-d}])/\mathrm{SL}_2(\mathbb{Z}[\sqrt{-d}])(p) = \mathrm{SL}_2(\mathbb{Z}[\sqrt{-d}]/(p)).$$

Therefore, we have

$$\tau_p(n) = \frac{\#\mathbb{V}[\mathbb{Z}/p\mathbb{Z}]}{\#\mathrm{SL}_2(\mathbb{Z}[\sqrt{-d}]/(p))}.$$

For any commutative ring R with identity, the allowable first columns of $\mathrm{SL}_2(R)$ is a set

$$U(R) = \{\text{pairs } (a, b) \mid \text{as ideals, } (a, b) = R\}.$$

We have that $\#U(R) = \#\mathbb{P}^1(R) \cdot \#R^*$. Furthermore,

$$\#\mathrm{SL}_2(R) = \#U(R) \cdot \#\,\mathrm{Stab}_*(\mathrm{SL}_2(R)) = \#U(R) \cdot \#R,$$

where we write Stab_* for the stabilizer of any one element of U. In our case, $R = \mathbb{Z}[\sqrt{-d}]/(p)$, this implies

$$\tau_p(n) = \frac{\#\mathbb{V}[\mathbb{Z}/p\mathbb{Z}]}{p^2 \#\mathbb{P}^1\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)\#\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)^*}.$$

We have

$$\#\mathbb{P}^1\left(\mathbb{Z}[\sqrt{-d}]/(p)\right) = \begin{cases} p^2+1 & \text{if } \left(\frac{-d}{p}\right) = -1 \\ p^2+2p+1 & \text{if } \left(\frac{-d}{p}\right) = 1 \end{cases}. \tag{6.25}$$

and

$$\#\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)^* = \begin{cases} p^2-1 & \text{if } \left(\frac{-d}{p}\right) = -1 \\ p^2-2p+1 & \text{if } \left(\frac{-d}{p}\right) = 1 \end{cases}. \tag{6.26}$$

It remains to compute $\#\mathbb{V}[\mathbb{Z}/p\mathbb{Z}]$. But we have

$$\begin{aligned}\#\mathbb{V}[\mathbb{Z}/p\mathbb{Z}] &= \#\{\lambda \in \mathrm{SL}_2\left(\mathbb{Z}[\sqrt{-d}]/(p)\right) : F_1(\lambda) = n\} \\ &= \#\{v \in U\left(\mathbb{Z}[\sqrt{-d}]/(p)\right) : F_1(v) = n\} \cdot \#\,\mathrm{Stab}_*\left(\mathrm{SL}_2\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)\right) \\ &= p^2 \#\{v \in U\left(\mathbb{Z}[\sqrt{-d}]/(p)\right) : F_1(v) = n\}.\end{aligned}$$

In the above, we use the notation $F_1(v) = F_1(\lambda)$ for any λ having bottom row v (upon which F_1 depends exclusively).

Therefore, it remains to compute

$$\#\left\{v \in U\left(\mathbb{Z}[\sqrt{-d}]/(p)\right) : F_1(v) = n\right\}.$$

If we assume that $M = I$, then we can write the equation $F_1(\lambda) = F_1(v) = n$ explicitly in terms of

$$\lambda = \begin{pmatrix} a_1 + a_2\sqrt{d}\mathbf{i} & b_1 + b_2\sqrt{d}\mathbf{i} \\ c_1 + c_2\sqrt{d}\mathbf{i} & d_1 + d_2\sqrt{d}\mathbf{i} \end{pmatrix}$$

as

$$c_1^2 + c_2^2 d + (c_1 d_2 - c_2 d_1) - n \equiv 0 \pmod{p}. \tag{6.27}$$

We count the number of solutions by evaluating the following exponential sum:

$$\begin{aligned}&\frac{1}{p}\sum_{s(p)}\sum_{c_1,c_2,d_1,d_2(p)} e_p\left(s(c_1^2 + c_2^2 d + c_1 d_2 - c_2 d_1 - n)\right) \\ =&\frac{1}{p}\sum_{s(p)}\sum_{c_1,c_2,d_1,d_2(p)} e_p\left(s(c_1 + d_2/2)^2 + sd(c_2 - d_1/2d)^2 - sd_2^2/4 - sd_1^2/4d - sn)\right) \\ =&\frac{1}{p}\sum_{s\equiv 0(p)} \cdots + \frac{1}{p}\sum_{s\not\equiv 0(p)} \cdots \\ =&p^3 + \frac{1}{p}\cdot\sum_{s\not\equiv 0(p)} p^2\left(\frac{s}{p}\right)\left(\frac{sd}{p}\right)\left(\frac{-s}{p}\right)\left(\frac{-s/d}{p}\right)\cdot e_p(-sn) \\ =&\begin{cases} p^3 + p(p-1) & \text{if } n \equiv 0 \pmod{p} \\ p^3 - p & \text{if } n \not\equiv 0 \pmod{p} \end{cases}\end{aligned}$$

where $\left(\frac{\cdot}{\cdot}\right)$ is the Legendre symbol and we obtained the second to last step by applying Gauss sums first to c_1, c_2, then to d_1, d_2.

To obtain $\#\mathbb{V}[\mathbb{Z}/p\mathbb{Z}]$ we need to subtract the contribution from solutions not in $U\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)$. It turns out if $n \equiv 0 \pmod p$ then all such are solutions to (6.27); if $n \not\equiv 0 \pmod p$ then none such are solutions. We thus arrive at the following result:

$$\#V[\mathbb{Z}/p\mathbb{Z}] = p^2 \cdot \begin{cases} p^3 + p(p-1) - 1 & \text{if } \left(\frac{-d}{p}\right) = -1 \text{ and } n \equiv 0 \pmod p \\ p^3 - p & \text{if } \left(\frac{-d}{p}\right) = -1 \text{ and } n \not\equiv 0 \pmod p \\ p^3 - p^2 - p + 1 & \text{if } \left(\frac{-d}{p}\right) = 1 \text{ and } n \equiv 0 \pmod p \\ p^3 - p & \text{if } \left(\frac{-d}{p}\right) = 1 \text{ and } n \not\equiv 0 \pmod p \end{cases}.$$

Now, if $M \neq I$, the effect of M on the equation (6.27) is to apply an invertible linear transformation to $\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)^2$ (recall that we are dealing only with good primes p). This takes $U\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)$ to $U\left(\mathbb{Z}[\sqrt{-d}]/(p)\right)$. Therefore, the number of solutions $\#\mathbb{V}[\mathbb{Z}/p\mathbb{Z}]$ is unaffected.

Therefore, we obtained the formula for $\tau_p(n)$:

$$\tau_p(n) = \begin{cases} \frac{p+1}{p^2+1} & \text{if } \left(\frac{-d}{p}\right) = -1 \text{ and } n \equiv 0 \pmod p \\ \frac{p}{p^2+1} & \text{if } \left(\frac{-d}{p}\right) = -1 \text{ and } n \not\equiv 0 \pmod p \\ \frac{1}{p+1} & \text{if } \left(\frac{-d}{p}\right) = 1 \text{ and } n \equiv 0 \pmod p \\ \frac{p}{p^2-1} & \text{if } \left(\frac{-d}{p}\right) = 1 \text{ and } n \not\equiv 0 \pmod p \end{cases}.$$

and indeed we have that $\tau_p(n) = \frac{1}{p} + O(\frac{1}{p^3})$ if $n \not\equiv 0 \pmod p$ and $\tau_p(n) = \frac{1}{p} + O(\frac{1}{p^2})$ if $n \equiv 0 \pmod p$ as desired. □

Proof of Lemma 6.5:
Recall from (6.11) that

$$\begin{aligned} B_p(n) &= \sum_{r(p)} \tau_p(r) c_p(r-n) \\ &= \tau_p(n)(p-1) + \sum_{\substack{r(p) \\ r \not\equiv n(p)}} -\tau_p(r) \\ &= \tau_p(n)(p-1) - (1 - \tau_p(n)) \\ &= p\tau_p(n) - 1 \end{aligned}$$

Now apply Lemma 6.6. □

We now combine everything to obtain an estimate of $\mathfrak{S}(n)$.

Lemma 6.7. *The term $\mathfrak{S}(n) \neq 0$ if and only if n is admissible, and when n is admissible, we have $\mathfrak{S}(n) \gg \frac{1}{\log n}$.*

Proof. We have already observed that n is admissible if and only if it is represented modulo all integers, which occurs if and only if $\mathfrak{S}(n) \neq 0$. If n is admissible, Lemma 6.3 demonstrates that $\mathfrak{S}(n) \asymp \prod_{p \text{ good}}(1 + B_p(n))$. Lemma 6.5 shows that $1 + B_p(n) = 1 + O(1/p)$ or $1 + O(1/p^2)$; the contribution from the latter converges, and the contribution from the former is bounded below by $1/\log n$. □

Finally, we show that the difference between $\mathfrak{S}_{Q_0}(n)$ and $\mathfrak{S}(n)$ is indeed small:

Lemma 6.8. *We have*

$$|\mathfrak{S}_{Q_0}(n) - \mathfrak{S}(n)| \ll \frac{\log n}{Q_0}.$$

Recall here that Q_0 is a small power of N.

Proof. Let L_1 be as in Lemma 6.4. Write $q = q_1q_2q_3$, where $q_1 = (q, L_1), q_2 = (q/q_1, n)$, so that $(q_3, L_1n) = 1$. Noting that $B_{q_1}(n)$ has a universal upper bound, and recalling that $B_q(n)$ is multiplicative with respect to q, we have

$$\begin{aligned}
|\mathfrak{S}_{Q_0}(n) - \mathfrak{S}(n)| &\leqslant \sum_{q>Q_0} |B_q(n)| \\
&= \sum_{q_1|L_1} |B_{q_1}(n)| \sum_{\substack{(q_2,L_1)=1\\ q_2|n}} |B_{q_2}(n)| \sum_{\substack{(q_3,L_1n)=1\\ q_1q_2q_3\geqslant Q_0}} |B_{q_3}(n)| \\
&\ll \sum_{q_1|L_1} \sum_{\substack{(q_2,L_1)=1\\ q_2|n}} \frac{1}{q_2} \sum_{\substack{(q_3,L_1n)=1\\ q_3\geqslant \frac{Q_0}{q_1q_2}}} \frac{1}{q_3^2} \ll \sum_{q_2|n} \frac{1}{q_2}\frac{q_2}{Q_0} \ll \frac{\log n}{Q_0}
\end{aligned}$$

as desired. □

Lemma 6.7 and Lemma 6.8 together imply Proposition 6.1. Therefore, by the discussion preceding Proposition 6.1, we have shown Theorem 4.4.

7. minor arcs

The aim of this section is to prove

$$\int_0^1 (1 - \mathfrak{T}(\theta))^2 |\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta \ll T^{2\delta-2}N^{1-\eta}. \tag{7.1}$$

By Plancherel's theorem, (7.1) leads to Theorem 4.5.

We bound the integral (7.1) above by $\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3$, where J is the depth of approximation (see (4.11)) and

$$\mathcal{I}_1 = \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \int_{\frac{r}{q}-\frac{1}{qJ}}^{\frac{r}{q}+\frac{1}{qJ}} |(1 - \mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta, \tag{7.2}$$

$$\mathcal{I}_2 = \sum_{Q_0\leqslant q<X} \sideset{}{'}\sum_{r(q)} \int_{\frac{r}{q}-\frac{1}{qJ}}^{\frac{r}{q}+\frac{1}{qJ}} |(1 - \mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta, \tag{7.3}$$

$$\mathcal{I}_3 = \sum_{X\leqslant q\leqslant J} \sideset{}{'}\sum_{r(q)} \int_{\frac{r}{q}-\frac{1}{qJ}}^{\frac{r}{q}+\frac{1}{qJ}} |(1 - \mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta. \tag{7.4}$$

The integrand is periodic on $\mathbb{R}$ modulo 1, and by Dirichlet's Theorem on Diophantine approximation, the domains of these integrals cover the circle $\mathbb{R}$ modulo 1.

The first integral $\mathcal{I}_1$ concerns small q in the range of the major arc analysis, the second integral $\mathcal{I}_2$ concerns q in the intermediate range $Q_0 \leqslant q < X$, and the last integral $\mathcal{I}_3$ concerns large q.

In Section 7.2 we show

$$\mathcal{I}_1 \ll T^{4\delta-4}N^{1-\eta}. \tag{7.5}$$

Then, in Sections 7.3 and 7.4 we divide $[Q_0, X]$ dyadically and prove

$$\mathcal{I}_Q := \sum_{Q<q\leqslant 2Q} \sum_{r(q)}{}' \int_{\frac{r}{q}-\frac{1}{qJ}}^{\frac{r}{q}+\frac{1}{qJ}} |\widehat{\mathcal{R}}_N^U(\theta)|^2 d\theta \ll T^{4\delta-4}N^{1-\eta}, \tag{7.6}$$

where $Q_0 \leqslant Q < X$ and $X \leqslant Q \leqslant J$ respectively. In doing this, we deal with the ranges of Q corresponding to $\mathcal{I}_1$ and $\mathcal{I}_2$ separately and this will give the desired upper bounds on those sums.

It is evident that whether or not M is fractional has little effect in the minor arc analysis: the main player here is the congruence subgroup $\Gamma(L)$ which gives rise to shifted quadratic forms. We can simply replace the shifted quadratic form by a constant multiple of the form, and the analysis will run in exactly the same way.

7.1 Lemmata for minor arcs

In this section, we include some lemmata which will be used in the minor arc analysis. The reader can choose to continue to the next section and refer back here for statements. These lemmata relate to the evaluation and bounds for exponential sums of the form

$$S(q, A, B, C, D, E) = \sum_{x,y(q)} e(Ax^2 + Bxy + Cy^2 + Dx + Ey). \tag{7.7}$$

and certain of their averages. For simplicity we assume q is odd. For $z \in \mathbb{Q}_p$, we define

$$\deg_{p^m}(z) = \max_{-\infty<k\leqslant m} \{k : p^{-k}z \in \mathbb{Z}_p\}.$$

We need the following lemma, which is a direct corollary of Gauss sums (see Page 13 of [10]).

Lemma 7.1. *For $a, b \in \mathbb{Z}$, we have*

$$\sum_{x\in\mathbb{Z}/p^m\mathbb{Z}} e_{p^m}(ax^2 + bx) =$$

$$\begin{cases} p^m \cdot \mathbf{1}\{p^m|b\} & \text{if } p^m|a \\ p^{m/2}(p^m, a)^{1/2} i^{\epsilon\left(\frac{p^m}{(p^m,a)}\right)} \left(\frac{\frac{a}{(p^{m-1},a)}}{p}\right) e_p\left(-\frac{b^2}{4a}\right) \cdot \mathbf{1}\{deg_{p^m}(b) \geqslant deg_{p^m}(a)\} & \text{if } p^m \nmid a \end{cases}, \tag{7.8}$$

where $\epsilon(n) = 0$ if $n \equiv 1 \pmod 4$ and $\epsilon(n) = 1$ if $n \equiv 3 \pmod 4$, and $\left(\frac{\cdot}{\cdot}\right)$ is the Legendre symbol.

The Legendre symbol $\left(\frac{a}{p}\right) = 1$ if a is a quadratic residue, and -1 if it is a quadratic non-residue. By convention we also let $\left(\frac{a}{p}\right) = 1$ if $a \equiv 0 \pmod p$. The Legendre symbol is multiplicative only on the set of nonzero congruence classes mod p.

Write $g(x, y) = Ax^2 + Bxy + Cy^2$ and $\Delta_g = B^2 - 4AC$. Let $k_g = \deg_{p^m}(A, B, C)$. If $k_g < m$ and $p^m|\Delta_g/p^{k_g}$, we say g is *degenerate* at p^m; in this case g is essentially a quadratic form of only one variable.

From Lemma 7.1, we obtain:

Lemma 7.2. *Let p be an odd prime. Let $k_g = deg_{p^m}(gcd(A,B,C))$. If $k_g = m$, then*

$$S(p^m, A, B, C, D, E) = p^{2m}\mathbf{1}\{p^m | \{D, E\}\}.$$

If $k_g < m$, then

$$\begin{aligned} S(p^m, A, B, C, D, E) =& p^m p^{\frac{k_g}{2}} \left(p^m, \frac{\Delta_g}{p^{k_g}}\right)^{\frac{1}{2}} i^{\epsilon(p^{m-k_g})} i^{\epsilon\left(\frac{p^m}{(p^m, \Delta_g/p^{k_g})}\right)} e_{p^m}\left(\frac{g(E,-D)}{\Delta_g}\right) \\ & \cdot (-1)^{\upsilon(g)} \chi(p^m, A, B, C, D, E) \end{aligned} \tag{7.9}$$

where $\upsilon(g) = 0$ if g is non-degenerate at p^m and $\left(\begin{array}{c} -\frac{\Delta_g}{(\Delta_g, p^{2m-2})} \\ p \end{array}\right) = 1$, or g is degenerate and the quadratic form $g(x,y)/p^{k_g}$ can represent nonzero quadratic residue mod p; $\upsilon(g) = 1$ otherwise. The function $\chi(p^m; A, B, C, D, E) = 1$ if $S(p^m, A, B, C, D, E) \neq 0$, and $\chi(p^m; A, B, C, D, E) = 0$ if $S(p^m, A, B, C, D, E) \neq 0$.

Proof. It is a case-by-case proof, and the statement of Lemma 7.2 is a synthesis of all cases.

If $k_g = m$, the proof is trivial. We thus assume If $k_g < m$. Then after a linear unimodular change of variables, we can rewrite

$$S(p^m, A, B, C, D, E) = S(p^m, A', 0, C', D', E') \tag{7.10}$$

where $\deg_{p^m}(A') = k_g$ and $C' = \frac{-\Delta_g}{4A'}$. For instance, if $\deg_{p^m}(A) = \deg_{p^m}(\gcd(A,B,C))$, then we can let $x' = x + \frac{B}{2A}y$, $y' = y$, then $Ax^2 + Bxy + Cy^2 + Dx + Ey = Ax'^2 + (C - \frac{B^2}{4A})y'^2 + Dx' + (E - \frac{BD}{2A})y'$. If, instead, $\deg_{p^m}(B) < \deg_{p^m}(A), \deg_{p^m}(C)$, then we can apply the change $x' = x + y$, $y' = x - y$ to reduce to the previous case.

Now we can evaluate $S(p^m, A, B, C, D, E) = S(p^m, A', 0, C', D', E')$ from Lemma 7.1,

We have

$$\begin{aligned} & S(p^m, A, B, C, D, E) = S(p^m, A', 0, C', D', E') \\ =& p^m (p^m, A')^{\frac{1}{2}} (p^m, C')^{\frac{1}{2}} i^{\epsilon\left(\frac{p^m}{(p^m, A')}\right)} i^{\epsilon\left(\frac{p^m}{(p^m, C')}\right)} \\ & \cdot \left(\begin{array}{c} \frac{A'}{(A', p^{m-1})} \\ p \end{array}\right) \left(\begin{array}{c} \frac{C'}{(C', p^{m-1})} \\ p \end{array}\right) \mathbf{1}\left\{\begin{array}{l} \deg_{p^m}(D') \geqslant \deg_{p^m}(A') \\ \deg_{p^m}(E') \geqslant \deg_{p^m}(C') \end{array}\right\} e_{p^m}\left(\frac{g(E,-D)}{\Delta_g}\right), \end{aligned} \tag{7.11}$$

We interpret (7.11) in an intrinsic way. First, while all other factors are nonzero, the factor

$$\mathbf{1}\left\{\begin{array}{l} \deg_{p^m}(D') \geqslant \deg_{p^m}(A') \\ \deg_{p^m}(E') \geqslant \deg_{p^m}(C') \end{array}\right\} \tag{7.12}$$

is the same as the indicator function indicating whether S is zero or not. So we have (7.12) $= \chi(p^m, A, B, C, D, E)$.

For the term A', we know $\deg_{p^m}(A') = \deg_{p^m}(\gcd(A,B,C)) = k_g < m$, and that $\deg_{p^m}(C') = \deg_{p^m}(\Delta_g/A')$.

If $p^m \nmid C'$, then $\left(\begin{array}{c} \frac{A'}{(A', p^{m-1})} \\ p \end{array}\right) \left(\begin{array}{c} \frac{C'}{(C', p^{m-1})} \\ p \end{array}\right) = \left(\begin{array}{c} -\frac{\Delta_g}{4(\Delta_g, p^{2m-2})} \\ p \end{array}\right)$.

If $p^m \mid C'$, then $\left(\frac{\frac{C'}{(C',p^{m-1})}}{p} \right) = 1$, and

$$\left(\frac{\frac{A'}{(A',p^{m-1})}}{p} \right) = \begin{cases} 1 & \text{if nonzero quadratic residue } \frac{A'}{(A',p^{m-1})} \text{ is represented by} \\ & g(x,y)/p^{k_g} \text{in } \mathbb{Z}/p\mathbb{Z} \\ -1 & \text{otherwise.} \end{cases} \tag{7.13}$$

□

We note here that the function $\chi(p^m, A, B, C, D, E)$ concerns whether the p^m-degrees of the x, y coefficients are bigger than or equal to that of the x^2, y^2 coefficients after diagonalizing the quadratic part of $Ax^2 + Bxy + Cy^2 + Dx + Ey$. We list the following two noteworthy properties of χ:

(i) $\chi(p^m, A, B, C, D, E)$ is invariant under scaling of the quadratic part or the linear part, i.e. for any $(r, p) = 1$,

$$\chi(p^m, A, B, C, D, E) = \chi(p^m, rA, rB, rC, D, E) = \chi(p^m, A, B, C, rD, rE). \tag{7.14}$$

(ii) $\chi(p^m, A, B, C, D, E)$ is invariant under changing variables of x, y. If $x = x_1 + a, y = y_1 + b$, then

$$\begin{aligned} Ax^2 + Bxy + Cy^2 + Dx + Ey = Ax_1^2 + Bx_1y_1 + Cy_1^2 + (2Aa + Bb + D)x_1 \\ + (2Cb + Ba + E)y_1 + Aa^2 + Bab + Cb^2 + Da + Eb. \end{aligned} \tag{7.15}$$

Comparing the coefficients of the quadratic parts and linear parts of (7.15), we have

$$\chi(p^m, A, B, C, D, E) = \chi(p^m, A, B, C, 2Aa + Bb + D, 2Cb + Ba + E). \tag{7.16}$$

In Section 7.2 we will encounter the exponential sum

$$\mathcal{S}_\gamma(q, u, r, \xi, \zeta) = \frac{1}{q^2} \sum_{x_0, y_0(q)} e_q \left(r\mathfrak{f}_{M\gamma}(Lux_0 + uu^*, Luy_0) + x_0\xi + y_0\zeta \right). \tag{7.17}$$

Write

$$\mathfrak{f}_{M\gamma}(x, y) = \tilde{\mathfrak{f}}_{M\gamma}(x, y) + \mathfrak{d}_\gamma = A''x^2 + B''xy + C''y^2 + \mathfrak{d}_\gamma.$$

The quadratic form has discriminant $\Delta\mathfrak{d}_\gamma^2$. We assume that M is integral, so that $\mathfrak{f}_{M\gamma}$ is primitive and integral by Lemma 3.1. If M is not integral, then one needs to multiply the curvature formula by a universal constant, to obtain integrality. By Lemma 3.1, the gcd of the coefficients of $\mathfrak{f}_{M\gamma}$ after this normalization is bounded for all γ, and consequently all the estimates from Lemma 7.3, 7.4, 7.5 stand, up to a constant factor.

We first give a bound for $\mathcal{S}_\gamma(q, u, r, \xi, \zeta)$:

Lemma 7.3. *Assume that $(r, q) = 1$. Then*

$$|S_\gamma(q, u, r, \xi, \zeta)| \leqslant \frac{|\Delta|^{1/2} u^2 L^2(q, \mathfrak{d}_\gamma^2)}{q}.$$

Proof of Lemma 7.3 for q odd. For the proof when q is even, see the discussion at the end of this section.

First we consider the case $q = p^m$. If $p^m \mid u^2L^2$, then we trivially bound $|\mathcal{S}_\gamma| \leqslant 1$ and we automatically get the lemma. We thus assume $p^m \nmid u^2L^2$, then we apply the second case of Lemma 7.2 to analyze $\mathcal{S}_\gamma$.

We write

$$\begin{aligned}\mathcal{S}_\gamma(p^m,u,r,\xi,\zeta) =& \frac{1}{p^{2m}} \sum_{x_0,y_0 (p^m)} e_{p^m}\left(r\mathfrak{f}_{M\gamma}\left(Lux_0+uu^*,Luy_0\right)+x_0\xi+y_0\zeta\right)\\ =& \frac{1}{p^{2m}} \sum_{x_0,y_0 (p^m)} e_{p^m}(rL^2u^2(A''x_0^2+B''x_0y_0+C''y_0^2)\\ &+(2rA''Lu^2u^*+\xi)x_0+(rB''Lu^2u^*+\zeta)y_0+ru^2A''u^{*2}+r\mathfrak{d}_\gamma) \end{aligned} \tag{7.18}$$

Therefore, having the primitivity of $\mathfrak{f}_{M\gamma}$ in mind and applying Lemma 7.2 to (7.18), we obtain (here $p^{k_g}=(p^m,u^2L^2)$):

$$\begin{aligned}&\mathcal{S}_\gamma(p^m,u,r,\xi,\zeta)\\ =&\frac{1}{p^m}e_{p^m}(r\mathfrak{d}_\gamma+ru^2A''u^{*2})(p^m,u^2L^2)^{\frac{1}{2}}(p^m,u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}\\ &i^{\epsilon\left(\frac{p^m}{(p^m,u^2L^2)}\right)}i^{\epsilon\left(\frac{p^m}{(p^m,u^2L^2\mathfrak{d}_\gamma^2\Delta)}\right)}\cdot(-1)^{\upsilon(ru^2L^2\mathfrak{f}_{M\gamma})}\\ &\cdot e_{p^m}\left(\frac{A''(rB''Lu^2u^*+\zeta)^2-B''(2rA''Lu^2u^*+\xi)(rB''Lu^2u^*+\zeta)+C''(2rA''Lu^2u^*+\xi)^2}{rL^2u^2(B''^2-4A''C'')}\right)\\ &\cdot\chi\left(p^m,ru^2L^2A'',ru^2L^2B'',ru^2L^2C'',2rA''Lu^2u^*+\xi,rB''Lu^2u^*+\zeta\right)\\ =&\frac{1}{p^m}e_{p^m}\left(r\mathfrak{d}_\gamma-\frac{u^*\xi}{L}\right)(p^m,u^2L^2)^{\frac{1}{2}}(p^m,u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}\\ &i^{\epsilon\left(\frac{p^m}{(p^m,u^2L^2)}\right)}i^{\epsilon\left(\frac{p^m}{(p^m,u^2L^2\mathfrak{d}_\gamma^2\Delta)}\right)}\cdot(-1)^{\upsilon(ru^2L^2\mathfrak{f}_{M\gamma})}\cdot e_{p^m}\left(\frac{\tilde{\mathfrak{f}}_{M\gamma}(\zeta,-\xi)}{ru^2L^2\Delta\mathfrak{d}_\gamma^2}\right)\\ &\cdot\chi\left(p^m,ru^2L^2A'',ru^2L^2B'',ru^2L^2C'',2rA''Lu^2u^*+\xi,rB''Lu^2u^*+\zeta\right)\end{aligned} \tag{7.19}$$

From (7.19) we thus have

$$|S_\gamma(p^m,u,r,\xi,\zeta)|\leqslant\frac{1}{p^m}(p^m,u^2L^2)^{\frac{1}{2}}(p^m,u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}.$$

Using the multiplicativity of $\mathcal{S}_\gamma$, we obtain

$$\begin{aligned}|\mathcal{S}_\gamma(q,u,r,\xi,\zeta)|\leqslant&\prod_{p_i^{n_i}\|q}\frac{(p^{n_i},u^2L^2)(p^{n_i},u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}}{p^{n_i}}\\ \leqslant&\frac{(q,u^2L^2)^{\frac{1}{2}}(q,u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}}{q}\leqslant\frac{|\Delta|^{1/2}u^2L^2(q,\mathfrak{d}_\gamma^2)}{q}.\end{aligned} \tag{7.20}$$

□

We will also encounter a certain average of such sums. Let

$$\mathcal{S}(q,u,\gamma,\xi,\zeta,\gamma',\xi',\zeta')=\sum_{r(q)}{}'\mathcal{S}_\gamma(q,u,r,\xi,\zeta)\overline{\mathcal{S}_{\gamma'}(q,u,r,\xi',\zeta')} \tag{7.21}$$

Set $q=p^m$. From (7.19), we can write

$$(7.21)=S_1\cdot S_2, \tag{7.22}$$

with S_1 and S_2 as follows. The factor S_1 consists of factors not involving r:

$$S_1 = \frac{1}{p^{2m}}(p^m, u^2L^2)(p^m, u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}(p^m, u^2L^2\mathfrak{d}_{\gamma'}^2\Delta)^{\frac{1}{2}} i^{2\epsilon\left(\frac{p^m}{(p^m, u^2L^2)}\right)} \\ i^{\epsilon\left(\frac{p^m}{(p^m, u^2L^2\mathfrak{d}_\gamma^2\Delta)}\right)} i^{\epsilon\left(\frac{p^m}{(p^m, u^2L^2\mathfrak{d}_{\gamma'}^2\Delta)}\right)} e\left(\frac{u^*(\xi'-\xi)}{L}\right) \tag{7.23}$$

For S_2, we have

$$S_2 = \sideset{}{'}\sum_{r(p^m)} e_{p^m}(r(\mathfrak{d}_\gamma - \mathfrak{d}_{\gamma'}))(-1)^{\upsilon(ru^2L^2\mathfrak{f}_{M\gamma})}(-1)^{\upsilon(ru^2L^2\mathfrak{f}_{M\gamma'})} \cdot e_{p^m}\left(\frac{\tilde{\mathfrak{f}}_{M\gamma}(\zeta, -\xi)}{ru^2L^2\Delta\mathfrak{d}_\gamma^2}\right) \\ e_{p^m}\left(-\frac{\tilde{\mathfrak{f}}_{M\gamma'}(\zeta', -\xi')}{ru^2L^2\Delta\mathfrak{d}_{\gamma'}^2}\right) \cdot \chi(*) \tag{7.24}$$

We can bound S_1 directly from (7.23):

$$|S_1| \leqslant \frac{(p^m, u^2L^2)(p^m, u^2L^2\mathfrak{d}_\gamma^2\Delta)^{\frac{1}{2}}(p^m, u^2L^2\mathfrak{d}_{\gamma'}^2\Delta)^{\frac{1}{2}}}{p^{2m}}. \tag{7.25}$$

We note that $(-1)^{\upsilon(ru^2L^2\mathfrak{f}_{M\gamma})}, (-1)^{\upsilon(ru^2L^2\mathfrak{f}_{M\gamma'})}$ are multiplicative over r. Moreover, from (7.14) and (7.16), we observe that with all other parameters fixed, χ is a periodic function over r with period dividing (L, p^m). Therefore the function $\chi(*)$ can be viewed as a function on $(\mathbb{Z}/(p^m, L)\mathbb{Z})^*$ bounded by 1, so can be written as at most (p^m, L) linearly combined multiplicative characters on $\mathbb{Z}/p^m\mathbb{Z}$ with coefficients bounded by 1. Therefore, the factor S_2 is a combination of at most (p^m, L) Kloosterman-Salié sums.

If $\mathfrak{d}_\gamma \neq \mathfrak{d}_{\gamma'}$, applying Kloosterman's elementary 3/4 bound for this type of sum (Lemma 3.4.1, [36]), we obtain

$$|S_2| \ll (p^m, L)p^{\frac{3}{4}m+\epsilon}(p^m, \mathfrak{d}_\gamma - \mathfrak{d}_{\gamma'})^{\frac{1}{4}}. \tag{7.26}$$

If $\mathfrak{d}_\gamma = \mathfrak{d}_{\gamma'}$ but $\mathfrak{f}_{M\gamma}(\zeta, -\xi) \neq \mathfrak{f}_{M\gamma'}(\zeta', -\xi')$, then we can use the last two factors in the summand of (7.24) to obtain a bound for S_2. It can be checked that if $S_2 \neq 0$, then the condition that $\chi(r; *) = 1$ in (7.24) leads to

$$\deg_{p^m}\left(\frac{\mathfrak{f}_{M\gamma}(\zeta, -\xi)}{ru^2L^2\Delta\mathfrak{d}_\gamma^2}\right), \deg_{p^m}\left(\frac{\mathfrak{f}_{M\gamma'}(\zeta', -\xi')}{ru^2L^2\Delta\mathfrak{d}_{\gamma'}^2}\right) \geqslant 0.$$

Therefore, the elementary Kloosterman 3/4 bound in this case gives

$$|S_2| \ll (p^m, L)p^{\frac{3}{4}m+\epsilon}(p^m, \mathfrak{f}_{M\gamma}(\zeta, -\xi) - \mathfrak{f}_{M\gamma'}(\zeta', -\xi'))^{\frac{1}{4}}. \tag{7.27}$$

Collecting (7.25), (7.26), (7.27), using the multiplicativity of $\mathcal{S}(q, u, r, \gamma, \xi, \zeta, \gamma', \xi', \zeta')$, and absorbing Δ, L in the $\ll$ relation, we obtain the following two lemmas in the case q is odd.

LEMMA 7.4. *If $\mathfrak{d}_\gamma \neq \mathfrak{d}_{\gamma'}$, then*

$$|\mathcal{S}(q, u, r, \gamma, \xi, \zeta, \gamma', \xi', \zeta')| \ll u^4q^{-\frac{5}{4}+\epsilon}(q, \mathfrak{d}_\gamma - \mathfrak{d}_{\gamma'})^{\frac{1}{4}}(q, \mathfrak{d}_\gamma^2)^{\frac{1}{2}}(q, \mathfrak{d}_{\gamma'}^2)^{\frac{1}{2}}.$$

LEMMA 7.5. *If $\mathfrak{d}_\gamma = \mathfrak{d}_{\gamma'}$ and $\mathfrak{f}_{M\gamma}(\zeta, -\xi) \neq \mathfrak{f}_{M\gamma'}(\zeta', -\xi')$, then*

$$|\mathcal{S}(q, u, r, \gamma, \xi, \zeta, \gamma', \xi', \zeta')| \ll u^4q^{-\frac{5}{4}+\epsilon}|\mathfrak{f}_{M\gamma}(\zeta, -\xi) - \mathfrak{f}_{M\gamma'}(\zeta', -\xi')|^{\frac{1}{4}}(q, \mathfrak{d}_\gamma^2).$$

We briefly explain how to extend Lemmas 7.1 through 7.5 when q is even. It is enough to consider $q = 2^m$ by multiplicativity. The extra complication arises in Lemma 7.1 when we complete squares for some exponential sums (e.g., $\sum_{x=0}^{7} e_8(x^2+x)$): we encounter certain "restricted" Gauss sums, meaning the sum index is restricted to certain congruence classes mod 2. This slightly alters the statement of Lemma 7.1 for $q = 2^m$. We can handle this by writing an indicator function of the allowed congruence classes. In Lemmas 7.3, 7.4 and 7.5, we can handle the extra indicator function by writing it as a linear combination of two additive characters to the modulus 2. We obtain a linear combination of more Kloosterman-Salié sums in Lemmas 7.4 and 7.5, and this eventually gives an extra constant factor to the bound on $|S_2|$. The rest of the proof is the same.

7.2 Minor arc analysis, part I

We begin by estimating $\mathcal{I}_1$. First we take the Fourier transform of $\mathcal{R}_N^U$ (defined at (4.8)):

$$\widehat{\mathcal{R}}_N^U(\theta) = \sum_{\substack{u<U \\ (u,L)=1}} \mu(u) \sum_{\gamma \in \mathfrak{F}_T} \mathcal{R}_{u,\gamma}(\theta), \tag{7.28}$$

where

$$\mathcal{R}_{u,\gamma}(\theta) = \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right) e(\mathfrak{f}_{M\gamma}(Lux+uu^*, Luy)\theta). \tag{7.29}$$

We will first give an L^∞ bound for $\mathcal{R}_{u,\gamma}$ (see (7.36)).

Write $\theta = \frac{r}{q} + \beta$ and rearrange the order of x, y according to the congruence classes mod q:

$$\begin{aligned}\mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right) = \sum_{x_0,y_0(q)} e\left(\mathfrak{f}_{M\gamma}(Lux_0+uu^*, Luy_0)\frac{r}{q}\right) \\ \cdot \left[\sum_{\substack{x\equiv x_0(q)\\ y\equiv y_0(q)}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right) e\left(\mathfrak{f}_{M\gamma}\left(Lux+uu^*, Luy\right)\beta\right)\right]\end{aligned} \tag{7.30}$$

Applying Poisson summation to the x, y sum in the bracket $[\cdot]$, we obtain

$$\begin{aligned}[\cdot] =& \sum_{\xi,\zeta\in\mathbb{Z}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi\left(\frac{Lu(x_0+qx)+uu^*}{X}\right)\psi\left(\frac{Lu(y_0+qy)}{X}\right) \\ & \cdot e\left(\mathfrak{f}_{M\gamma}\left(Lu(x_0+qx)+uu^*, Lu(y_0+qy)\right)\beta\right) e(-x\xi-y\zeta)dxdy \\ =& \frac{X^2}{q^2L^2u^2} \sum_{\xi,\zeta\in\mathbb{Z}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y) e\left(\mathfrak{f}_{M\gamma}\left(Xx, Xy\right)\beta - \frac{X\xi}{quL}x - \frac{X\zeta}{quL}y\right) e\left(\frac{u^*\xi}{Lq}\right) dxdy \cdot \\ & \cdot e_q(x_0\xi + y_0\zeta)\end{aligned} \tag{7.31}$$

Plugging (7.31) back into (7.30), we have

$$\mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right) = \frac{X^2}{L^2u^2} \sum_{\xi,\zeta\in\mathbb{Z}} \mathcal{S}_\gamma(q,u,r,\xi,\zeta)\mathcal{J}_\gamma(\beta;q,u,\xi,\zeta), \tag{7.32}$$

where

$$\mathcal{S}_\gamma(q,u,r,\xi,\zeta)=\frac{1}{q^2}\sum_{x_0,y_0(q)} e_q\left(r\mathfrak{f}_{M\gamma}\left(Lux+uu^*,Luy_0\right)+x_0\xi+y_0\zeta\right) \tag{7.33}$$

and

$$\mathcal{J}_\gamma(\beta;q,u,\xi,\zeta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)e\left(\mathfrak{f}_{M\gamma}\left(Xx,Xy\right)\beta-\frac{X\xi}{quL}x-\frac{X\zeta}{quL}y\right)e\left(\frac{u^*\xi}{Lq}\right)dxdy \tag{7.34}$$

Note that the sum in (7.32) is principally supported on a few terms, since the $\mathcal{J}_\gamma$ term decays quickly. We will use non-stationary and stationary phase methods to give bounds for the $\mathcal{J}_\gamma$ terms. We review the statements here, for reference.

Proposition 7.6 [37, Page 24, Non-stationary phase]. *Let ϕ be a smooth compactly supported function on $(-\infty,\infty)$ and f be a function which, for x in the support of ϕ, satisfies*

(i) $|f'(x)|>A>0$,
(ii) $A\geqslant|f^{(2)}(x)|,\dots,|f^{(n)}(x)|$.

Then

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

Proposition 7.7 [37, Page 25, Stationary phase]. *Let f be a quadratic polynomial of two variables x and y whose homogeneous part has discriminant $-D$ with $D>0$. Let $\phi(x,y)$ be a smooth compactly supported function on $\mathbb{R}^2$, then*

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

We apply the non-stationary phase to $\mathcal{J}_\gamma$. We can obtain a bound A as required in the statement by taking

$$\frac{X\xi}{qu}\text{ or }\frac{X\zeta}{qu}\gg T^2X^2|\beta|.$$

(Note that the discriminant of $\mathfrak{f}_{M\gamma}$ is bounded above by T^4.) Using the former for example, the value of A is then XU/quL, which is >1 since $u<U$, $q<Q_0$ and by (4.1) and (5.1).

Therefore, the main contribution of the ξ,ζ sum in (7.32) comes from the ξ,ζ terms such that $\frac{X\xi}{qu}\ll T^2X^2|\beta|$ and $\frac{X\zeta}{qu}\ll T^2X^2|\beta|$, or in other words the terms ξ,ζ such that

$$\xi,\zeta\ll quT^2X|\beta|\ll uT^2X/J=u<U,$$

where we used $|\beta|\leqslant\frac{1}{qJ}$ and $J=T^2X$ (by (4.11)).

For the terms $\xi,\zeta\ll u$, we have an upper bound for $\mathcal{J}_\gamma$ using the stationary phase:

$$|\mathcal{J}_\gamma(\beta;q,u,\xi,\zeta)|\ll\min\left\{1,\frac{1}{T^2X^2|\beta|}\right\} \tag{7.35}$$

Lemma 7.3 and (7.35) together lead to a bound for $\mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right)$ and hence for $\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)$:

$$\left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|\ll\frac{T^{2\delta-2}U}{|\beta|}. \tag{7.36}$$

Now we are ready to give an estimate for $\mathcal{I}_1$. We rewrite $\mathcal{I}_1$ as

$$\mathcal{I}_1 = \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \int_{-\frac{1}{qJ}}^{\frac{1}{qJ}} \left|\left(1-\mathfrak{T}\left(\frac{r}{q}+\beta\right)\right)\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 d\beta. \tag{7.37}$$

We now split the integral $\int_{-\frac{1}{qJ}}^{\frac{1}{qJ}}$ above into three parts $\int_{-\frac{K_0}{N}}^{\frac{K_0}{N}}$, $\int_{\frac{K_0}{N}}^{\frac{1}{qJ}}$ and $\int_{-\frac{1}{qJ}}^{-\frac{K_0}{N}}$. For each integral we use (7.36) to bound the $\widehat{\mathcal{R}}_N^U$ term. In the first integral, we use

$$\left|\left(1-\mathfrak{T}\left(\frac{r}{q}+\beta\right)\right)\right|^2 = \frac{N^2\beta^2}{K_0^2},$$

and in the second and third integral, we trivially bound $\left|\left(1-\mathfrak{T}\left(\frac{r}{q}+\beta\right)\right)\right|^2$ above by 1. Altogether, we have

Lemma 7.8.

$$\mathcal{I}_1 \ll \frac{U^2T^{4\delta-4}NQ_0^2}{K_0} \tag{7.38}$$

Since $K_0 \gg Q_0^3 \gg Q_0^2U^2N^\epsilon$ (see (7.53)), we have $\mathcal{I}_1 \ll T^{4\delta-4}N^{1-\epsilon}$.

7.3 Minor arc analysis, part II

In this section we give an upper bound for

$$\mathcal{I}_Q = \sum_{Q<q\leqslant 2Q} \int_{-\frac{1}{qJ}}^{\frac{1}{qJ}} \sideset{}{'}\sum_{r(q)} \left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 d\beta \tag{7.39}$$

for $Q_0 < Q < X$ and show the following.

Lemma 7.9.

$$\mathcal{I}_2 \ll T^{4\delta-4}N^{1-\eta}$$

Proof. Going back to (7.28), we apply Cauchy-Schwarz to the u sum to get an upper bound for $\widehat{\mathcal{R}}_N^U(\frac{r}{q}+\beta)$:

$$\left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 \ll U\sum_{u<U}\sum_{\gamma\in\mathfrak{F}_T}\sum_{\gamma'\in\mathfrak{F}_T}\mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right)\overline{\mathcal{R}_{u,\gamma'}\left(\frac{r}{q}+\beta\right)} \tag{7.40}$$

Using (7.32), we obtain

$$\begin{aligned}\sum_{r(q)}{}' \left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 \ll & U\sum_{u<U}\frac{X^4}{u^4}\sum_{\xi,\zeta\in\mathbb{Z}}\sum_{\xi',\zeta'\in\mathbb{Z}}\sum_{\gamma\in\mathfrak{F}_T}\sum_{\gamma'\in\mathfrak{F}_T}\mathcal{S}(q,u,\gamma,\xi,\zeta,\gamma',\xi',\zeta') \\ & \cdot \mathcal{J}_\gamma(\beta;q,u,\xi,\zeta)\overline{\mathcal{J}_{\gamma'}(\beta;q,u,\xi',\zeta')}\end{aligned} \tag{7.41}$$

By the non-stationary phase, the main contribution to (7.41) comes from the terms $\xi,\zeta,\xi',\zeta' \ll U$, and for these terms, we have

$$\mathcal{J}_\gamma(\beta;q,u,\xi,\zeta)\overline{\mathcal{J}_{\gamma'}(\beta;q,u,\xi',\zeta')} \ll \min\left\{1, \frac{1}{T^4X^4\beta^2}\right\} \tag{7.42}$$

Using Lemma 7.4 together with (7.42), we obtain

$$\sum_{r(q)}{}' \left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 \ll \sum_{\gamma\in\mathfrak{F}_T}\sum_{\gamma'\in\mathfrak{F}_T} U^6X^4q^{-\frac{5}{4}+\epsilon}(q,\mathfrak{d}_\gamma-\mathfrak{d}_{\gamma'})^{\frac{1}{4}}(q,\mathfrak{d}_\gamma^2)^{\frac{1}{2}}(q,\mathfrak{d}_{\gamma'}^2)^{\frac{1}{2}}\min\left\{1,\frac{1}{T^4X^4\beta^2}\right\} \tag{7.43}$$

Observe that $qJ\leqslant T^2X^2$, so that

$$\int_{-\frac{1}{qJ}}^{\frac{1}{qJ}}\min\left\{1,\frac{1}{T^4X^4\beta^2}\right\}d\beta \ll \frac{1}{T^2X^2} \tag{7.44}$$

Plug (7.43) and (7.44) into (7.39), and we obtain

$$\mathcal{I}_Q \ll \frac{N^\epsilon U^6X^2}{T^2Q^{\frac{5}{4}}}\sum_{Q<q\leqslant 2Q}\sum_{\gamma\in\mathfrak{F}_T}\sum_{\gamma'\in\mathfrak{F}_T}(q,\mathfrak{d}_\gamma-\mathfrak{d}_{\gamma'})^{\frac{1}{4}}(q,\mathfrak{d}_\gamma^2)^{1/2}(q,\mathfrak{d}_{\gamma'}^2)^{1/2} \tag{7.45}$$

We split (7.45) into two parts $\mathcal{I}_Q^{(=)}$ and $\mathcal{I}_Q^{(\neq)}$ according to whether $\mathfrak{d}_\gamma=\mathfrak{d}_{\gamma'}$ or not. We first estimate $\mathcal{I}_Q^{(=)}$:

$$\begin{aligned}
\mathcal{I}_Q^{(=)} \ll& \frac{N^\epsilon U^6X^2}{QT^2}\sum_{Q<q\leqslant 2Q}\sum_{\gamma\in\mathfrak{F}_T}\sum_{\substack{\gamma'\in\mathfrak{F}_T\\ \mathfrak{d}_{\gamma'}=\mathfrak{d}_\gamma}}(q,\mathfrak{d}_\gamma^2)\\
\ll& \frac{N^\epsilon U^6X^2}{QT^2}\sum_{\gamma\in\mathfrak{F}_T}\sum_{\substack{\gamma'\in\mathfrak{F}_T\\ \mathfrak{d}_{\gamma'}=\mathfrak{d}_\gamma}}\sum_{d|\mathfrak{d}_\gamma^2}d\sum_{Q<q\leqslant 2Q}\mathbf{1}\{d|q\}\\
\ll& \frac{N^\epsilon U^6X^2}{QT^2}\sum_{\gamma\in\mathfrak{F}_T}\sum_{\substack{\gamma'\in\mathfrak{F}_T\\ \mathfrak{d}_{\gamma'}=\mathfrak{d}_\gamma}}\sum_{d|\mathfrak{d}_\gamma^2}Q\\
\ll& \frac{N^\epsilon U^6X^2T^{2\delta}}{T^2}\sum_{\substack{\gamma'\in\mathfrak{F}_T\\ \mathfrak{d}_{\gamma'}=\mathfrak{d}_\gamma}}1\\
\ll& \frac{N^{1+\epsilon}U^6T^{4\delta-4}}{T^{\eta_0}},
\end{aligned} \tag{7.46}$$

where we have bounded the number of divisors of $\mathfrak{d}_\gamma^2$ by N^ϵ and in the last step we used Lemma 5.1 to estimate the sum (using modulus T in the statement of the Lemma), with reference to (4.1).

Now we estimate $\mathcal{I}_Q^{(\neq)}$. We introduce a new parameter H and we further split $\mathcal{I}_Q^{(\neq)}$ into two $\mathcal{I}_Q^{(\neq,\geqslant)}$ and $\mathcal{I}_Q^{(\neq,<)}$ according to $(\mathfrak{d}_\gamma,\mathfrak{d}_{\gamma'})\geqslant H$ or not.

We first estimate $\mathcal{I}_Q^{(\neq,\geqslant)}$. Recall (7.43). Then we have

$$\mathcal{I}_Q^{(\neq,\geqslant)} \ll \frac{N^\epsilon U^6 X^2}{QT^2} \sum_{Q<q\leqslant 2Q} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ (\mathfrak{d}_{\gamma'},\mathfrak{d}_\gamma)\geqslant H}} (q,\mathfrak{d}_\gamma^2)^{\frac{1}{2}}(q,\mathfrak{d}_{\gamma'}^2)^{\frac{1}{2}}$$

$$\ll \frac{N^\epsilon U^6 X^2}{QT^2} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{h|\mathfrak{d}_\gamma \\ h\geqslant H}} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}_{\gamma'}\equiv\mathfrak{d}_\gamma(h)}} \sum_{q_1|\mathfrak{d}_\gamma^2} \sum_{q_2|\mathfrak{d}_{\gamma'}^2} (q_1q_2)^{\frac{1}{2}} \sum_{Q<q\leqslant 2Q} \mathbf{1}\{[q_1,q_2]|q\}. \tag{7.47}$$

Notice that $[q_1,q_2]\geqslant (q_1q_2)^{\frac{1}{2}}$. Therefore,

$$(7.47) \ll \frac{N^\epsilon U^6 X^2}{T^2} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{h|\mathfrak{d}_\gamma \\ h\geqslant H}} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}_{\gamma'}\equiv\mathfrak{d}_\gamma(h)}} 1 \tag{7.48}$$

Again using Lemma 5.1, we have

$$(7.48) \ll \frac{N^\epsilon U^6 X^2}{T^2 H^{\eta_0}} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{h|\mathfrak{d}_\gamma \\ h\geqslant H}} 1 \ll \frac{N^{1+\epsilon} U^6 T^{4\delta-4}}{H^{\eta_0}} \tag{7.49}$$

Now we estimate $\mathcal{I}_Q^{(\neq,<)}$. Using (7.43) and replacing $(q,\mathfrak{d}_\gamma^2)^{\frac{1}{2}}, (q,\mathfrak{d}_{\gamma'}^2)^{\frac{1}{2}}$ by $(q,\mathfrak{d}_\gamma), (q,\mathfrak{d}_{\gamma'})$, we have

$$\mathcal{I}_Q^{(\neq,<)} \ll \frac{N^\epsilon U^6 X^2}{T^2 Q^{\frac{5}{4}}} \sum_{Q<q\leqslant 2Q} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ (\mathfrak{d}_{\gamma'},\mathfrak{d}_\gamma)<H}} (q,\mathfrak{d}_\gamma-\mathfrak{d}_{\gamma'})^{\frac{1}{4}}(q,\mathfrak{d}_\gamma)(q,\mathfrak{d}_{\gamma'})$$

$$\ll \frac{N^\epsilon U^6 X^2}{T^2 Q^{\frac{5}{4}}} \sum_{\gamma\in\mathfrak{F}_T} \sum_{d_1|\mathfrak{d}_\gamma} \sum_{d_3\leqslant 2Q} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}'_\gamma\equiv\mathfrak{d}_\gamma(d_3)}} \sum_{d_2|\mathfrak{d}_{\gamma'}} d_1d_2d_3^{\frac{1}{4}} \sum_{Q<q\leqslant 2Q} \mathbf{1}\{[d_1,d_2,d_3]|q\} \tag{7.50}$$

Writing $h=(\mathfrak{d}_\gamma,\mathfrak{d}_{\gamma'})$, then $\frac{d_1}{h}, \frac{d_2}{h}, \frac{d_3}{h}$ are mutually relatively prime. Since $h<H$, we have the estimate

$$\sum_{Q<q\leqslant 2Q} \mathbf{1}\{[d_1,d_2,d_3]|q\} \leqslant \frac{QH^2}{d_1d_2d_3} \tag{7.51}$$

Therefore,

$$(7.50) \ll \frac{N^\epsilon U^6 X^2 H^2}{T^2 Q^{\frac{1}{4}}} \sum_{\gamma\in\mathfrak{F}_T} \sum_{d_1|\mathfrak{d}_\gamma} \sum_{d_3\leqslant 2Q} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}'_\gamma\equiv\mathfrak{d}_\gamma(d_3)}} \sum_{d_2|\mathfrak{d}_{\gamma'}} d_3^{-\frac{3}{4}}$$

$$\ll \frac{N^\epsilon U^6 X^2 H^2 T^{4\delta}}{T^2 Q^{\frac{1}{4}}} \sum_{d_3\leqslant 2Q} d_3^{-\frac{3}{4}-\eta_0}$$

$$\ll \frac{N^{1+\epsilon} U^6 H^2 T^{4\delta-4}}{Q^{\eta_0}} \tag{7.52}$$

To make the terms at (7.46), (7.49), (7.52), and later on at (7.69) $\ll T^{4\delta-4}N^{1-\eta}$ for an appropriate positive η, we can set

$$\boxed{H = Q_0^{\frac{\eta_0}{4}}, U = H^{\frac{\eta_0}{20}}} \tag{7.53}$$

Thus we have proven Lemma 7.9. □

7.4 Minor arc analysis, part III

In this section we give an upper bound for $\mathcal{I}_Q$ when Q is large, i.e. $X < Q \leqslant J$. We keep all notation from the previous sections. Namely, we show the following.

Lemma 7.10.

$$\mathcal{I}_3 \ll T^{4\delta-4}N^{1-\eta}.$$

Proof. Recall

$$\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right) = \sum_{\substack{u<U\\(u,L)=1}} \mu(u) \sum_{\gamma\in\mathfrak{F}_T} \mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right), \tag{7.54}$$

where

$$\mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right) = \sum_{x,y,\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right) e\left(\mathfrak{f}_{M\gamma}\left(Lux+uu^*,uLy\right)\left(\frac{r}{q}+\beta\right)\right). \tag{7.55}$$

We insert extra harmonics by writing $e_q\left(r\mathfrak{f}_{M\gamma}\left(Lux+uu^*,uLy\right)\right)$ into its Fourier expansion:

$$\begin{aligned}
&e_q\left(r\mathfrak{f}_{M\gamma}\left(Lux+uu^*,uLy\right)\right)\\
=&\frac{1}{q^2}\sum_{m(q)}\sum_{n(q)}\sum_{s(q)}\sum_{t(q)} e_q\left(r\mathfrak{f}_{M\gamma}\left(Lus+uu^*,Lut\right)+sm+tn\right) e\left(-\frac{mx}{q}-\frac{ny}{q}\right)\\
=&\sum_{m(q)}\sum_{n(q)} \mathcal{S}_\gamma(q,u,r,m,n) e\left(-\frac{mx}{q}-\frac{ny}{q}\right)
\end{aligned} \tag{7.56}$$

Inserting (7.56) into (7.55), we obtain

$$\mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right) = \sum_{m(q)}\sum_{n(q)} \mathcal{S}_\gamma(q,u,r,m,n)\lambda_\gamma\left(\beta,X,u,\frac{m}{q},\frac{n}{q}\right) \tag{7.57}$$

where

$$\lambda_\gamma(\beta,X,u,s,t) = \sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right)\psi\left(\frac{Luy}{X}\right) e\left(\mathfrak{f}_{M\gamma}\left(Lux+uu^*,Luy\right)\beta - sx - ty\right) \tag{7.58}$$

Now we apply Cauchy-Schwarz to the u sum for $\widehat{\mathcal{R}}_N^U$ (see (7.40), (7.54)), and insert it back

into $\mathcal{I}_Q$ at (7.39). We have

$$\begin{aligned}
\mathcal{I}_Q &= \sum_{Q<q\leqslant 2Q} {\sum_{r(q)}}' \int_{-\frac{1}{qJ}}^{\frac{1}{qJ}} \left|\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right|^2 d\beta \\
&\ll U \sum_{u<U} \sum_{Q<q\leqslant 2Q} {\sum_{r(q)}}' \int_{-\frac{1}{qJ}}^{\frac{1}{qJ}} \left|\sum_{\gamma\in\mathfrak{F}_T} \mathcal{R}_{u,\gamma}\left(\frac{r}{q}+\beta\right)\right|^2 d\beta \\
&\ll U \sum_{u<U} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\gamma'\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \sum_{m,n,m',n'(q)} \left({\sum_{r(q)}}' \mathcal{S}_\gamma(q,u,r,m,n)\overline{\mathcal{S}_{\gamma'}(q,u,r,m',n')}\right) \\
&\quad \cdot \int_{-\frac{1}{qJ}}^{\frac{1}{qJ}} \lambda_\gamma\left(\beta,X,u,\frac{m}{q},\frac{n}{q}\right)\overline{\lambda_{\gamma'}\left(\beta,X,u,\frac{m'}{q},\frac{n'}{q}\right)} d\beta \\
&= U \sum_{u<U} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\gamma'\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \sum_{m,n,m',n'(q)} \mathcal{S}(q,u,\gamma,m,n,\gamma',m',n') \\
&\quad \cdot \int_{-\frac{1}{qJ}}^{\frac{1}{qJ}} \lambda_\gamma\left(\beta,X,u,\frac{m}{q},\frac{n}{q}\right)\overline{\lambda_{\gamma'}\left(\beta,X,u,\frac{m'}{q},\frac{n'}{q}\right)} d\beta
\end{aligned} \tag{7.59}$$

Applying Poisson summation and non-stationary phase to λ_γ and $\lambda_{\gamma'}$, we see that the main contribution to (7.59) comes from the terms m,n,m',$n' \ll \frac{qu}{X}$. For these terms, we use the trivial bound:

$$\left|\lambda_\gamma\left(\beta,X,u,\frac{m}{q},\frac{n}{q}\right)\right|, \left|\lambda_{\gamma'}\left(\beta,X,u,\frac{m'}{q},\frac{n'}{q}\right)\right| \ll \frac{X^2}{u^2} \tag{7.60}$$

From (7.59) and (7.60) we have

$$\mathcal{I}_Q \ll \frac{UX^4}{QJ} \sum_{u<U} \frac{1}{u^4} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\gamma'\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \sum_{m,n,m',n'\ll\frac{uq}{X}} |\mathcal{S}(q,u,\gamma,m,n,\gamma',m',n')| \tag{7.61}$$

We split (7.61) into $\mathcal{I}_Q^{(=)}$ and $\mathcal{I}_Q^{(\neq)}$ according to whether $\mathfrak{d}_\gamma = \mathfrak{d}_{\gamma'}$ or not:

$$\mathcal{I}_Q^{(=)} = \frac{UX^4}{QJ} \sum_{u<U} \frac{1}{u^4} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}_{\gamma'}=\mathfrak{d}_\gamma}} \sum_{Q<q\leqslant 2Q} \sum_{m,n,m',n'\ll\frac{uq}{X}} |\mathcal{S}(q,u,\gamma,m,n,\gamma',m',n')| \tag{7.62}$$

and

$$\mathcal{I}_Q^{(\neq)} = \frac{UX^4}{QJ} \sum_{u<U} \frac{1}{u^4} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}_{\gamma'}\neq\mathfrak{d}_\gamma}} \sum_{Q<q\leqslant 2Q} \sum_{m,n,m',n'\ll\frac{uq}{X}} |\mathcal{S}(q,u,\gamma,m,n,\gamma',m',n')| \tag{7.63}$$

We first deal with $\mathcal{I}_Q^{(\neq)}$. Using Lemma 7.4 to bound $|\mathcal{S}|$, we obtain:

$$\mathcal{I}_Q^{(\neq)} \ll \frac{UX^4}{QJ} \sum_{u<U} \frac{1}{u^4} \sum_{\gamma\in\mathfrak{F}_T} \sum_{\substack{\gamma'\in\mathfrak{F}_T \\ \mathfrak{d}_{\gamma'}\neq\mathfrak{d}_\gamma}} \sum_{Q<q\leqslant 2Q} \sum_{m,n,m',n'\ll\frac{uq}{X}} u^4 q^{-\frac{5}{4}+\epsilon} (q,\mathfrak{d}_\gamma-\mathfrak{d}_{\gamma'})^{\frac{1}{4}} (q,\mathfrak{d}_\gamma^2)^{\frac{1}{2}} (q,\mathfrak{d}_{\gamma'}^2)^{\frac{1}{2}} \tag{7.64}$$

Bounding $(q, \mathfrak{d}_\gamma - \mathfrak{d}_{\gamma'})$, $(q, \mathfrak{d}_\gamma^2)$ $(q, \mathfrak{d}_{\gamma'}^2)$ by T, T^2, T^2 respectively, we obtain

$$\mathcal{I}_Q^{(\neq)} \ll U^6 T^{4\delta+\frac{1}{4}} X^{-1} Q^{\frac{11}{4}+\epsilon} \ll T^{4\delta-4} N^{1+\epsilon} \left(U^6 T^{\frac{9}{4}} X^{-3} Q^{\frac{11}{4}} \right). \tag{7.65}$$

Since $Q \leqslant J = T^2 X$, the term in the parentheses above is $\ll U^2 T^{\frac{17}{4}} X^{-\frac{1}{4}}$, and thus we have obtained a significant power saving for $\mathcal{I}_Q^{(\neq)}$.

Now we deal with $\mathcal{I}_Q^{(=)}$. We split $\mathcal{I}_Q^{(=)}$ into $\mathcal{I}_Q^{(=,=)}$ and $\mathcal{I}_Q^{(=,\neq)}$ according to whether $\mathfrak{f}_{M\gamma}(n,-m) = \mathfrak{f}_{M\gamma'}(n',-m')$ or not. We first give an upper bound for $\mathcal{I}_Q^{(=,\neq)}$. From Lemma 7.5,

$$\begin{aligned}\mathcal{I}_Q^{(=,\neq)} \ll & \frac{UX^4}{QJ} \sum_{u<U} \frac{1}{u^4} \sum_{m,n,m',n' \ll \frac{uq}{X}} \sum_{\gamma \in \mathfrak{F}_T} \sum_{\substack{\gamma' \in \mathfrak{F}_T \\ \mathfrak{d}_{\gamma'} = \mathfrak{d}_\gamma \\ \mathfrak{f}_{M\gamma}(\xi,-\zeta) \neq \mathfrak{f}_{M\gamma'}(\xi',-\zeta')}} \sum_{Q<q\leqslant 2Q} u^4 q^{-\frac{5}{4}} (q, \mathfrak{d}_\gamma^2)^{\frac{1}{2}} (q, \mathfrak{d}_{\gamma'}^2)^{\frac{1}{2}} \\ & \cdot |\mathfrak{f}_{M\gamma}(n,-m) - \mathfrak{f}_{M\gamma'}(n',-m')|^{\frac{1}{4}}. \end{aligned} \tag{7.66}$$

We bound $(q, \mathfrak{d}_\gamma^2), (q, \mathfrak{d}_{\gamma'}^2)$ by T^2 and $|\mathfrak{f}_{M\gamma}(n,-m) - \mathfrak{f}_{M\gamma'}(n',-m')|$ by $\frac{T^2u^2q^2}{X^2}$, so that we have

$$\mathcal{I}_Q^{(=,\neq)} \ll U^{\frac{13}{2}} Q^{\frac{13}{4}+\epsilon} T^{4\delta+\frac{1}{2}} X^{-\frac{3}{2}} \ll T^{4\delta-4} N^{1+\epsilon} (U^{\frac{13}{2}} T^9 X^{-\frac{1}{4}}), \tag{7.67}$$

where we have used $Q \leqslant T^2 X$. Again we obtain a significant power saving from (7.67).

Finally we estimate $\mathcal{I}_Q^{(=,=)}$. From Lemma 7.4 and (7.62), we have

$$\mathcal{I}_Q^{(=,=)} \ll \frac{N^\epsilon UX^4}{QJ} \sum_{u<U} \frac{1}{u^4} \sum_{\gamma \in \mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \sum_{m,n \ll \frac{uq}{X}} \frac{(\mathfrak{d}_\gamma^2, q)}{q} u^4 \sum_{\substack{\gamma' \in \mathfrak{F}_T \\ \mathfrak{d}_{\gamma'} = \mathfrak{d}_\gamma}} \sum_{\substack{m',n' \ll \frac{uq}{X} \\ \mathfrak{f}_{M\gamma'}(n',-m') = \mathfrak{f}_{M\gamma}(n,-m)}} 1 \tag{7.68}$$

To analyze (7.68), we introduce the following two lemmata, the proofs of which are minor adaptions of the proofs of Lemma 3.15 and Lemma 3.16 from [37].

Lemma 7.11. *Fix $\gamma \in \mathfrak{F}_T$. Then we have*

$$\sum_{\substack{\gamma' \in \mathfrak{F}_T \\ \mathfrak{d}_{\gamma'} = \mathfrak{d}_\gamma}} \sum_{\substack{m',n' \ll \frac{uq}{X} \\ \mathfrak{f}_{M\gamma'}(n',-m') = \mathfrak{f}_{M\gamma}(n,-m)}} 1 \ll N^\epsilon \left(\tilde{\mathfrak{f}}_{M\gamma}(n,-m), \mathfrak{d}_\gamma^2 \right)^{\frac{1}{2}}$$

Lemma 7.12. *For any $\gamma \in \mathfrak{F}_T$, $d | \mathfrak{d}_\gamma^2$ and any integer $W > 0$, we have*

$$\sum_{\substack{m,n \leqslant W \\ \tilde{\mathfrak{f}}_{M\gamma}(n,-m) \equiv 0 (d)}} 1 \ll W^2 d^{-\frac{1}{2}} + W.$$

We now return to (7.68). From Lemma 7.11 and Lemma 7.12, we have

$$\begin{aligned}
\mathcal{I}_Q^{(=,=)} \ll & \frac{N^\epsilon UX^4}{QJ} \sum_{u<U} \sum_{\gamma\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \sum_{m,n\ll \frac{uq}{X}} \frac{(\mathfrak{d}_\gamma^2, q)}{q} \left(\tilde{\mathfrak{f}}_{M\gamma}(m,-n), \mathfrak{d}_\gamma^2\right)^{\frac{1}{2}} \\
\ll & \frac{N^\epsilon UX^4}{QJ} \sum_{u<U} \sum_{\gamma\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \frac{(\mathfrak{d}_\gamma^2, q)}{q} \sum_{d|\mathfrak{d}_\gamma^2} d^{\frac{1}{2}} \sum_{m,n\ll\frac{uq}{X}} \mathbf{1}\{\tilde{\mathfrak{f}}_{M\gamma}(m,-n)\equiv 0(d)\} \\
\ll & \frac{N^\epsilon UX^4}{QJ} \sum_{u<U} \sum_{\gamma\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \frac{(\mathfrak{d}_\gamma^2, q)}{q} \sum_{d|\mathfrak{d}_\gamma^2} d^{\frac{1}{2}} \left(\frac{u^2q^2}{X^2 d^{\frac{1}{2}}} + \frac{uq}{X}\right) \\
\ll & \frac{N^\epsilon U^4X^4}{QJ} \sum_{\gamma\in\mathfrak{F}_T} \sum_{Q<q\leqslant 2Q} \frac{(\mathfrak{d}_\gamma^2, q)}{q} \cdot \frac{T^2q}{X} \\
\ll & \frac{N^\epsilon U^4X^3T^2}{QJ} \sum_{\gamma\in\mathfrak{F}_T} \sum_{d|\mathfrak{d}_\gamma^2} d \sum_{\substack{Q<q\leqslant 2Q \\ q\equiv 0(d)}} 1 \\
\ll & N^\epsilon U^4X^2T^{2\delta} \ll N^{1+\epsilon}T^{4\delta-4}(U^4T^{2-2\delta}) \\
\ll & N^{1-\eta}T^{4\delta-4}
\end{aligned} \tag{7.69}$$

Thus we have a power saving here, too.

Put together, (7.64), (7.66), (7.69) lead to the desired bound in Lemma 7.10. □

With the bounds on $\mathcal{I}_1$, $\mathcal{I}_2$, and $\mathcal{I}_3$ that we have obtained here, Theorem 4.5 and the main Theorem 1.6 follow.

8. Spectral Gap for a Class of Kleinian Groups

In this section, we prove Theorem 1.3, which will in particular imply that a familial group $\mathcal{A}$ has a geometric spectral gap. Theorem 1.3 concerns more generally an infinite-covolume, geometrically finite, Zariski dense Kleinian group $\mathcal{A} < \mathrm{PSL}_2(\mathcal{O}_K)$ containing a Zariski dense subgroup $\Gamma < \mathrm{PSL}_2(\mathbb{Z})$.

We first simplify the situation by moving to SL_2 instead of PSL_2. In particular, if let $\mathcal{A}'$ be the preimage of $\mathcal{A}$ in SL_2, then the quotients $\mathcal{A}'(q)\backslash\mathbb{H}$ and $\mathcal{A}(q)\backslash\mathbb{H}$ are the same for all $q\in\mathbb{Z}^+$. Therefore, their geometric spectral theories agree. The properties of being geometrically finite, infinite-covolume, Zariski dense and having a Zariski dense surface subgroup are preserved.

Assume also that $\mathcal{A}$ is not itself contained in $\mathrm{SL}_2(\mathbb{Z})$ (in which case it has a spectral gap in the senses described below by [8]).

Assume also that Γ has a multiplicative structure, in the sense that for any $q=\prod_i p_i^{n_i}$,

$$\Gamma/\Gamma(q) \cong \prod_i \Gamma/\Gamma(p_i^{n_i}).$$

For, if Γ does not have this multiplicative structure, we replace Γ by $\widehat{\Gamma} := \Gamma\cap\Lambda$, where Λ is a principal congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, so that $\widehat{\Gamma}$ has a multiplicative structure. The existence of this subgroup is guaranteed by the strong approximation property. Then $\widehat{\Gamma}$ still has Zariski closure $\mathrm{SL}_2(\mathbb{R})$ as it is finite index in Γ (See (0.13), Chapter I of [22]).

As a byproduct, we prove a version of strong approximation for $\mathcal{A}$, as follows.

Theorem 8.1. *Let $\mathcal{A}$ and Γ be as above. There exists an integer P_{bad} depending on $\mathcal{A}$, such that if $q \in \mathbb{Z}$, with $q = q_{bad} \cdot q_1$ where $(q_1, P_{bad}) = 1$, we have*

(i)
$$\mathcal{A}/\mathcal{A}(q) \cong \mathcal{A}/\mathcal{A}(q_{bad}) \times \mathcal{A}/\mathcal{A}(q_1)$$

(ii)
$$\begin{aligned}\mathcal{A}/\mathcal{A}(q_1) &= SL_2(\mathcal{O}_K)/SL_2(\mathcal{O}_K)(q_1)\\ &\cong \prod_{p_i^{n_i}||q_1} SL_2(\mathcal{O}_K)/SL_2(\mathcal{O}_K)(p_i^{n_i})\end{aligned}$$

(iii) *For each $p|P_{bad}$, there exists $m_p \geqslant 1$ such that for all $k_p \geqslant m_p$,*
$$\mathcal{A}\left(p^{m_p}\right)\Big/\mathcal{A}\left(p^{k_p}\right) = SL\left(2,\mathcal{O}_K\right)\left(p^{m_p}\right)\Big/SL\left(2,\mathcal{O}_K\right)\left(p^{k_p}\right).$$
Moreover, we can choose m_p so that $m_p \leqslant m'_p + \iota_p$, where m'_p is the smallest positive power m of p such that for all $k'_p \geqslant m'_p$,
$$\Gamma(p^{m'_p})/\Gamma(p^{k'_p}) = SL_2(\mathbb{Z})(p^{m'_p})/SL_2(\mathbb{Z})(p^{k'_p}).$$
and ι_p is the smallest non-negative integer such that
$$p^{\iota_p}\mathfrak{sl}(2,\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathcal{O}_K) \subset \mathrm{Span}_{\mathbb{Z}_p}(\mathcal{A}\cdot\mathfrak{sl}(2,\mathbb{Z}_p)).$$
In this notation, the action of $\mathcal{A}$ is the restriction of the adjoint action of the Lie group SL on its Lie algebra $\mathfrak{sl}$, i.e. conjugation.

(iv) *If $p|P_{bad}$, then p can only be possibly one of the following:*
   (a) *$p = 2, 3$, or*
   (b) *p is such that $\Gamma/\Gamma(p) \neq SL_2(\mathbb{Z}/p\mathbb{Z})$, or*
   (c) *p is a common factor of all curvatures in the associated orbit $\mathcal{A}\cdot\mathbb{P}^1(\mathbb{R})$ (after scaling all raw curvatures by $\frac{1}{\sqrt{-\Delta}}$).*

Note that in the case of a familial group $\mathcal{A}$ (which is the object of this paper), Γ can be taken to be the principal congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ contained in $\mathcal{A}$, in which case the bad primes of the second kind in the theorem above are simply those dividing the level of this congruence subgroup.

We begin with the definitions of geometric and combinatorial spectral gaps for $\mathcal{A}$. Let Δ be the hyperbolic Laplacian operator associated to the metric $ds^2 = \frac{dx^2+dy^2+dz^2}{z^2}$ on $\mathbb{H}^3$:
$$\Delta = -z^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + z\frac{\partial}{\partial z}.$$
For any integer $q > 1$, let $\mathcal{A}(q)$ denote the kernel of reduction modulo q. The operator Δ is symmetric and positive definite on $L^2(\mathcal{A}(q)\backslash\mathbb{H}^3)$ with the standard inner product. From Lax-Phillips [20], the discrete spectrum consists of finitely many eigenvalues
$$\lambda_0(q) = \delta(2-\delta) < \lambda_1(q) \leqslant \lambda_2(q) \cdots .$$
From [33], we have $\lambda_0(q) \neq \lambda_1(q)$. If there exists $\epsilon > 0$ independent of q such that $\lambda_1(q) - \lambda_0(q) \geqslant \epsilon$ for all q then ϵ is called a geometric spectral gap and $\mathcal{A}$ is said to have a *geometric spectral gap*.

We now recall the definition of a combinatorial spectral gap for $\mathcal{A}$. Suppose $\mathcal{A}$ has a finite symmetric generating set S. Let

$$\lambda_n'(\mathcal{A}, S) \leqslant \cdots \leqslant \lambda_1'(\mathcal{A}, S) \leqslant \lambda_0'(\mathcal{A}, S) = 1$$

denote the eigenvalues of the averaging operator $T_{\mathcal{A},S} := 1 - \Delta_{\mathcal{A},S}/|S|$ where $\Delta_{\mathcal{A},S}$ is the discrete Laplacian operator

$$(\Delta_{\mathcal{A},S} f)(g) = \sum_{h \in S} (f(g) - f(hg)).$$

We say that $\mathcal{A}$ has a *combinatorial spectral gap* if there is a finite symmetric collection of generators S and a positive ϵ such that

$$\lambda_1'(\mathcal{A}/\mathcal{A}(q), \overline{S}) < 1 - \epsilon$$

for all positive integers q, where ϵ is independent of q (here $\overline{S}$ denotes the image of S modulo q). Informally, a spectral gap for $\mathcal{A}/\mathcal{A}(q)$ gives a measure of how quickly a random walk on the Cayley graph of $\mathcal{A}/\mathcal{A}(q)$ reaches the whole graph. A spectral gap for $\mathcal{A}$ indicates a uniform rate for all q.

We now give an overview of the proof of Theorem 1.3. Let T be an element of $\mathcal{A}$ which does not normalize $\mathrm{SL}_2(\mathbb{R})$, i.e., $T \notin \mathrm{SL}_2(\mathbb{R}) \cup i\mathrm{SL}_2(\mathbb{R}) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Write $\Gamma' = T\Gamma T^{-1}$, and let $\mathcal{A}' = \langle \Gamma, \Gamma' \rangle < \mathcal{A}$. We first prove a combinatorial spectral gap for $\mathcal{A}'$, using ideas similar to those of Varjú in the appendix of [6], some of which have also been used by Sarnak in [26], Shalom in [30], and Kassabov-Lubotzky-Nikolov in [16]. We then convert this to a combinatorial spectral gap for $\mathcal{A}$. Finally, we use the fact that a combinatorial spectral gap for $\mathcal{A}$ implies a geometric spectral gap if the Hausdorff dimension of the limit set of $\mathcal{A}$ is greater than 1 via a variant of [4, Theorem 1.2], which states that geometric and combinatorial spectral gaps co-occur as long as the Hausdorff dimension of the limit set of the group is greater than 1.

Proposition 8.2. *$\mathcal{A}'$ has a combinatorial spectral gap.*

As a Zariski-dense subgroup of $\mathrm{SL}_2(\mathbb{Z})$, Γ is known to have a spectral gap (see [8]), and therefore so does Γ'. We will show that $\mathcal{A}'/\mathcal{A}'(q)$ is made up of a bounded number of copies of $\Gamma/\Gamma(q)$ and $\Gamma'/\Gamma'(q)$, which will imply a spectral gap for $\mathcal{A}'/\mathcal{A}'(q)$. To be precise, we quote a Lemma of Varjú:

Lemma 8.3 [6, Lemma A.4]. *Let G be a finite group with a finite symmetric generating set S. Suppose $G_1, \ldots, G_k < G$, and that for each $g \in G$, there exist $g_i \in G_i$ such that $g = g_1 g_2 \ldots g_k$. Then,*

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

As a consequence, we have immediately:

Lemma 8.4. *Suppose G is a group with a finite symmetric generating set S and a tuple $(G_1, G_2, \ldots, G_k)$ of subgroups of G such that:*

(i) *each G_i has a spectral gap;*

(ii) *$S \cap G_i \neq \emptyset$ for each G_i;*

(iii) *for each integer q, for each $g \in \rho_q(G)$, it is possible to write $g = g_1 g_2 \cdots g_k$ where $g_i \in \rho_q(G_i)$, i.e.*

$$\rho_q(G) = \prod_{i=1}^{k} \rho_q(G_i).$$

Then G has a combinatorial spectral gap.

To verify the hypotheses of Lemma 8.4 for $G = \mathcal{A}'$, we will use $k = 2k_0$, and the tuple $(G_1, G_2, \ldots, G_{2k_0}) = (\Gamma, \Gamma', \Gamma, \Gamma', \ldots, \Gamma')$. Write

$$A_k(q) = \{g_1 h_1 \cdots g_k h_k : g_1, \cdots g_k \in \Gamma/\Gamma(q), h_1 \cdots h_k \in \Gamma^{'}/\Gamma^{'}(q)\}.$$

Then, for the third hypothesis of Lemma 8.4, we need to show:

Lemma 8.5. *There exists some k_0 such that $A_{k_0}(q) = \mathcal{A}^{'}/\mathcal{A}^{'}(q)$ for every q.*

Our approach is to break q into prime powers, and prove a universal bound for prime powers for all but finitely many 'bad primes'. We therefore break the proof into two lemmata dealing with the good primes and bad primes, respectively. The first lemma uses some geometric arguments to construct elements of $\mathcal{A}'/\mathcal{A}'(p^m)$ in terms of Γ and Γ'. The second lemma works prime-by-prime, and uses the Lie algebra $\mathfrak{sl}_2$ to lift to higher powers of p uniformly.

Lemma 8.6. *There exists a finite set of primes $\mathcal{S}$ such that, for $p \notin \mathcal{S}$, and for all $m \geqslant 1$, we have*

$$A_{14}(p^m) = \mathcal{A}'/\mathcal{A}'(p^m).$$

Proof. Throughout the proof we assume $p \notin \mathcal{S}$, and we augment $\mathcal{S}$ as necessary while preserving its finiteness.

Consider $C_T = T^{-1} \cdot \mathbb{P}^1(\mathbb{R})$. If C_T is a line, let γ be the identity matrix. Otherwise it is a circle, and we write $r\sqrt{d}$ for its radius, $x_0 + \sqrt{-d}y_0$ for its center, and let

$$\gamma = \begin{pmatrix} 1 & -x_0 \\ 0 & 1 \end{pmatrix}.$$

Note that x_0, y_0, r are rational numbers which may be written with denominator b, the curvature of C_T (formulae for these integers in terms of the entries of T are given in [31, Proposition 3.7]). Then the intersection points of $\gamma T^{-1} \cdot \mathbb{P}^1(\mathbb{R})$ with the imaginary axis are of the form $\sqrt{-d}s$ where, in the case that C_T is a line, $\sqrt{-d}s$ is the height of the line, and if C_T is a circle, $s = y_0 \pm r$, and $r\sqrt{d}$ is the radius of $T^{-1} \cdot \mathbb{P}^1(\mathbb{R})$. In any case, choose such an s, and remark that γ and s are defined over $\mathbb{Q}$.

Consider reduction modulo $p^m\mathcal{O}_K$ on the projective line:

$$\rho_{p^m} : \mathbb{P}^1(\mathcal{O}_K) \to \mathbb{P}^1(\mathcal{O}_K/(p^m)).$$

Then the reduction map

$$\rho_{p^m} : \mathrm{SL}_2(\mathcal{O}_K) \to \mathrm{SL}_2(\mathcal{O}_K/(p^m))$$

is equivariant with respect to reduction on the projective line. Let $\mathcal{S}$ contain any primes where

$$\rho_{p^m} : \Gamma \to \mathrm{SL}_2(\mathbb{Z}/(p^m))$$

is not surjective for some $m \geqslant 1$ (there are finitely many such, by strong approximation for Γ). We allow for p to be inert, split, or ramified.

Let $\mathcal{S}$ also contain any primes appearing in denominators of γ, so that $\rho_{p^m}(\gamma)$ is defined, and has a lift in Γ. By expanding $\mathcal{S}$, we may assume p and s are coprime.

Therefore s is invertible modulo p^m and there is a representation $\phi_s : \mathcal{O}_K/(p^m) \to M(2, \mathbb{Z}/(p^m))$ given by

$$x + y\sqrt{-d} \mapsto \begin{pmatrix} x & -yds \\ ys^{-1} & x \end{pmatrix}.$$

In particular, the eigenvalues of the matrix are $x \pm y\sqrt{-d}$, and the determinant is the norm $N(x + y\sqrt{-d})$. It has exactly two fixed points modulo p^m, namely $\pm s\sqrt{-d}$.

Let x and y be a solution to $x^2 + dy^2 \equiv 1 \pmod{p^m}$ having $\gcd(xy, p) = 1$. The existence of such is a consequence of an argument with Gauss sums [9, Exercise 13(v), p. 32], if $p \geqslant 5$. Therefore let $2, 3 \in \mathcal{S}$. Therefore, $x + y\sqrt{-d}$ is of norm 1 modulo p^m, so that $\phi_s(x + y\sqrt{-d})$ is in $\mathrm{SL}_2(\mathbb{Z}/(p^m))$, and therefore has a lift, call it T_0, in Γ. We guarantee that neither of $(x \pm y\sqrt{-d})^2$ are equivalent to integers modulo p^m (i.e. in the subring $\mathbb{Z}/(p^m) \subset \mathcal{O}_K/(p^m)$), since $p \nmid 2xy$ by construction.

Therefore $T\gamma^{-1}T_0\gamma T^{-1}$, considered modulo p^m, has a fixed point in $\mathbb{P}^1(\mathbb{Z}/(p^m))$. Since $\mathrm{SL}_2(\mathbb{Z}/(p^m))$ is transitive on $\mathbb{P}^1(\mathbb{Z}/(p^m))$, we can conjugate this fixed point to ∞ modulo p^m. Therefore, we find an element T_1 in $\Gamma T\Gamma T^{-1}\Gamma$ which fixes ∞ modulo p^m.

So T_1 has the form

$$T_1 \equiv \begin{pmatrix} a_0 & b \\ 0 & a_1 \end{pmatrix} \pmod{p^m},$$

where $a_0, a_1 \in \mathcal{O}_K$, $a_0 a_1 \equiv 1 \pmod{p^m}$. As a_0 and a_1 are the eigenvalues of T_1 and hence T_0, they are $x \pm y\sqrt{-d}$. In particular, we have arranged that $a_0^2 \notin \mathbb{Z}/(p^m)$.

Now take

$$T_{2,n} = T_1 \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} T_1^{-1} \equiv \begin{pmatrix} 1 & na_0^2 \\ 0 & 1 \end{pmatrix}.$$

We know $a_0^2 \notin \mathbb{Z}/(p^m)$ and a_0^2 is invertible. Now, this implies that $a_0^2\mathbb{Z}/(p^m) + \mathbb{Z}/(p^m) = \mathcal{O}_K/(p^m)$. This implies that all upper triangular matrices are in $\Gamma T\Gamma T^{-1}\Gamma T\Gamma T^{-1}\Gamma$ modulo p^m.

The rest of the proof follows Varjú. Specifically, an exactly analogous argument shows that the lower triangular matrices with 1's on the diagonal are also in $\Gamma T\Gamma T^{-1}\Gamma T\Gamma T^{-1}\Gamma$ modulo p^m. Therefore, in $A_7(p^m)$ we obtain all elements of the form

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

This includes any matrix γ with lower-left entry not congruent to 0 modulo p, since it is possible to solve for a, b, c modulo p^m in that circumstance. As this is more than half of the group $\rho_{p^m}(\mathcal{A}')$, the emma is proved. □

Lemma 8.7. *Let p be any prime. Then there exists some positive integers k_p and m_p such that*

$$A_{k_p}(p^m) = \mathcal{A}'/\mathcal{A}'(p^m)$$

for all $m \geqslant m_p$.

Proof. Let SL_2 act on $\mathfrak{sl}_2$ via the standard adjoint action of a Lie group on its Lie algebra by conjugation, i.e.

$$\mathrm{SL}_2 \times \mathfrak{sl}_2 \to \mathfrak{sl}_2, \quad g \times v \mapsto g \cdot v := gvg^{-1}.$$

We will first find a $\mathbb{Q}_p$-basis of $\mathfrak{sl}(2, \mathbb{Q}_p \otimes_{\mathbb{Q}} K_d)$ formed of elements from $\mathfrak{sl}(2, \mathbb{Q})$ and $\Gamma T \cdot \mathfrak{sl}(2, \mathbb{Q})$. Using this basis, we will apply an inductive argument to show that, for all $m \geqslant m_p$ (where m_p will

be defined below), a finite-index subgroup of $\mathrm{SL}_2(\mathcal{O}_K)/\mathrm{SL}_2(\mathcal{O}_K)(p^m)$, whose index is independent of m, is contained in $A_4(p^m)$.

To find the aforementioned basis, we begin with the standard basis for the real Lie algebra $\mathfrak{sl}(2,\mathbb{R})$:

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The above is also a $\mathbb{Q}_p$-basis for $\mathfrak{sl}(2,\mathbb{Q}_p)$ for any p, and a $\mathbb{Q}$-basis for $\mathfrak{sl}(2,\mathbb{Q})$.

First, we remark that $\Gamma(v)$ spans $\mathrm{SL}_2(\mathbb{R})(v)$ over $\mathbb{R}$ for any non-zero $v \in \mathfrak{sl}(2,\mathbb{C})$. For, since Γ is Zariski dense in SL_2, and the adjoint action is Zariski continuous, the Zariski closure of the orbit $\Gamma(v)$ in $\mathfrak{sl}(2,\mathbb{R})$ is $\mathrm{SL}_2(\mathbb{R})(v)$.

Next, we claim that the orbit $\mathrm{SL}_2(\mathbb{R})(v)$ must be of real dimension 3. This follows from irreducibility of the adjoint action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathfrak{sl}(2,\mathbb{R})$ in the case that $v \in \mathfrak{sl}(2,\mathbb{R})$. In fact, the same elementary irreducibility argument shows that the orbit $\mathrm{SL}_2(\mathbb{R})(v)$ for any $v \in \mathfrak{sl}(2,\mathbb{C})$ is at least 3-dimensional (any v can be conjugated to be diagonal, hence λH with $\lambda \in \mathbb{C}$; then conjugations and linear combinations yield λR and λL).

Furthermore, for $v \notin \mathfrak{sl}(2,\mathbb{R})$, we have $\mathrm{SL}_2(\mathbb{R})(v) \cap \mathfrak{sl}(2,\mathbb{R}) = \{0\}$. By dimensional considerations, then, in this case

$$\mathrm{Span}_{\mathbb{R}}(\Gamma(v), \mathfrak{sl}(2,\mathbb{R})) = \mathfrak{sl}(2,\mathbb{C}).$$

Next, we show that the stabilizer of $\mathfrak{sl}(2,\mathbb{R})$ under the adjoint action of $\mathrm{SL}_2(\mathbb{C})$ is exactly $\mathrm{SL}_2(\mathbb{R}) \cup i\mathrm{SL}_2(\mathbb{R})$. For, suppose M is in the stabilizer. Then, taking $\mathfrak{m} \in \mathfrak{sl}(2,\mathbb{R}) \cap \mathrm{SL}_2(\mathbb{R})$ (for example, an elliptic element of order 2 with fixed points on $\mathbb{P}^1(\mathbb{R})$), we find that it must stabilize the circle $M(\mathbb{P}^1(\mathbb{R}))$, which is only possible if $M(\mathbb{P}^1(\mathbb{R})) = \mathbb{P}^1(\mathbb{R})$. Hence the stabilizer of $\mathfrak{sl}(2,\mathbb{R})$ is contained in the stabilizer of $\mathbb{P}^1(\mathbb{R})$ under the $\mathrm{SL}_2(\mathbb{C})$ action on $\mathbb{P}^1(\mathbb{C})$.

We have assumed $T \notin \mathrm{SL}_2(\mathbb{R}) \cup i\mathrm{SL}_2(\mathbb{R}) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Hence, by simplicity, $T(\mathfrak{sl}(2,\mathbb{R})) \cap \mathfrak{sl}(2,\mathbb{R}) = \{0\}$. In particular, we may take any $w \in \mathfrak{sl}(2,\mathbb{Q})$, and obtain $T(w) \notin \mathfrak{sl}(2,\mathbb{R})$. We may now conclude that for some appropriate choice of $\gamma_2, \gamma_3, \gamma_4 \in \Gamma$, we have:

$$\mathrm{Span}_{\mathbb{R}}\{H, R, L, w_2 = \gamma_2(T(w)), w_3 = \gamma_3(T(w)), w_4 = \gamma_4(T(w))\} = \mathfrak{sl}(2,\mathbb{C}).$$

Let W denote this basis, where we have chosen $w \in \mathfrak{sl}(2,\mathbb{Z})$.

We may actually conclude that W is a $\mathbb{Q}$-basis of $\mathfrak{sl}(2,K_d)$, which is 3 K_d-dimensional, hence 6 $\mathbb{Q}$-dimensional. We may extend scalars and find that W is also a $\mathbb{Q}_p$-basis of $\mathfrak{sl}(2,\mathbb{Q}_p \otimes_{\mathbb{Q}} K_d)$.

We have therefore found the desired $\mathbb{Q}_p$-basis of $\mathrm{SL}_2(\mathbb{Q}_p \otimes_{\mathbb{Q}} K_d)$, namely W.

Next we define m_p. Since Γ is Zariski dense, for each p we can find a positive m'_p such that for all $m \geqslant m'_p$, $\Gamma(p^m)$ is dense in $\mathrm{SL}_2(\mathbb{Z}_p)(p^m)$. For technical reasons, we take $m_p = m'_p + \iota_p$ where ι_p is the smallest non-negative integer so that

$$p^{\iota_p}\mathfrak{sl}(2,\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathcal{O}_K) \subset \mathrm{Span}_{\mathbb{Z}_p}(W). \tag{8.1}$$

This ι_p is necessarily finite. In the case that W is a $\mathbb{Z}_p$-integral basis of $\mathfrak{sl}(2,\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathcal{O}_K)$, then $\iota_p = 0$, and the technical condition may be dropped in the sense that $m_p = m'_p$.

Next, we prove the following **claim:** For any $g \in \mathrm{SL}_2(\mathcal{O}_K)(p^{m_p})/\mathrm{SL}_2(\mathcal{O}_K)(p^m)$ and any $m \geqslant m_p$, we may express g as

$$g \equiv L_1(\gamma_2 T L_2 T^{-1}\gamma_2^{-1})(\gamma_3 T L_3 T^{-1}\gamma_3^{-1})(\gamma_4 T L_4 T^{-1}\gamma_4^{-1}) \pmod{p^m}$$

for some $L_1, L_2, L_3, L_4 \in \Gamma$.

This would imply $g \in A_4(p^m)$.

We prove this by induction. The base case $m = m_p$ is trivial. Suppose for $m = k$ we can find $L_{1,k}, L_{2,k}, L_{3,k}, L_{4,k} \in \Gamma$ such that

$$g \equiv L_{1,k}(\gamma_2 T L_{2,k} T^{-1} \gamma_2^{-1})(\gamma_3 T L_{3,k} T^{-1} \gamma_3^{-1})(\gamma_4 T L_{4,k} T^{-1} \gamma_4^{-1}) \pmod{p^k}$$

Then

$$g = L_{1,k}(\gamma_2 T L_{2,k} T^{-1} \gamma_2^{-1})(\gamma_3 T L_{3,k} T^{-1} \gamma_3^{-1})(\gamma_4 T L_{4,k} T^{-1} \gamma_4^{-1}) + p^k u$$

for some $u \in \mathfrak{sl}(2, \mathbb{Z}_p \otimes_{\mathbb{Z}} \mathcal{O}_K)$. Therefore, using the basis W, and the fact that $\mathcal{A}$, R, L give a $\mathbb{Z}_p$-integral basis for $\mathfrak{sl}(2, \mathbb{Z}_p)$, we can find $u_1 \in \mathfrak{sl}(2, \mathbb{Z}_p)$, and $t_2, t_3, t_4 \in \mathbb{Q}_p$ so that

$$u = u_1 + t_2 w_2 + t_3 w_3 + t_4 w_4.$$

If W forms a $\mathbb{Z}_p$-integral basis for $\mathfrak{sl}(2, \mathbb{Z}_p \otimes_{\mathbb{Z}} \mathcal{O}_K)$, then $t_i \in \mathbb{Z}_p$ for $i = 2, 3, 4$. Otherwise,

$$t_i \in p^{-\iota_p} \mathbb{Z}_p,$$

(by (8.1)). This implies

$$t_i p^k w \in p^{k-\iota_p} \mathfrak{sl}(2, \mathbb{Z}_p).$$

Since $\Gamma(p^{k-\iota_p})$ is dense in $\mathrm{SL}_2(\mathbb{Z}_p)(p^{k-\iota_p})$, we can find $\beta_1, \beta_2, \beta_3, \beta_4 \in \Gamma$ such that $\beta_i \equiv I \pmod{p^{k-\iota_p}}$ and

$$\begin{aligned}
\beta_1 &\equiv I + p^k u_1 \pmod{p^{k+1}} \\
\beta_2 &\equiv I + t_2 p^k w \pmod{p^{k+1}} \\
\beta_3 &\equiv I + t_3 p^k w \pmod{p^{k+1}} \\
\beta_4 &\equiv I + t_4 p^k w \pmod{p^{k+1}}
\end{aligned}$$

Then we set $L_{i,k+1} = L_{i,k}\beta_i$ for $i = 1, 2, 3, 4$. This is enough to prove the statement for $m = k+1$ (here, we rely on the fact that $k, k - \iota_p \geqslant 1$):

$$\begin{aligned}
&L_{1,k+1}(\gamma_2 T L_{2,k+1} T^{-1} \gamma_2^{-1})(\gamma_3 T L_{3,k+1} T^{-1} \gamma_3^{-1})(\gamma_4 T L_{4,k+1} T^{-1} \gamma_4^{-1}) \pmod{p^{k+1}} \\
\equiv &L_{1,k}(\gamma_2 T L_{2,k} T^{-1} \gamma_2^{-1})(\gamma_3 T L_{3,k} T^{-1} \gamma_3^{-1})(\gamma_4 T L_{4,k} T^{-1} \gamma_4^{-1}) + p^k u \equiv g \pmod{p^{k+1}}.
\end{aligned}$$

This completes the induction. Therefore, we have $g \in A_4(p^m)$ for any $g \in \mathrm{SL}_2(\mathcal{O}_K)(p^{m_p})/\mathrm{SL}_2(\mathcal{O}_K)(p^m)$ and any $m \geqslant m_p$.

Now, $[\mathrm{SL}_2(\mathcal{O}_K) : \mathrm{SL}_2(\mathcal{O}_K)(p^{m_p})] \leqslant p^{6m_p}$. It must be that $A_1(p^m)$ contains something outside $\mathrm{SL}_2(\mathcal{O}_K)(p^{m_p})/\mathrm{SL}_2(\mathcal{O}_K)(p^m)$. Therefore, $A_{4+1}(p^m)$ contains at least two cosets; $A_{4+2}(p^m)$ contains at least 3 cosets and so forth. So if we set $k_p = 4 + p^{6m_p}$, we have $A_{k_p}(p^m) = \mathcal{A}'/\mathcal{A}'(p^m)$ for all $m \geqslant m_p$. □

Proof of Lemma 8.5. For each p and m, there is a k_p such that

$$\mathcal{A}'/\mathcal{A}'(p^m) = A_{k_{p,m}}(p^m).$$

For $p \in \mathcal{S}$, this $k_{p,m}$ is uniform with respect to p (Lemma 8.6), while for any fixed $p \notin \mathcal{S}$, this $k_{p,m}$ is uniform for $m \geqslant m_p$ (Lemma 8.7). As $\mathcal{S}$ is finite, the supremum of the $k_{p,m}$ is finite, say k_0. Therefore,

$$\mathcal{A}'/\mathcal{A}'(p^m) = A_{k_0}(p^m)$$

for any p, m.

We have assumed Γ and therefore $\Gamma^{'}$ have a multiplicative structure. In other words, for any $q = \prod_i p_i^{n_i}$, we have

$$\Gamma/\Gamma(q) \cong \prod_i \Gamma/\Gamma(p_i^{n_i}),$$
$$\Gamma^{'}/\Gamma^{'}(q) \cong \prod_i \Gamma^{'}/\Gamma^{'}(p_i^{n_i}).$$

A direct corollary is that $\mathcal{A}^{'}$ also has a multiplicative structure

$$\mathcal{A}^{'}/\mathcal{A}^{'}(q) \cong \prod_i \mathcal{A}^{'}/\mathcal{A}^{'}(p_i^{n_i}), \tag{8.2}$$

since $\mathcal{A}^{'}$ is generated by Γ and $\Gamma^{'}$, and that A_{k_0} has a multiplicative structure:

$$A_{k_0}(q) \cong \prod_i A_{k_0}(p_i^{n_i}).$$

These isomorphisms are compatible so that the composition of isomorphisms

$$A_{k_0}(q) \cong \prod_i A_{k_0}(p_i^{n_i}) = \prod_i \mathcal{A}'/\mathcal{A}'(p_i^{n_i}) \cong \mathcal{A}^{'}/\mathcal{A}^{'}(q)$$

is the identity map. Therefore,

$$A_{k_0}(q) = \mathcal{A}^{'}/\mathcal{A}^{'}(q)$$

as desired. □

Proof of Proposition 8.2. We verify the hypotheses of Lemma 8.4 for $G = \mathcal{A}'$, $S = S' \cup TS'T^{-1}$, where S is a finite set of generators for Γ, $k = 2k_0$, and $(G_1, G_2, \ldots, G_{2k_0}) = (\Gamma, \Gamma', \Gamma, \Gamma', \ldots, \Gamma')$. The group Γ has a spectral gap as a Zariski dense subgroup of $\mathrm{SL}_2(\mathbb{Z})$, by [8]; hence Γ' does also. The second hypothesis is immediate, and the third is verified by Lemma 8.5. Therefore $\mathcal{A}'$ has a combinatorial spectral gap. □

Next, we wish to pass from $\mathcal{A}'$ to $\mathcal{A}$.

Proposition 8.8. *$\mathcal{A}$ has a combinatorial spectral gap.*

Before proving this, we note that our main spectral theorem follows immediately.

Proof of Theorem 1.3. The theorem follows from the fact that $\mathcal{A}$ has a combinatorial spectral gap (Proposition 8.8) and a version of [4, Theorem 1.2] for $\mathrm{SL}_2(\mathcal{O}_K)$ giving equivalence of geometric and spectral gaps when the Hausdorff dimension of the limit set of the group is greater than 1, which would follow from the arguments in [4] modified as described in the paragraph preceding Theorem 2.1 in [4]. □

To prove Proposition 8.8, we recall an equivalent condition for a combinatorial spectral gap to the one given at the beginning of this section. Given a graph G and subset V, write ∂V for the set of edges joining V to its complement in G. Then define the *expansion ratio* of G to be

$$h_G := \min_{V \subset G, |V| \leqslant \frac{1}{2}|G|} \frac{|\partial V|}{|V|}.$$

Let ϵ_G be the gap between the two biggest eigenvalues of the discrete Laplacian operator on G. It is known the expansion ratio of G is related to ϵ_G by the inequalities [19, Propositions 3.2.31,

3.2.33]:

$$\frac{h_G^2}{2M_G^2} \leqslant \epsilon_G \leqslant 2M_G h_G,$$

where M_G is the maximum valence of vertices in G. In particular, $h_{\rho_q G}$ is bounded away from 0 uniformly with respect to q if and only if G satisfies a combinatorial spectral gap.

Proof of Proposition 8.8. We will demonstrate the existence of a positive constant h such that for any positive integer q, and any $V \subset \mathcal{A}/\mathcal{A}(q)$ with $|V| \leqslant \frac{1}{2}|\mathcal{A}/\mathcal{A}(q)|$, we have

$$|\partial V| \geqslant h|V|. \tag{8.3}$$

We use the corresponding property for $\mathcal{A}'$ (which has a combinatorial spectral gap by Proposition 8.2). Let h_0 be such that for any positive integer q and any $V \subset \mathcal{A}'/\mathcal{A}'(q)$ with $|V| \leqslant \frac{1}{2}|\mathcal{A}'/\mathcal{A}'(q)|$, we have

$$|\partial V| \geqslant h_0|V|. \tag{8.4}$$

By the strong approximation property for $\mathcal{A}$ and $\mathcal{A}'$, there is a universal M such that the index $[\mathcal{A}/\mathcal{A}(q) : \mathcal{A}'/\mathcal{A}'(q)] \leqslant M$. Let S be a finite generating set for $\mathcal{A}$, which is symmetric under inverses (this exists since we assume $\mathcal{A}$ is geometrically finite, hence finitely generated). We say two cosets $a\mathcal{A}'/\mathcal{A}'(q)$ and $a'\mathcal{A}'/\mathcal{A}'(q)$ are connected if there exists some $s \in S$ such that $sa\mathcal{A}'/\mathcal{A}'(q) = a'\mathcal{A}'/\mathcal{A}'(q)$. By the symmetry of S, this connectedness is an equivalence relation.

Fix q. Let $a_1\mathcal{A}'/\mathcal{A}'(q), \cdots, a_l\mathcal{A}'/\mathcal{A}'(q)$ be the cosets of $\mathcal{A}'/\mathcal{A}'(q)$ in $\mathcal{A}/\mathcal{A}(q)$, with $l \leqslant M$. If $l = 1$, then (8.3) follows trivially from (8.4), with $h = h_0$, for this value of q. Therefore, assume $l \geqslant 2$.

Let $V_i = V \cap a_i\mathcal{A}'/\mathcal{A}'(q)$ and define

$$\kappa = \max\left\{\big||V_i| - |V_j|\big| : a_i\mathcal{A}'/\mathcal{A}'(q) \text{ and } a_j\mathcal{A}'/\mathcal{A}'(q) \text{ are connected}\right\}.$$

Case 1: $\kappa \leqslant \frac{|V|}{10l^2}$. Then we have

$$\max\{|V_i|\} - \min\{|V_i|\} \leqslant l\kappa \leqslant \frac{|V|}{10l}.$$

From this, one finds that for each i,

$$\frac{9}{10}\frac{|V|}{l} \leqslant |V_i| \leqslant \frac{11}{10}\frac{|V|}{l}. \tag{8.5}$$

Case 1a: If $|V| \leqslant \frac{10}{22}|\mathcal{A}/\mathcal{A}(q)|$, then each $|V_i| \leqslant \frac{1}{2}|\mathcal{A}'/\mathcal{A}'(q)|$. Applying (8.4), we have $|Eg(V_i, a_i\mathcal{A}'/\mathcal{A}'(q) - V_i)| \geqslant h_0|V_i|$. Therefore,

$$|\partial V| \geqslant \sum_i |Eg(V_i, a_i\mathcal{A}'/\mathcal{A}'(q) - V_i)| \geqslant h_0|V|. \tag{8.6}$$

Case 1b: If $|V| \geqslant \frac{10}{22}|\mathcal{A}/\mathcal{A}(q)|$, then from (8.5) we have

$$\frac{9}{22}\frac{|\mathcal{A}'/\mathcal{A}'(q)|}{l} \leqslant |V_i| \leqslant \frac{11}{22}\frac{|\mathcal{A}'/\mathcal{A}'(q)|}{l}.$$

And then it can be worked out that

$$|\partial V| \geqslant \frac{9h_0}{22}|\mathcal{A}/\mathcal{A}(q)| \geqslant \frac{9h_0}{11}|V| \tag{8.7}$$

Case 2: $\kappa \geqslant \frac{|V|}{10l^2}$. There exists $s \in S$ such that $sa_i\mathcal{A}'/\mathcal{A}'(q) = a_j\mathcal{A}'/\mathcal{A}'(q)$ and $\big||V_i| - |V_j|\big| = \kappa$. Since multiplication by s is a bijection between $a_i\mathcal{A}'/\mathcal{A}'(q)$ and $a_j\mathcal{A}'/\mathcal{A}'(q)$, by the pigeon hole principle, multiplication by s must map at least κ elements from the bigger set, say V_i, to $a_j\mathcal{A}'/\mathcal{A}'(q) - V_j$, so we have at least

$$|\partial V| \geqslant \kappa = \frac{|V|}{10l^2} \geqslant \frac{|V|}{10M^2} \tag{8.8}$$

Combining (8.6),(8.7) and (8.8), we find we can set $h = \min\{\frac{9h_0}{11}, \frac{1}{10M^2}\}$.

□

Lastly, we prove the statement of explicit strong approximation, with reference to the proof of the spectral gap just completed.

Proof of Theorem 8.1. First, we isolate the primes p for which $\mathcal{A}/\mathcal{A}(p) \neq \mathrm{SL}_2(\mathcal{O}_K)/\mathrm{SL}_2(\mathcal{O}_K)(p)$. Lemma 8.6 shows that $\mathcal{A}/\mathcal{A}(p) = \mathrm{SL}_2(\mathcal{O}_K)/\mathrm{SL}_2(\mathcal{O}_K)(p)$ for 'good' primes, but in the course of the proof, we throw a variety of primes into $\mathcal{S}$, for which we do not prove this; they are to be dealt with as bad primes. The first class of primes placed in $\mathcal{S}$ are those arising from the denominator of γ. The denominator of γ is always a divisor of the curvature of $T^{-1} \cdot \mathbb{P}^1(\mathbb{R})$ ([31, Proposition 3.7]). Therefore, by choice of T (applying an element of Γ to T^{-1}), we can avoid any prime not dividing all curvatures in $\mathcal{A} \cdot \mathbb{P}^1(\mathbb{R})$. The second class of primes removed are those not coprime to s. However, by choice of s, we can again avoid any odd prime not dividing the curvature b (since $r = 1/b$, so that $s = (y \pm 1)/b$ for some integer y [31, Proposition 3.7]). Other primes moved to $\mathcal{S}$ during the proof are those p for which $\Gamma/\Gamma(p^m) \neq \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$ for some $m \geqslant 1$, and the special primes $p = 2, 3$. Note that if $\Gamma/\Gamma(p) = \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, then by Lemma 3 on page IV-23 of J-P.Serre in [29] one automatically has that $\Gamma/\Gamma(p^m) = \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$. Hence the statement in part 3(b) of Theorem 8.1 is equivalent to $\Gamma/\Gamma(p^m) \neq \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$ for some $m \geqslant 1$. For all primes not contained in $\mathcal{S}$, the proof demonstrates that $\mathcal{A}'/\mathcal{A}'(p) = \mathrm{SL}_2(\mathcal{O}_K)/\mathrm{SL}_2(\mathcal{O}_K)(p)$, which implies $\mathcal{A}/\mathcal{A}(p) = \mathrm{SL}_2(\mathcal{O}_K)/\mathrm{SL}_2(\mathcal{O}_K)(p)$.

Now let P_{bad} be the product of the primes of $\mathcal{S}$ as above. We obtain parts (1) and (2) immediately from the fact that $\mathcal{A}/\mathcal{A}(p) = \mathrm{SL}_2(\mathcal{O}_K)/\mathrm{SL}_2(\mathcal{O}_K)(p)$ for all other primes. Part (4) is by definition.

It remains to prove part (3). Let $p|P_{\mathrm{bad}}$. In the course of the proof of Lemma 8.7, we find that $\mathcal{A}'(p^{m_p})/\mathcal{A}'(p^k) = \mathrm{SL}_2(\mathcal{O}_K)(p^{m_p})/\mathrm{SL}_2(\mathcal{O}_K)(p^k)$, where by judicious choice of the basis W in the proof, $m_p = m'_p + \iota_p$ where ι_p is as defined as the smallest non-negative integer so that

$$p^{\iota_p}\mathfrak{sl}(2, \mathbb{Z}_p \otimes_{\mathbb{Z}} \mathcal{O}_K) \subset \mathrm{Span}_{\mathbb{Z}_p}(\mathcal{A}' \cdot \mathfrak{sl}(2, \mathbb{Z}_p)).$$

However, if the goal is only that $\mathcal{A}(p^{m_p})/\mathcal{A}(p^k) = \mathrm{SL}_2(\mathcal{O}_K)(p^{m_p})/\mathrm{SL}_2(\mathcal{O}_K)(p^k)$, and not a spectral gap for $\mathcal{A}$, the proof of Lemma 8.7 can be modified for $\mathcal{A}$ instead of $\mathcal{A}'$, as follows. Using the same justification, we find that $\mathrm{Span}_{\mathbb{Q}_p}(\mathcal{A}(\mathfrak{sl}(2, \mathbb{Q}_p)))$ is of dimension 6, hence we can find a $\mathbb{Q}_p$-basis of $\mathfrak{sl}(2, \mathbb{Q}_p \otimes_{\mathbb{Q}} K_d)$ of the form

$$H, R, L, w_2 = a_2(w), w_3 = a_3(w), w_4 = a_4(w),$$

where $w \in \mathfrak{sl}(2, \mathbb{Z}_p)$. We may choose w and a_i such that we have a $\mathbb{Z}_p$-basis for $\mathrm{Span}_{\mathbb{Z}_p}(\mathcal{A}(\mathfrak{sl}(2, \mathbb{Z}_p)))$. Running the rest of the proof with a_i in place of $\gamma_i T$, we no longer obtain a spectral gap but we obtain surjectivity with the stated ι_p. □

9. Example packings

As discussed in the introduction, Kontorovich and Nakamura present a collection of examples which satisfy the hypotheses of Theorem 1.6. Here we first present one explicit example appearing in Kontorovich and Nakamura satisfying the hypotheses of Theorem 1.6. Second, we verify that the hypotheses hold for the entire family of K-Apollonian packings.

9.1 A cuboctohedral packing

The packing presented here is neither the Apollonian packing, nor any K-Apollonian packing, but it appears as an example of a super-integral polyhedral packing of Kontorovich and Nakamura [18]. The packing is shown in Figure 1, where cuboctahedral symmetry is evident.

Define

$$G_1 = \left\langle c_1(z) = \overline{z} + \sqrt{-6}, \quad c_2(z) = \frac{\overline{z}}{\frac{-\sqrt{-6}}{6}\overline{z} + 1}, \quad c_3(z) = \frac{(1+\sqrt{-6})\overline{z} - 3\sqrt{-6}}{\frac{\sqrt{-6}}{3}\overline{z} + 1 - \sqrt{-6}} \right\rangle,$$
$$G_2 = \left\langle a_1(z) = -\overline{z}, \quad a_2(z) = -\overline{z} + 6, \quad a_3(z) = \frac{\overline{z}}{\overline{z} - 1}, \quad a_4(z) = \frac{5\overline{z} - 12}{2\overline{z} - 5} \right\rangle.$$

Define $\mathcal{A}''$ as a group generated by the fourteen reflections:

$$\mathcal{A}'' = \langle a_1, \quad a_2, \quad a_3, \quad a_4, \quad c_1a_3c_1, \quad c_1a_4c_1, \quad c_2a_4c_2, \quad c_3a_3c_3, \quad c_1c_3a_3c_3c_1, \quad c_3a_1c_3,$$
$$c_3c_2a_4c_2c_3, \quad c_2c_3a_3c_3c_2, \quad c_2c_3a_1c_3c_2, \quad c_1c_2c_3a_1c_3c_2c_1 \rangle.$$

Note that

$$G_2 < \mathcal{A}'' < G_1G_2G_1^{-1} < M\left(\mathrm{PGL}_2(\mathbb{Z}[\sqrt{-6}]) \rtimes \mathfrak{c}\right)M^{-1}, \quad M = \begin{pmatrix} \sqrt{-6} & 0 \\ 0 & 1 \end{pmatrix},$$

These 14 reflections correspond to the 14 faces of a cuboctahedron. The fundamental domain therefore consists of hyperbolic upper half 3-space minus 14 tangent geodesic hemispheres. This shows that $\mathcal{A}''$ is of infinite covolume but geometrically finite.

Let $\mathcal{A} = \mathcal{A}'' \cap \mathrm{PSL}_2(\mathcal{O}_K)$. The limit set of $\mathcal{A}''$ is shown in Figure 1. Since $[\mathcal{A}'' : \mathcal{A}]$ is finite, this limit set is the closure of a union of finitely many K-rational Möbius images of a single circle orbit; in this case, of $\mathcal{A}C$ where $C = \widehat{\mathbb{R}} + \sqrt{-6}$. Therefore we aim to demonstrate that $\mathcal{A}$ is an infinite-covolume, geometrically finite, Zariski dense, familial Kleinian group.

The geometric finiteness and infinite covolume are inherited by $\mathcal{A}$ from $\mathcal{A}''$, as it is finite index. By arguments exactly analogous to those in [32, Theorems 9.3-9.4], the limit sets of $\mathcal{A}''$ and $\mathcal{A}$ have Hausdorff dimension greater than 1 and are Zariski dense.

It simply remains to prove the following lemma.

Lemma 9.1. *The group G_2 is a congruence subgroup of* $\mathrm{PGL}_2(\mathbb{Z})$.

This implies $G_2 \cap \mathcal{A}$ is a congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$.

Proof of Lemma 9.1. We will show that G_2 contains the principal congruence subgroup $\Gamma(6)$. Let

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

These matrices generate $\mathrm{PSL}_2(\mathbb{Z})$. We will use the fact that $\Gamma(6)$ is the subgroup generated by the following elements [15, p. 1357]: $L^6, R^6, L^2R^3L^{-2}R^{-3}, L^3R^2L^{-3}R^{-2}$. It suffices now to verify

that

$$\begin{aligned}(a_1a_2)^{-1} &= L^6,\\ (a_1a_3)^{-6} &= R^6,\\ (a_1a_4)^{-1}(a_1a_3)^4 &= L^2R^3L^{-2}R^{-3},\\ (a_1a_2)^{-1}a_1a_4(a_1a_3)^2 &= L^3R^2L^{-3}R^{-2}.\end{aligned}$$

□

Finally, we apply Theorem 8.1. The potential bad primes are exactly $p = 2, 3$, since the curvatures of the packing are coprime and the congruence subgroup is of level 6. Letting $T_0 = a_3a_1 \in \mathcal{A}$ and $V = c_1a_3c_1a_1 \in \mathcal{A}$, and using the notation H, L, R for the basis of $\mathfrak{sl}(2,\mathbb{Z})$ as in the proof of Lemma 8.7, one can compute the following elements of $\mathcal{A}\cdot\mathfrak{sl}(2,\mathbb{Z})$:

$$VHV^{-1},\quad VLV^{-1},\quad VRV^{-1},\quad T_0VRV^{-1}T_0^{-1},\quad T_0^{-1}VRV^{-1}T_0.$$

These are enough to verify that $\iota_2 \leqslant 1$ and $\iota_3 = 0$. Therefore the modulus of the congruence obstruction divides 12. As experimental confirmation, computing curvatures $\leqslant 159$ appearing in the limit set packing (Figure 1), we find that the curvatures missing in this range are exactly those $\equiv 7, 9, 11 \pmod{12}$ as well as 13 and 16.

9.2 K-Apollonian packings

In this section we show that all K-Apollonian circle packings satisfy the hypotheses of Theorem 1.6. For an example of a K-Apollonian packing, see Figure 2.

The (strong) K-Apollonian groups defined in [32] are shown there to be finitely generated Zariski dense subgroups of $\mathrm{PSL}_2(\mathcal{O}_K)$ containing congruence subgroups (either $\Pi(2)$ or Γ^3 in the notation of [32, Section 10]). They are of infinite covolume since they are of infinite index, and each packing contains the horizontal line $\widehat{\mathbb{R}} + \sqrt{\Delta}/2$. Therefore all the hypotheses of Theorem 1.6 are satisfied save geometric finiteness. For that, it suffices to consider the remark following Theorem 1.6.

However, it may be useful to give an explicit description of a group associated to the packing. For each imaginary quadratic field K, we may use an adaptation of the weak K-Apollonian group given in [32, Theorem 9.2]:

$$\mathcal{A}' = \left\langle S = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},\quad T = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix},\quad V = \begin{pmatrix}-1 & \tau\\ 0 & 1\end{pmatrix}\right\rangle < \mathrm{PGL}_2(\mathcal{O}_K)$$

This group has the K-Apollonian packing as a limit set, and this limit set is of the form $\mathcal{A}'\widehat{\mathbb{R}} = \mathcal{A}'(\widehat{\mathbb{R}} + \sqrt{\Delta}/2)$. It has the following fundamental domain, given here as a list of the boundaries in $\widehat{\mathbb{C}}$ of its geodesic walls:

$$\begin{aligned}&A : \Re(z) = 0,\ \Im(z) \leqslant \Im(\tau)/2\\ &B : \Im(z) = \Im(\tau)/2,\ 0 \leqslant \Re(z) \leqslant 1\\ &C : \Re(z) = 1,\ \Im(z) \leqslant \Im(\tau)/2\\ &D : |z - 1/2| = 1/2.\end{aligned}$$

It is straightforward to verify that this region satisfies the Poincaré Polyhedron Theorem and is therefore a fundamental domain for $\mathcal{A}$; it is therefore geometrically finite and of infinite covolume. It has $\mathrm{PSL}_2(\mathbb{Z}) < \mathcal{A}$ (in the form of the first two generators above). It is Zariski dense by the same arguments as in [32, Section 10].

In order to apply Theorem 1.6, we need only pass to the finite-index subgroup $\mathcal{A} = \mathcal{A}' \cap \mathrm{PSL}_2(\mathcal{O}_K)$, by replacing V with

$$V_0 = VST^{-1}SV = \begin{pmatrix} \tau - 1 & -\tau^2 \\ 1 & -\tau - 1 \end{pmatrix}.$$

The curvatures of the K-Apollonian circle packings are primitive integral (after scaling by $\sqrt{-\Delta}$). Therefore, with this choice of group, Theorem 8.1 tells us immediately that the only primes of bad reduction for strong approximation are 2 and 3. In fact, it tells us more. Write L, R, H for the usual generators of $\mathfrak{sl}(2, \mathbb{Z})$ as in the proof of Lemma 8.7. Then following matrices are among $\mathcal{A} \cdot \mathfrak{sl}(2, \mathbb{Z})$:

$$V_0 R V_0^{-1}, \quad SV_0 R V_0^{-1} S, \quad TV_0 R V_0^{-1} T^{-1}, \quad STV_0 R V_0^{-1} T^{-1} S, \quad TSV_0 R V_0^{-1} ST^{-1}.$$

Using these suffices to verify that for $\Delta \equiv 0 \pmod 4$, $\iota_2 \leqslant 2$ and $\iota_3 = 0$; while for $\Delta \not\equiv 0 \pmod 4$, $\iota_2 \leqslant 1$ and $\iota_3 = 0$. Then Theorem 8.1 tells us that the modulus of the congruence obstruction for K-Apollonian packings is a divisor of 24 in all cases, and in fact a divisor of 12 if $\Delta \not\equiv 0 \pmod 4$. This is in accordance with [32, Conjecture 1.4], which gives an explicit prediction for the modulus for the congruence obstruction.

10. Notations

Table 1: Table of Notation used in Sections 2 through 7

$\mathcal{A}$	a familial Kleinian group in $\mathrm{PSL}_2(K)$, assumed from Section 3 onwards to be in $\mathrm{PSL}_2(\mathbb{Z}[\sqrt{-d}])$
$\mathcal{A}(q)$	elements of $\mathcal{A}$ congruent to identity modulo q
β	$\theta - \frac{r}{q}$; $\lvert\beta\rvert \leqslant \frac{K_0}{N}$
$B_q(n)$	$\frac{1}{[\mathcal{A}:\mathcal{A}(q)]} \sum_{\gamma_0 \in \mathcal{A}/\mathcal{A}(q)} c_q\left(\mathfrak{f}_{M\gamma_0}(Lx+1, Ly) - n\right)$
C	a circle tangent to the real line
$\widehat{\mathbb{C}}$	the extended complex plane
$c_q(n)$	$\sum_{r(q)}{}' e\left(\frac{rn}{q}\right)$
δ	hausdorff dimension of limit set of $\mathcal{A}$
Δ	discriminant of K
$\epsilon(n)$	0 if $n \equiv 1 \pmod 4$ and 1 if $n \equiv 3 \pmod 4$
$\mathfrak{d}_\gamma$	$2\frac{\Im(\overline{C_{M\cdot\gamma}} D_{M\cdot\gamma})}{\sqrt{-\Delta}}$ (i.e., the shift of the shifted form)
$e(x)$	$e^{2\pi i x}$
$e_q(x)$	$e^{\frac{2\pi i x}{q}}$
ϵ	small positive number
η	small positive number depending on M, $\mathcal{A}$, and C
$\mathcal{E}_N(n)$	minor arcs (error term) defined in (4.14)
$\mathcal{E}_N^U(n)$	modification of error term defined in (4.16)
$f \ll g$	$f = O(g)$
$f \asymp g$	$f \ll g$ and $g \ll f$
$\mathfrak{F}, \mathfrak{F}_T$	growing region in $\mathcal{A}$ defined in (4.3)
$\widehat{\mathfrak{f}}_{M\gamma}(a,c)$	shifted binary form $\sqrt{-\Delta}\lvert C_{M\cdot\gamma} a + D_{M\gamma} c\rvert^2 + 2\Im(\overline{C_{M\gamma}} D_{M\gamma})$
F_1, F_2, F_3	defined in (6.18) and (6.21)
γ	element of $\mathcal{A}$
h	$(\mathfrak{d}_\gamma, \mathfrak{d}_{\gamma'})$
$\mathcal{I}_1$	$\sum_{q<Q_0} \sum_{r(q)}{}' \int_{r/q-1/qJ}^{r/q+1/qJ} \lvert(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)\rvert^2 d\theta$
$\mathcal{I}_2$	$\sum_{Q_0 \leqslant q<X} \sum_{r(q)}{}' \int_{r/q-1/qJ}^{r/q+1/qJ} \lvert(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)\rvert^2 d\theta$
$\mathcal{I}_3$	$\sum_{X \leqslant q \leqslant J} \sum_{r(q)}{}' \int_{r/q-1/qJ}^{r/q+1/qJ} \lvert(1-\mathfrak{T}(\theta))\widehat{\mathcal{R}}_N^U(\theta)\rvert^2 d\theta$
$\mathcal{I}_Q$	$\sum_{Q<q\leqslant 2Q} \int_{-1/qJ}^{1/qJ} \sum_{r(q)}{}' \left\lvert\widehat{\mathcal{R}}_N^U\left(\frac{r}{q}+\beta\right)\right\rvert^2 d\beta$
$\mathcal{J}_\gamma(\beta; q, u, \xi, \zeta)$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y) e\left(\mathfrak{f}_{M\gamma}(Xx, Xy)\,\beta - \frac{X\xi}{quL}x - \frac{X\zeta}{quL}y\right) e\left(\frac{u^*\xi}{Lq}\right) dxdy$
$\Im$	imaginary part
J	$T^2 X$, depth of approximation; see (4.11)
K	$\mathbb{Q}(\sqrt{-d})$
$\kappa(\cdot)$	curvature of circle $\cdot$
$\mathcal{K}$	the set of curvatures in integral packing
$\mathcal{K}_a$	$\{n \in \mathbb{Z} \mid \forall q \in \mathbb{Z}, \exists k \in \mathcal{K}, \text{such that } n \equiv k (\mathrm{mod}\ q)\}$
$\mathcal{K}_a(N)$	$\mathcal{K}_a \cap [0, N]$
K_0	small power of N given in (5.1), depending on spectral gap

L	the level of the congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ contained in $\mathcal{A}$
L_0	positive integer such that $\mathcal{K}_a$ is union of some congruence classes mod L_0
$\lambda_\gamma(\beta, X, u, s, t)$	$\sum_{x,y\in\mathbb{Z}} \psi\left(\frac{Lux+uu^*}{X}\right) \psi\left(\frac{Luy}{X}\right) e\left(\mathfrak{f}_{M\gamma}\left(Lux+uu^*, Luy\right)\beta - sx - ty\right)$
M	Moebius transformation in $\mathrm{PSL}_2(K)$
$\mathcal{M}_N(n)$	major arcs (main term) defined in (4.13)
$\mathcal{M}_N^U(n)$	modification of main term defined in (4.15)
$\mathfrak{M}(n)$	$\frac{K_0}{N}\sum_{\gamma\in\mathfrak{F}} \hat{\mathfrak{t}}\left(\frac{K_0}{N}(\mathfrak{f}_{M\gamma}(Lx+1, Ly) - n)\right)$
N	a growing parameter; see Section 4
$\mathcal{O}_K$	ring of integers in K
p, p_i	prime numbers
$p^j \| n$	$p^j \mid n$ and $p^{j+1} \nmid n$
ψ	smooth function supported on $[1,2]$, with $\psi \geqslant 0$ and $\int_{\mathbb{R}} \psi(x)dx = 1$
P_{bad}	product of bad primes
q	positive integer
Q_0	small power of N given in (5.1), depending on spectral gap
$\widehat{\mathbb{R}}$	the extended real line, manifest as the horizontal axis in $\widehat{\mathbb{C}}$
$\Re$	real part
$\mathcal{R}_N(n)$	representation number of n in packing defined in (4.4)
$\mathcal{R}_N^U(n)$	modification of $\mathcal{R}_N(n)$ defined in (4.8)
$\widehat{\mathcal{R}}_N^U(\frac{r}{q}+\beta)$	$\sum_{u<U}\mu(u)\sum_{\gamma\in\mathfrak{F}_T}\mathcal{R}_{u,\gamma}(\frac{r}{q}+\beta)$
$\frac{r}{q}$	rational number of small denominator
$\sum_{r(q)}'$	sum over all $0 \leqslant r < q$ where $(r,q)=1$
$\mathfrak{S}_{Q_0}(n)$	$\sum_{q<Q_0}\frac{1}{[\mathcal{A}:\mathcal{A}(q)]}\sum_{\gamma_0\in\mathcal{A}/\mathcal{A}(q)} c_q\left(\mathfrak{f}_{M\cdot\gamma_0}(Lx+1, Ly) - n\right)$
$\mathfrak{S}(n)$	$\sum_{q=1}^{\infty} B_q(n)$
$S(q, A, B, C, D, E)$	$\sum_{x,y(q)} e(Ax^2 + Bxy + Cy^2 + Dx + Ey)$
$\mathcal{S}_\gamma(q, u, r, \xi, \zeta)$	$\frac{1}{q^2}\sum_{x_0,y_0(q)} e_q\left(r\mathfrak{f}_{M\gamma}\left(Lux_0+uu^*, Luy_0\right) + x_0\xi + y_0\zeta\right)$
$\mathcal{S}(q, u, \gamma, \xi, \zeta, \gamma', \xi', \zeta')$	$\sum_{r(q)}' \mathcal{S}_\gamma(q, u, r, \xi, \zeta)\overline{\mathcal{S}_{\gamma'}(q, u, r, \xi', \zeta')}$
T	$N^{1/200}$; see Section 4
T_1, T_2	growing parameters used to define $\mathfrak{F}_T$ in (4.3)
$\mathfrak{t}(x)$	$\max\{0, 1-\lvert x\rvert\}$, a hat function used in definition of major arcs
$\mathfrak{T}$	spike function in (4.12) used to define major arcs
$\tau_q(r)$	$\frac{1}{[\mathcal{A}:\mathcal{A}(q)]}\sum_{\gamma_0\in\mathcal{A}/\mathcal{A}(q)} \mathbf{1}\{\mathfrak{f}_{M\gamma_0}(Lx+1, Ly) = r\}$
θ	number in $[0,1]$
Θ	max of Θ_1, Θ_2 in Lemma 5.2 and Lemma 5.3 in context of $\mathcal{A}$
U	small power of N; see Section 4
u	positive number less than U
u^*	integer such that $uu^* \equiv 1(L)$
X	$N^{99/200}$; see Section 4
$\#\cdot$	cardinality of finite set $\cdot$
$\mathbf{1}\{\cdot\}$	characteristic function
$\lvert\lvert\cdot\rvert\rvert$	Frobenius norm
$(\cdot,\cdot)$	$\gcd(\cdot,\cdot)$

References

1 I. Agol. Tameness of hyperbolic 3-manifolds. arXiv:math/0405568.

2 C.J. Bishop and P.W. Jones. Hausdorff dimension and Kleinian groups. *Acta Math.*, 179(1):1–39, 1997.

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

4 J. Bourgain, A. Gamburd, and P. Sarnak. Generalization of Selberg's 3/16 theorem and affine sieve. *Acta Math.*, 207(2):255–290, 2011.

5 J. Bourgain and A. Kontorovich. On representations of integers in thin subgroups of $\mathrm{SL}_2(\mathbb{Z})$. *Geom. Funct. Anal.*, 20(5):1144–1174, 2010.

6 J. Bourgain and A. Kontorovich. On the local-global conjecture for integral Apollonian gaskets. *Invent. Math.*, 196(3):589–650, 2014. With an appendix by P. P. Varjú.

7 J. Bourgain and A. Kontorovich. On Zaremba's conjecture. *Annals of Math.*, 180:137–196, 2014.

8 J. Bourgain and P. Varjú. Expansion in $\mathrm{SL}_d(\mathbb{Z}/q\mathbb{Z})$, q arbitrary. *Ivent. Math.*, 188:151–173, 2012.

9 J.W.S. Cassels. *Rational Quadratic Forms.* Dover Publications, Inc, Mineola, New York, 1978.

10 H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. Revised and with a preface by H. L. Montgomery.

11 The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 7.3)*, 2016. `http://www.sagemath.org`.

12 E. Fuchs. Strong approximation in the Apollonian group. *J. Number Theory*, 131(12):2282–2302, 2011.

13 E. Fuchs and K. Sanden. Some experiments with Apollonian circle packings. *Exp. Math.*, 20(4):380–399, 2011.

14 R.L. Graham, J.C. Lagarias, C.L. Mallows, A.R. Wilks, and C.H. Yan. Apollonian circle packings: number theory. *J. Number Theory*, 100(1):1–45, 2003.

15 T. Hsu. Identifying congruence subgroups of the modular group. *Proc. Amer. Math. Soc.*, 124(5):1351–1359, 1996.

16 M. Kassabov, A. Lubotzky, and N. Nikolov. Finite simple groups as expanders. *Proc. Natl. Acad. Sci. USA*, 103(16):6116 – 6119, 2006.

17 Inkang Kim. Counting, mixing and equidistribution of horospheres in geometrically finite rank one locally symmetric manifolds. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2015(704):85–133, 2015.

18 A. Kontorovich and K. Nakamura. The superPAC: geometry to arithmetic of integral sphere packings. preprint, 2017.

19 E. Kowalski. Expander graphs (lecture notes). https://people.math.ethz.ch/ kowalski/expander-graphs.pdf.

20 P.D. Lax and R.S Phillips. The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces. *J. Funct. Anal.*, 46(3):280–350, 1982.

21 Michael Magee. Quantitative spectral gap for thin groups of hyperbolic isometries. *Journal of the European Mathematical Society*, 017(1):151–187, 2015.

22 Gregori A Margulis. *Discrete subgroups of semisimple Lie groups*, volume 17. Springer Science & Business Media, 1991.

23 A. Rapinchuk. Strong approximation for algebraic groups. In E. Breuillard and H. Oh, editors, *Thin groups and superstrong approximation*, volume 61 of *Math. Sci. Res. Inst. Publ.*, pages 269–288, Cambridge, 2014. Cambridge Univ. Press.

24 A. Salehi Golsefidy and X. Zhang. Inducing super-approximation. preprint.

25 P. Sarnak. The arithmetic and geometry of some hyperbolic three manifolds. *Acta Math.*, 151:253–295, 1983.

26 P. Sarnak. *Some Applications of Modular Forms*, volume 99 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1990.

27 P. Sarnak. Letter to J. Lagarias about integral Apollonian packings. http://web.math.princeton.edu/sarnak/AppolonianPackings.pdf, 2007.

28 A. Selberg. On discontinuous groups in higher-dimensional symmetric spaces. In *Contributions to function theory (internat. Colloq. Function Theory, Bombay, 1960)*, pages 147–164. Tata Institute of Fundamental Research, Bombay, 1960.

29 J-P. Serre. *Abelian ℓ-Adic Representations and Elliptic Curves*. The Advanced Book Program. Addison-Wesley Publishing Company, INC, New York, NY, 1989.

30 Y. Shalom. Bounded generation and Kazhdan's property (T). *Publ. Math. Inst. Hautes Études Sci.*, 90:145 – 168, 1999.

31 Katherine E. Stange. Visualizing the arithmetic of quadratic imaginary fields. To appear in Int. Math. Res. Not., 2017 `http://dx.doi.org/10.1093/imrn/rnx006`.

32 Katherine E. Stange. The Apollonian structure of Bianchi groups, 2015. To appear in Trans. Amer. Math. Soc. `http://arxiv.org/abs/1505.03121`.

33 D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.*, 153(3-4):259–277, 1984.

34 I. Vinogradov. Effective bisector estimate with application to Apollonian circle packings. *Int Math Res Notices*, 12:3217–3262, 2014.

35 X. Zhang. On representation of integers from thin subgroups of $\mathrm{SL}(2,\mathbb{Z})$ with parabolics. preprint, arXiv:1610.00770.

36 X. Zhang. *On the Local-global Principle for Integral Apollonian-3 Circle Packings*. ProQuest LLC, Ann Arbor, MI, 2014. Thesis (Ph.D.)–State University of New York at Stony Brook.

37 X. Zhang. On the local-global principle for integral apollonian 3-circle packings. *J. Reine Angew. Math.*, DOI: https://doi.org/10.1515/crelle-2015-0042, 2015.

Elena Fuchs efuchs@math.ucdavis.edu
Department of Mathematics, UC Davis, One Shields Avenue, Davis, CA 95616

Katherine E. Stange kstange@math.colorado.edu
Department of Mathematics, University of Colorado, Campus Box 395, Boulder, Colorado 80309-0395

Xin Zhang xz87@illinois.edu
Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801