

EXPONENTIAL MIXING VIA ADDITIVE COMBINATORICS

OSAMA KHALIL

ABSTRACT. We prove that the geodesic flow on a geometrically finite locally symmetric space of negative curvature is exponentially mixing with respect to the Bowen-Margulis-Sullivan measure. The approach is based on constructing a suitable anisotropic Banach space on which the infinitesimal generator of the flow admits an essential spectral gap. A key step in the proof involves estimating certain oscillatory integrals against the Patterson-Sullivan measure. For this purpose, we prove a general result of independent interest asserting that Fourier transforms of measures on \mathbb{R}^d , which do not concentrate near proper affine subspaces, enjoy polynomial decay outside of a sparse set of frequencies. As an intermediate step, we show that the L^q -dimension ($1 < q \leq \infty$) of iterated self-convolutions of such measures tend towards that of the ambient space. Our analysis also yields polynomial bounds on the Patterson-Sullivan mass of neighborhoods of certain proper subvarieties of the boundary at infinity which are saturated along the vertical foliation.

1. INTRODUCTION

1.1. Exponential mixing and Pollicott-Ruelle resonances. Let \mathcal{X} be the unit tangent bundle of a quotient of a real, complex, quaternionic, or a Cayley hyperbolic space by a discrete, geometrically finite, non-elementary group of isometries Γ . Denote by g_t the geodesic flow on \mathcal{X} and by m^{BMS} the Bowen-Margulis-Sullivan probability measure of maximal entropy for g_t . Let δ_Γ be the critical exponent of Γ . We refer the reader to Section 2 for definitions. The following is the main result of this article in its simplest form.

Theorem 1.1. *The geodesic flow on \mathcal{X} is exponentially mixing with respect to m^{BMS} . More precisely, there exists $\sigma_0 = \sigma_0(\mathcal{X}) > 0$ such that for all $f \in C_c^3(\mathcal{X})$, $g \in C_c^2(\mathcal{X})$ and $t \geq 0$,*

$$\int_{\mathcal{X}} f \circ g_t \cdot g \, dm^{\text{BMS}} = \int_{\mathcal{X}} f \, dm^{\text{BMS}} \int_{\mathcal{X}} g \, dm^{\text{BMS}} + \|f\|_{C^3} O_g(e^{-\sigma_0 t}).$$

The implicit constant depends on g through its C^2 -norm and the injectivity radius of its support.

The results also hold for functions with unbounded support and controlled growth in the cusp; cf. Section 8. Theorem 1.1 follows immediately from the following more precise result showing that the correlation function admits a finite resonance expansion.

Theorem 1.2. *There exists $\sigma > 0$ such that the following holds. There exist finitely many complex numbers $\lambda_1, \dots, \lambda_N$ with $-\sigma < \text{Re}(\lambda_i) < 0$, finite-rank projectors Π_i , and nilpotent matrices \mathcal{N}_i acting on the range of Π_i for each i , such that for all $f \in C_c^3(\mathcal{X})$ with $\int_{\mathcal{X}} f \, dm^{\text{BMS}} = 0$, $g \in C_c^2(\mathcal{X})$ and $t \geq 0$, we have*

$$\int_{\mathcal{X}} f \circ g_t \cdot g \, dm^{\text{BMS}} = \sum_{i=1}^N e^{t\lambda_i} \int_{\mathcal{X}} g \cdot e^{t\mathcal{N}_i} \Pi_i(f) \, dm^{\text{BMS}} + \|f\|_{C^3} O_g(e^{-\sigma t}).$$

The implicit constant depends on g through its C^2 -norm and the injectivity radius of its support.

Remark 1.3. The constant σ in Theorem 1.2 depends only on non-concentration parameters of Patterson-Sullivan (PS) measures near proper generalized sub-spheres of the boundary at infinity; cf. Corollary 7.3 for details. In particular, Theorem 1.2 implies that σ does not change if we replace Γ with a finite index subgroup. The interested reader is referred to [MN20, MN21] for

recent developments on a closely related problem yielding uniform resonance-free regions for the Laplacian operator on random covers of convex cocompact hyperbolic surfaces.

The “eigenvalues” λ_i above are known as *Pollicott-Ruelle resonances*. Theorem 1.1 follows from the above result by taking σ_0 to be the absolute value of the largest real part of the λ_i ’s. The reader is referred to Section 8 for a more precise discussion of the operators Π_i and \mathcal{N}_i .

Given two bounded functions f and g on \mathcal{X} , the associated correlation function is defined by

$$\rho_{f,g}(t) := \int_{\mathcal{X}} f \circ g_t \cdot g \, dm^{\text{BMS}}, \quad t \in \mathbb{R}.$$

Its (one-sided) Laplace transform is defined for any $z \in \mathbb{C}$ with positive real part $\text{Re}(z)$ as follows:

$$\hat{\rho}_{f,g}(z) := \int_0^\infty e^{-zt} \rho_{f,g}(t) \, dt.$$

Theorem 1.2 implies that, for suitably smooth f and g , $\hat{\rho}_{f,g}$ admits a meromorphic continuation to the half plane $\text{Re}(z) > -\sigma$ with the only possible poles occurring at $\{\lambda_i\}$.

Among the motivations for Theorem 1.2 is the closely related Jakobson-Naud conjecture asserting that the size of the essential spectral gap of the hyperbolic Laplacian operator for convex cocompact hyperbolic surfaces is exactly half the critical exponent [JN12].

1.2. L^q -flattening of measures on \mathbb{R}^d under convolution. The key new ingredient in our proof of Theorem 1.1 is the statement that conditional measures of m^{BMS} along the strong unstable foliation enjoy polynomial Fourier decay outside of a very sparse set of frequencies; cf. Corollary 1.7.

The key step in the proof is an L^q -flattening result for convolutions of measures on \mathbb{R}^d of independent interest. Roughly speaking, it states that the L^q -dimension (Def. 1.4) of a measure μ improves under iterated self-convolutions unless μ is concentrated near proper affine hyperplanes in \mathbb{R}^d at many scales. The proof of this result provided in Section 6 can be read independently of the rest of the article.

We formulate here a special case of our results under the following non-concentration condition and refer the reader to Definition 6.1 for a much weaker condition under which these results hold.

We need some notation before stating the result. Let \mathcal{D}_k denote the dyadic partition of \mathbb{R}^d by translates of the cube $2^{-k}[0,1]^d$ by $2^{-k}\mathbb{Z}^d$. We recall the notion of L^q -dimension of measures.

Definition 1.4. For $q > 1$, the L^q -dimension of a Borel probability measure μ on \mathbb{R}^d , denoted $\dim_q \mu$, is defined to be

$$\dim_q \mu := \liminf_{k \rightarrow \infty} \frac{-\log_2 \sum_{P \in \mathcal{D}_k} \mu(P)^q}{(q-1)k}.$$

The *Frostman exponent* of μ , denoted $\dim_\infty \mu$, is defined to be

$$\dim_\infty \mu := \liminf_{k \rightarrow \infty} \frac{\log_2 \max_{P \in \mathcal{D}_k} \mu(P)}{-k}.$$

We say that Borel measure μ on \mathbb{R}^d is *uniformly affinely non-concentrated* if for every $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ so that $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$ and for all $x \in \text{supp}(\mu)$, $0 < r \leq 1$, and every affine hyperplane $W < \mathbb{R}^d$, we have

$$\mu(W^{(\varepsilon r)} \cap B(x, r)) \leq \delta(\varepsilon) \mu(B(x, r)), \quad (1.1)$$

where $W^{(r)}$ and $B(x, r)$ denote the r -neighborhood of W and the r -ball around x respectively.

The following is our main result on flattening under convolution with non-concentrated measures.

Theorem 1.5. *Let $1 < q < \infty$ and $\eta > 0$ be given. Then, there exists $\varepsilon = \varepsilon(q, \eta) > 0$ such that if μ is any compactly supported Borel probability measure on \mathbb{R}^d which is uniformly affinely non-concentrated, then*

$$\dim_q(\mu * \nu) > \dim_q \nu + \varepsilon,$$

for every compactly supported probability measure ν on \mathbb{R}^d with $\dim_q \nu \leq d - \eta$.

In particular, $\dim_\infty \mu^{*n}$ converges to d at a rate depending only on the non-concentration parameters of μ , and, hence, the same holds for $\dim_q \mu^{*n}$ for all $q > 1$.

Remark 1.6. We refer the reader to Section 6 where a quantitative form of Theorem 1.5 is obtained under a much weaker *non-uniform* non-concentration condition; cf. Def. 6.1. This quantitative form is necessary for our applications and the weaker hypothesis is essential in the presence of cusps.

The L^2 -dimension case of Theorem 1.5 has the following immediate corollary asserting that the Fourier transform of affinely non-concentrated measures enjoys polynomial decay outside of a very sparse set of frequencies.

Corollary 1.7. *Let μ be as in Theorem 1.5 and $\hat{\mu}$ be its Fourier transform. Then, for every $\varepsilon > 0$, there is $\tau > 0$, depending only on the non-concentration parameters of μ , such that for all $T \geq 1$,*

$$|\{\|\xi\| \leq T : |\hat{\mu}(\xi)| > T^{-\tau}\}| \leq C_{\varepsilon, \mu} T^\varepsilon,$$

where $|\cdot|$ denotes the Lebesgue measure on \mathbb{R}^d , and $C_{\varepsilon, \mu} \geq 1$ is a constant depending on ε , the diameter of the support of μ , and its non-concentration parameters.

Remark 1.8. (1) A large class of dynamically defined measures, which includes self-conformal measures, is known to be affinely non-concentrated; cf. [RS20, Proposition 4.7 and Corollary 4.9] for measures on the real line and the results surveyed in [DFSU21, Section 1.3] for measures in higher dimensions under suitable irreducibility hypotheses¹. In particular, Theorem 1.5 applies to these measures generalizing prior known special cases for certain self-similar measures on \mathbb{R} by different methods; cf. [FL09, MS18].

(2) In [BY], it was observed that the proofs of Theorem 1.5, and its quantitative form Theorem 6.3, go through under the following weaker form of (1.1) allowing the ball on the right side to have a larger radius:

$$\mu(W^{(\varepsilon r)} \cap B(x, r)) \leq \delta(\varepsilon) \mu(B(x, cr)), \quad (1.2)$$

where $c \geq 1$ is a fixed constant. This property holds for instance for certain self-similar measures which do not satisfy (1.1), e.g. in the absence of separation conditions.

(3) Our proof in fact shows that Theorem 1.5 holds for *projections* of non-concentrated measures; cf. Theorem 6.3 and Corollary 6.4. Beyond the intrinsic interest in the study of projections of fractal measures, this stronger form is essential in our proof of exponential mixing outside the case of real hyperbolic spaces; cf. Section 1.5 for further discussion.

Corollary 1.7 generalizes the work of Kaufman [Kau84] and Tsujii [Tsu15] for self-similar measures on \mathbb{R} by different methods. Theorem 1.5 was obtained for measures on the real line by Rossi and Shmerkin in [RS20] under the uniform non-concentration hypothesis above. Their work builds crucially on a 1-dimensional inverse theorem due to Shmerkin in [Shm19] which was the key ingredient in his groundbreaking solution of Furstenberg's intersection conjecture. Proposition 6.14 can be regarded as a higher dimensional substitute for Shmerkin's inverse theorem. A similar higher dimensional inverse theorem for L^q -dimension was announced by Shmerkin in his ICM survey [Shm21, Section 3.8.3].

In Section 7, we show that Corollary 1.7 applies to PS measures when \mathcal{X} is real hyperbolic (and to certain projections of these measures in the other cases, see discussion in Section 1.5 below).

¹The results referenced in [DFSU21] require the open set condition, while [RS20] does not.

For convex cocompact hyperbolic surfaces, Bourgain and Dyatlov showed that PS measures in fact have polynomially decaying Fourier transform [BD17]. Their methods are different to ours and are based on Bourgain's sum-product estimates. Their result was extended to convex cocompact Schottky real hyperbolic 3-manifolds in [LNP21] by similar methods. These results imply Corollary 1.7 in these special cases, however Corollary 1.7 also applies to measures whose Fourier transform does not tend to 0 at infinity (e.g. the coin tossing measure on the middle 1/3 Cantor set). In forthcoming work, we apply our methods to generalize these results to hyperbolic manifolds of any dimension which are not necessarily of Schottky type.

1.3. Polynomial decay near proper subvarieties. Theorem 1.5 has the following important consequence regarding polynomial decay of the PS mass of neighborhoods of certain proper subvarieties of the boundary at infinity, which are saturated along the vertical foliation. This result is of independent interest. Denote by N^+ the expanding horospherical group associated to g_t for $t > 0$, the orbits of which give rise to the strong unstable foliation. Let N_r^+ be the r -ball around identity in N^+ (cf. Section 2.4 for definition of the metric on N^+). Let N_{ab}^+ denote the abelianization $N^+/[N^+, N^+]$. Finally, let $\Omega \subseteq \mathcal{X}$ be the non-wandering set for the geodesic flow; i.e. the closure of the set of its periodic orbits.

Theorem 1.9. *Let $x \in \Omega$. Then, there exist $C, \kappa > 0$ such that for all $\varepsilon > 0$ the following holds. Let $\mathcal{L} \subset N^+$ be the preimage of any proper affine subspace of the abelianization N_{ab}^+ and let $\mathcal{L}^{(\varepsilon)}$ be its ε -neighborhood. Then,*

$$\mu_x^u(\mathcal{L}^{(\varepsilon)} \cap N_1^+) \leq C\varepsilon^\kappa \mu_x^u(N_1^+).$$

The constants C and κ can be chosen to be uniform as x varies in any fixed compact set.

We refer the reader to Theorem 6.23 for a more general version of this result. Theorem 1.9 was obtained in [DFSU21, Lemma 3.8] in the case of real hyperbolic spaces by completely different methods. It is worth noting that our proof of exponential mixing only uses Theorem 1.9 in the case when the space is *not* real hyperbolic; cf. Remark 4.15 for further discussion.

1.4. Prior results. In the case Γ is convex cocompact, Theorem 1.1 is a special case of the results of [Sto11] which extend the arguments of Dolgopyat [Dol98] to Axiom A flows under certain assumptions on the regularity of the foliations and the holonomy maps. The special case of convex cocompact hyperbolic surfaces was treated in earlier work of Naud [Nau05]. The extension to frame flows on convex cocompact manifolds was treated in [SW20, CS22].

In the case of real hyperbolic manifolds with δ_Γ strictly greater than half the dimension of the boundary at infinity, Theorem 1.1 was obtained in [EO21], with much more precise and explicit estimates on the size of the essential spectral gap. The methods of [EO21] are unitary representation theoretic, building on the work of Lax and Phillips in [LP82], for which the restriction on the critical exponent is necessary. Earlier instances of the results of [EO21] under more stringent assumptions on the size of δ_Γ were obtained by Mohammadi and Oh in [MO15], albeit the latter results are stronger in that they in fact hold for the frame flow rather than the geodesic flow.

The case of real hyperbolic geometrically finite manifolds with cusps and arbitrary critical exponent was only recently resolved independently in [LP23] where a symbolic coding of the geodesic flow was constructed. This approach builds on extensions of Dolgopyat's method to suspension flows over shifts with infinitely many symbols; cf. [AM16, AGY06]. The extension of their result to frame flows was carried out in [LPS23].

Finally, we refer the reader to [DG16] and the references therein for a discussion of the history of the microlocal approach to the problem of spectral gaps via anisotropic Sobolev spaces.

1.5. Outline of the argument. For the convenience of the reader, we give a brief outline of the article. Section 2 recalls some basic facts about geometrically finite manifolds. We then recall several key results for our proof that were obtained in the prequel article [Kha23]. This includes Proposition 2.2 on uniform doubling for the conditional measures of m^{BMS} along the strong unstable foliation, and a Margulis inequality and its consequences on quantitative recurrence away from the cusp in Section 2.8. In Section 3, we recall from [Kha23] the construction of anisotropic Banach spaces arising as completions of spaces of smooth functions with respect to a dynamically relevant norm. We also a spectral gap result on resolvents of the geodesic flow in Theorem 3.3.

Theorem 3.3, roughly speaking, implies the existence of (a possibly non-converging) resonance expansion of the correlation function as in Theorem 1.2, where the real part of the resonances λ_i may a priori tend to 0 as $i \rightarrow \infty$. The bulk of the article is dedicated to ruling out this possibility.

To this end, a crucial Dolgopyat-type estimate, Theorem 4.2, is established in Section 4, which in particular rules out this accumulation of resonances on the imaginary axis. This estimate requires auxiliary technical results proved in Sections 5 and 7. Theorem 4.2 provides a contraction estimate on the norm of resolvents with large imaginary parts. A sketch of its proof is given in Section 4.1. Theorems 1.1 and 1.2 are deduced from this result in Section 8. The principle behind Theorem 4.2, due to Dolgopyat, is to exploit the non-joint integrability of the stable and unstable foliations via certain oscillatory integral estimates; cf. [Dol98, Liv04, GLP13, GPL22, BDL18].

A major difficulty in implementing this philosophy lies in estimating these oscillatory integrals against Patterson-Sullivan measures, which are *fractal* in nature in general. In particular, we cannot argue using the standard integration by parts method in previous works on exponential mixing of SRB measures using the method of anisotropic spaces, see for instance [Liv04, GLP13, GPL22, BDL18], where the unstable conditionals are of Lebesgue class.

We deal with this difficulty using Corollary 6.4 by taking advantage of the fact that the estimate in question is an *average* over oscillatory integrals. This idea is among the main contributions of this article. We hope this method can be fruitful in establishing rates of mixing of hyperbolic flows in greater generality.

In the case of variable curvature (i.e. when \mathcal{X} is not real hyperbolic), the action of the derivative of the geodesic flow on the strong unstable distribution is non-conformal which causes significant additional difficulties in the analysis, particularly in the presence of cusps. We deal with this difficulty by working with the *projection* of the unstable conditionals to the directions of slowest expansion and show that these projections also satisfy the conclusion of Corollary 6.4. See Remarks 4.15 and 4.16 for further discussion.

In Section 5, we obtain a linearization of the so-called temporal distance function. In Section 7, we verify the non-concentration hypotheses of Corollary 1.7 (more precisely, we verify the weaker hypothesis of Corollary 6.4) for the projection of the unstable conditionals of m^{BMS} onto the directions with weakest expansion. This allows us to apply Corollary 6.4 towards estimating the oscillatory integrals arising in Section 4.

Finally, Section 6 is dedicated to the proof of Theorem 1.5 and Corollary 1.7. Among the key ingredients in the proof are the asymmetric Balog-Szemerédi-Gowers Lemma due to Tao and Vu (Theorem 6.10) as well as Hochman’s inverse theorem for the entropy of convolutions (Theorem 6.12). Theorem 1.9 in its general form, Theorem 6.23, is deduced from these results in Section 6.10. This section can be read independently from the rest of the article.

Acknowledgements. The author thanks the Hausdorff Research Institute for Mathematics for its hospitality during the trimester program “Dynamics: Topology and Numbers” in Spring 2020 where part of this research was conducted. The author is partially supported by NSF grant number DMS-2247713. The author thanks Hee Oh, Peter Sarnak, and Pablo Shmerkin for helpful discussions regarding this project. The author also thanks Félix Lequen and the referees for detailed reading of the article and for numerous corrections and comments that significantly improved the exposition.

2. PRELIMINARIES

We recall here some background and definitions on geometrically finite manifolds.

2.1. Geometrically finite manifolds. The standard reference for the material in this section is [Bow93]. Let G be the group of orientation preserving isometries of a d -dimensional real, complex, quaternionic or Cayley hyperbolic space, denoted $\mathbb{H}_{\mathbb{K}}^d$, where $d \geq 2$ for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$, and $d = 2$ for $\mathbb{K} = \mathbb{O}$.

Fix a basepoint $o \in \mathbb{H}_{\mathbb{K}}^d$. Then, G acts transitively on $\mathbb{H}_{\mathbb{K}}^d$ and the stabilizer K of o is a maximal compact subgroup of G . We shall identify $\mathbb{H}_{\mathbb{K}}^d$ with $K \backslash G$. Denote by $A = \{g_t : t \in \mathbb{R}\}$ a 1-parameter subgroup of G inducing the geodesic flow on the unit tangent bundle of $\mathbb{H}_{\mathbb{K}}^d$. Let $M < K$ denote the centralizer of A inside K so that the unit tangent bundle $T^1\mathbb{H}_{\mathbb{K}}^d$ may be identified with $M \backslash G$.

Let $\Gamma < G$ be an infinite discrete subgroup of G . The limit set of Γ , denoted Λ_{Γ} , is the set of limit points of the orbit $\Gamma \cdot o$ on $\partial\mathbb{H}_{\mathbb{K}}^d$. Note that the discreteness of Γ implies that all such limit points belong to the boundary. Moreover, this definition is independent of the choice of o in view of the negative curvature of $\mathbb{H}_{\mathbb{K}}^d$. We often use Λ to denote Λ_{Γ} when Γ is understood from context. We say Γ is *non-elementary* if Λ_{Γ} is infinite.

The *hull* of Λ_{Γ} , denoted $\text{Hull}(\Lambda_{\Gamma})$, is the smallest convex subset of $\mathbb{H}_{\mathbb{K}}^d$ containing all the geodesics joining points in Λ_{Γ} . The convex core of the manifold $\mathbb{H}_{\mathbb{K}}^d/\Gamma$ is the smallest convex subset containing the image of $\text{Hull}(\Lambda_{\Gamma})$. We say $\mathbb{H}_{\mathbb{K}}^d/\Gamma$ is *geometrically finite* (resp. *convex cocompact*) if the closed 1-neighborhood of the convex core has finite volume (resp. is compact), cf. [Bow93]. The non-wandering set for the geodesic flow, denoted Ω , is the closure of the set of vectors in the unit tangent bundle whose orbit accumulates on itself.

Denote by N^+ the expanding horospherical subgroup of G associated to g_t , $t \geq 0$. Given $g \in G$, we denote by g^+ the coset of P^-g in the quotient $P^- \backslash G$, where $P^- = N^-AM$ is the stable parabolic group associated to $\{g_t : t \geq 0\}$. Similarly, g^- denotes the coset P^+g in $P^+ \backslash G$. Since M is contained in P^{\pm} , such a definition makes sense for vectors in the unit tangent bundle $M \backslash G$. Geometrically, for $v \in M \backslash G$, v^+ (resp. v^-) is the forward (resp. backward) endpoint of the geodesic determined by v on the boundary of $\mathbb{H}_{\mathbb{K}}^d$. Given $x \in G/\Gamma$, we say x^{\pm} belongs to Λ if the same holds for any representative of x in G ; this notion being well-defined since Λ is Γ invariant.

Notation. Throughout the remainder of the article, we fix a discrete non-elementary geometrically finite group Γ of isometries of some (irreducible) rank one symmetric space $\mathbb{H}_{\mathbb{K}}^d$ and denote by X the quotient G/Γ , where G is the isometry group of $\mathbb{H}_{\mathbb{K}}^d$.

2.2. Conformal Densities and the BMS Measure. The *critical exponent*, denoted δ_{Γ} , is defined to be the infimum over all real number $s \geq 0$ such that the Poincaré series $P_{\Gamma}(s, o) := \sum_{\gamma \in \Gamma} \exp(-s \text{dist}(o, \gamma \cdot o))$ converges. We shall simply write δ for δ_{Γ} when Γ is understood from context. The Busemann function is defined as follows: given $x, y \in \mathbb{H}_{\mathbb{K}}^d$ and $\xi \in \partial\mathbb{H}_{\mathbb{K}}^d$, let $\gamma : [0, \infty) \rightarrow \mathbb{H}_{\mathbb{K}}^d$ denote a geodesic ray terminating at ξ and define

$$\beta_{\xi}(x, y) = \lim_{t \rightarrow \infty} \text{dist}(x, \gamma(t)) - \text{dist}(y, \gamma(t)).$$

A Γ -invariant conformal density of dimension s is a collection of Radon measures $\{\nu_x : x \in \mathbb{H}_{\mathbb{K}}^d\}$ on the boundary satisfying

$$\gamma_* \nu_x = \nu_{\gamma x}, \quad \text{and} \quad \frac{d\nu_y}{d\nu_x}(\xi) = e^{s\beta_{\xi}(x, y)}, \quad \forall x, y \in \mathbb{H}_{\mathbb{K}}^d, \xi \in \partial\mathbb{H}_{\mathbb{K}}^d, \gamma \in \Gamma.$$

Given a pair of conformal densities $\{\mu_x\}$ and $\{\nu_x\}$ of dimensions s_1 and s_2 respectively, we can form a Γ invariant measure on $T^1\mathbb{H}_{\mathbb{K}}^d$, denoted by $m^{\mu, \nu}$ as follows: for $x = (\xi_1, \xi_2, t) \in T^1\mathbb{H}_{\mathbb{K}}^d$ in Hopf coordinates, we set

$$dm^{\mu, \nu}(\xi_1, \xi_2, t) = e^{s_1\beta_{\xi_1}(o, x) + s_2\beta_{\xi_2}(o, x)} d\mu_o(\xi_1) d\nu_o(\xi_2) dt. \quad (2.1)$$

The measure $m^{\mu, \nu}$ is quasi-invariant by the geodesic flow, and it is invariant if and only if $s_1 = s_2$.

When Γ is geometrically finite and $\mathbb{K} = \mathbb{R}$, Patterson [Pat76] and Sullivan [Sul79] showed the existence of a unique (up to scaling) Γ -invariant conformal density of dimension δ_Γ , denoted $\{\mu_x^{\text{PS}} : x \in \mathbb{H}_{\mathbb{R}}^d\}$. Geometric finiteness also implies that the measure $m^{\mu^{\text{PS}}, \mu^{\text{PS}}}$ descends to a finite measure of full support on Ω and is the unique measure of maximal entropy for the geodesic flow. This measure is called the Bowen-Margulis-Sullivan (BMS for short) measure and is denoted m^{BMS} .

Since the fibers of the projection from G/Γ to $T^1\mathbb{H}_{\mathbb{K}}^d/\Gamma$ are compact and parametrized by the group M , we can lift such a measure to G/Γ , also denoted m^{BMS} , by taking locally the product with the Haar probability measure on M . Since M commutes with the geodesic flow, this lift is invariant under the group A . By slight abuse of notation, we shall also use Ω to denote the lift of the non-wandering set to G/Γ . We refer the reader to [Rob03] and [PPS15] and references therein for details of the construction in much greater generality than that of $\mathbb{H}_{\mathbb{K}}^d$.

2.3. Stable and unstable foliations and leafwise measures. The fibers of the projection $G \rightarrow T^1\mathbb{H}_{\mathbb{K}}^d$ are given by the compact group M , which is the centralizer of A inside the maximal compact group K . In particular, we may lift m^{BMS} to a measure on G/Γ , also denoted m^{BMS} , and given locally by the product of m^{BMS} with the Haar probability measure on M . The leafwise measures of m^{BMS} on N^+ orbits are given as follows:

$$d\mu_x^u(n) = e^{\delta_\Gamma \beta_{(nx)^+}(o, nx)} d\mu_o^{\text{PS}}((nx)^+). \quad (2.2)$$

They satisfy the following equivariance properties for all $t \in \mathbb{R}$, $n \in N^+$ and $m \in M$:

$$\mu_{g_t x}^u = e^{\delta t} \text{Ad}(g_t)_* \mu_x^u, \quad (n)_* \mu_{nx}^u = \mu_x^u, \quad \mu_{mx}^u = \text{Ad}(m)_* \mu_x^u, \quad (2.3)$$

where $(n)_* \mu_{nx}^u$ is the pushforward of μ_{nx}^u under the map $u \mapsto un$ from N^+ to itself, and the last property follows since M normalizes N^+ and leaves m^{BMS} invariant.

2.4. Cygan metrics. We recall the definition of the Cygan metric on N^+ , denoted d_{N^+} . These metrics are right invariant under translation by N^+ , and satisfy the following convenient scaling property under conjugation by g_t . For all $r > 0$, if N_r^+ denotes the ball of radius r around identity in that metric and $t \in \mathbb{R}$, then

$$\text{Ad}(g_t)(N_r^+) = N_{e^t r}^+. \quad (2.4)$$

To define the metric, we need some notation which we use throughout the article. For $x \in \mathbb{K}$, denote by \bar{x} its \mathbb{K} -conjugate and by $|x| := \sqrt{\bar{x}x}$ its modulus. This modulus extends to a norm on \mathbb{K}^n by setting $\|u\|^2 := \sum_i |u_i|^2$, for $u = (u_1, \dots, u_n) \in \mathbb{K}^n$.

We let $\text{Im}\mathbb{K}$ denote those $x \in \mathbb{K}$ such that $\bar{x} = -x$. For example, $\text{Im}\mathbb{K}$ is the pure imaginary numbers and the subspace spanned by the quaternions i, j and k in the cases $\mathbb{K} = \mathbb{C}$ and $\mathbb{K} = \mathbb{H}$ respectively. For $u \in \mathbb{K}$, we write $\text{Re}(u) = (u + \bar{u})/2$ and $\text{Im}(u) = (u - \bar{u})/2$.

The Lie algebra \mathfrak{n}^+ of N^+ splits under $\text{Ad}(g_t)$ into eigenspaces as $\mathfrak{n}_\alpha^+ \oplus \mathfrak{n}_{2\alpha}^+$, where $\mathfrak{n}_{2\alpha}^+ = 0$ if and only if $\mathbb{K} = \mathbb{R}$. Moreover, we have the identification $\mathfrak{n}_\alpha^+ \cong \mathbb{K}^{d-1}$ and $\mathfrak{n}_{2\alpha}^+ \cong \text{Im}(\mathbb{K})$ as real vector spaces; cf. [Mos73, Section 19]. Given $(u, s) \in \mathfrak{n}_\alpha^+ \oplus \mathfrak{n}_{2\alpha}^+$, we define a quasi-norm on \mathfrak{n}^+ by $\|(u, s)\|' := \left(\|u\|^4 + |s|^2 \right)^{1/4}$. The distance of $n := \exp(u, s)$ to identity is then defined to be

$$d_{N^+}(n, \text{id}) := \|(u, s)\|'. \quad (2.5)$$

Given $n_1, n_2 \in N^+$, we set $d_{N^+}(n_1, n_2) = d_{N^+}(n_1 n_2^{-1}, \text{id})$.

2.5. Local stable holonomy. In this section, we recall the definition of (stable) holonomy maps. We give a simplified discussion of this topic which is sufficient in our homogeneous setting. Let $x = u^- y$ for some $y \in \Omega$ and $u^- \in N_2^-$. Since the product map $N^- \times A \times M \times N^+ \rightarrow G$ is

a diffeomorphism near identity, we can choose the norm on the Lie algebra so that the following holds. We can find maps $p^- : N_1^+ \rightarrow P^- = N^-AM$ and $u^+ : N_2^+ \rightarrow N^+$ so that

$$nu^- = p^-(n)u^+(n), \quad \forall n \in N_2^+. \quad (2.6)$$

Then, it follows by (2.2) that for all $n \in N_2^+$, we have $d\mu_y^u(u^+(n)) = e^{\delta\beta_{(nx)^+}(u^+(n)y, nx)} d\mu_x^u(n)$. Moreover, by further scaling the metrics if necessary, we can ensure that these maps are diffeomorphisms onto their images. In particular, writing $\Phi(nx) = u^+(n)y$, we obtain the following change of variables formula: for all $f \in C(N_2^+)$,

$$\int f(n) d\mu_x^u(n) = \int f((u^+)^{-1}(n)) e^{-\delta\beta_{\Phi^{-1}(ny)^+}(ny, \Phi^{-1}(ny))} d\mu_y^u(n). \quad (2.7)$$

Remark 2.1. To avoid cluttering the notation with auxiliary constants, we shall assume that the N^- component of $p^-(n)$ belongs to N_2^- for all $n \in N_2^+$ whenever u^- belongs to N_1^- .

2.6. Notational convention. Throughout the article, given two quantities A and B , we use the Vinogradov notation $A \ll B$ to mean that there exists a constant $C \geq 1$, possibly depending on Γ and the dimension of G , such that $|A| \leq CB$. In particular, this dependence on Γ is suppressed in all of our implicit constants, except when we wish to emphasize it. The dependence on Γ may include for instance the diameter of the complement of our choice of cusp neighborhoods inside Ω and the volume of the unit neighborhood of Ω . We write $A \ll_{x,y} B$ to indicate that the implicit constant depends on parameters x and y . We also write $A = O_x(B)$ to mean $A \ll_x B$.

2.7. Doubling Properties of Leafwise Measures. The goal of this section is to recall important doubling properties of the leafwise measures. Define the following exponents:

$$\begin{aligned} \Delta &:= \min \{ \delta, 2\delta - k_{\max}, k_{\min} \}, \\ \Delta_+ &:= \max \{ \delta, 2\delta - k_{\min}, k_{\max} \}. \end{aligned} \quad (2.8)$$

where k_{\max} and k_{\min} denote the maximal and minimal ranks of parabolic fixed points of Γ respectively. If Γ has no parabolic points, we set $k_{\max} = k_{\min} = \delta$, so that $\Delta = \Delta_+ = \delta$.

Proposition 2.2 ([Kha23, Prop. 3.1]). *For every $0 < \sigma \leq 5$, $x \in N_2^- \Omega$ and $0 < r \leq 1$, we have*

$$\mu_x^u(N_{\sigma r}^+) \ll \begin{cases} \sigma^\Delta \cdot \mu_x^u(N_r^+) & \forall 0 < \sigma \leq 1, 0 < r \leq 1, \\ \sigma^{\Delta_+} \cdot \mu_x^u(N_r^+) & \forall \sigma > 1, 0 < r \leq 5/\sigma. \end{cases}$$

2.8. Margulis Functions In Infinite Volume. We recall the construction of Margulis functions on Ω which allow us to obtain quantitative recurrence estimates to compact sets.

Theorem 2.3 ([Kha23, Theorem 4.1]). *Let $\Delta > 0$ denote the constant in (2.8). For every $0 < \beta < \Delta/2$, there exists a proper function $V_\beta : N_1^- \Omega \rightarrow \mathbb{R}_+$ such that the following holds. There is a constant $c \geq 1$ such that for all $x \in N_1^- \Omega$ and $t \geq 0$,*

$$\frac{1}{\mu_x^u(N_1^+)} \int_{N_1^+} V_\beta(g_t nx) d\mu_x^u(n) \leq ce^{-\beta t} V_\beta(x) + c.$$

The following proposition summarizes the main geometric properties of the functions V_β .

Proposition 2.4 ([Kha23, Prop. 4.2]). *The functions V_β in Theorem 2.3 satisfy the following:*

(1) *For every x in the unit neighborhood of Ω , we have that*

$$\text{inj}(x)^{-1} \ll_\Gamma V_\beta^{2/\beta}(x), \quad e^{\text{dist}(x,o)} \ll V_\beta(x)^{O_\beta(1)},$$

where $\text{inj}(x)$ denotes the injectivity radius at x and o is our fixed basepoint.

(2) *For all $g \in G$ and all $x \in X$, $\|g\|^{-\beta} V_\beta(x) \leq V_\beta(gx) \leq \|g^{-1}\|^\beta V_\beta(x)$.*

We also recall the following crucial exponential recurrence result.

Theorem 2.5 ([Kha23, Theorem 7.13]). *For every $\varepsilon > 0$, there exists $r_0 \asymp_\beta 1/\varepsilon$ such that the following holds for all $m \in \mathbb{N}, r \geq r_0, 0 < \theta < 1$ and $x \in N_1^- \Omega$. Let $H = e^{3\beta r_0}$, and let χ_H be the indicator function of the set $\{x : V(x) > H\}$. Then,*

$$\mu_x^u \left(n \in N_1^+ : \sum_{1 \leq \ell \leq m} \chi_H(g_{r\ell} n x) > \theta m \right) \leq e^{-(\beta\theta - \varepsilon)m} V(x) \mu_x^u(N_1^+).$$

3. ANISOTROPIC BANACH SPACES, TRANSFER OPERATORS, AND FRACTAL MOLLIFIERS

In this section, we define the Banach spaces on which the transfer operator and resolvent associated to the geodesic flow have good spectral properties. The transfer operator, denoted \mathcal{L}_t , acts on continuous functions as follows: for a continuous function f , let

$$\mathcal{L}_t f := f \circ g_t.$$

For $z \in \mathbb{C}$, the resolvent $R(z) : C_c(X) \rightarrow C(X)$ is defined formally as follows:

$$R(z)f := \int_0^\infty e^{-zt} \mathcal{L}_t f \, dt. \quad (3.1)$$

If Γ is not convex cocompact, we fix a choice of $\beta > 0$ so that Theorem 2.3 holds and set $V = V_\beta$. If Γ is convex cocompact, we take $V = V_\beta \equiv 1$ and we may take β as large as we like in this case. Note that the conclusion of Theorem 2.3 holds trivially with this choice of V . In particular, we shall use its conclusion throughout the argument regardless of whether Γ admits cusps.

Denote by $C_c^{k+1}(X)^M$ the subspace of $C_c^{k+1}(X)$ consisting of M -invariant functions, where M is the centralizer of the geodesic flow inside the maximal compact group K . In particular, $C_c^{k+1}(X)^M$ is naturally identified with the space of C_c^{k+1} functions on the unit tangent bundle of $\mathbb{H}_{\mathbb{K}}^d/\Gamma$.

3.1. Anisotropic Banach Spaces. We construct a Banach space of functions on X containing C^∞ functions and having the desired spectral properties for the action of \mathcal{L}_t and $R(z)$.

Given $r \in \mathbb{N}$, let \mathcal{V}_r^- denote the space of all C^r vector fields on N^+ pointing in the direction of the Lie algebra \mathfrak{n}^- of N^- and having norm at most 1. More precisely, \mathcal{V}_r^- consists of all C^r maps $v : N^+ \rightarrow \mathfrak{n}^-$, with C^r norm at most 1. Similarly, we denote by \mathcal{V}_r^0 the set of C^r vector fields $v : N^+ \rightarrow \mathfrak{a} := \text{Lie}(A)$, with C^r norm at most 1. Note that if $\omega \in \mathfrak{a}$ is the vector generating the flow g_t , i.e. $g_t = \exp(t\omega)$, then each $v \in \mathcal{V}_r^0$ is of the form $v(n) = \phi(n)\omega$, for some $\phi \in C^r(N^+)$ such that $\|\phi\|_{C^r(N^+)} \leq 1$. Define $\mathcal{V}_r = \mathcal{V}_r^- \cup \mathcal{V}_r^0$.

For $v \in \text{Lie}(G)$, denote by L_v the differential operator on $C^1(X)$ given by differentiation with respect to the vector field generated by v . Hence, for $\varphi \in C^1(G/\Gamma)$,

$$L_v \varphi(x) = \lim_{s \rightarrow 0} \frac{\varphi(\exp(sv)x) - \varphi(x)}{s}.$$

For each $k \in \mathbb{N}$, we define a norm on $C^k(N^+)$ as follows. Letting \mathcal{V}^+ be the unit ball in the Lie algebra of N^+ , $0 \leq \ell \leq k$, and $\phi \in C^k(N^+)$, we define $c_\ell(\phi)$ to be the supremum of $|L_{v_1} \cdots L_{v_\ell}(\phi)|$ over N^+ and all tuples $(v_1, \dots, v_\ell) \in (\mathcal{V}^+)^\ell$. We define $\|\phi\|_{C^k}$ to be $\sum_{\ell=0}^k c_\ell(\phi)/(2^\ell \ell!)$.

Following [GL06, GL08], we define a norm on $C_c^{k+1}(X)$ as follows. Given $f \in C_c^{k+1}(X)$, k, ℓ non-negative integers, $\gamma = (\gamma_1, \dots, \gamma_\ell) \in \mathcal{V}_{k+\ell}^\ell$ (i.e. ℓ tuple of $C^{k+\ell}$ vector fields) and $x \in X$, define

$$e_{k,\ell,\gamma}(f; x) := \frac{1}{V(x)} \sup \frac{1}{\mu_x^u(N_1^+)} \left| \int_{N_1^+} \phi(n) L_{\gamma_1(n)} \cdots L_{\gamma_\ell(n)}(f)(g_s n x) \, d\mu_x^u(n) \right|, \quad (3.2)$$

where the supremum is taken over all $s \in [0, 1]$ and all functions $\phi \in C^{k+\ell}(N_1^+)$ which are compactly supported in the interior of N_1^+ and having $\|\phi\|_{C^{k+\ell}(N_1^+)} \leq 1$.

For $\gamma \in \mathcal{V}_{k+\ell+1}^\ell$, we define $e'_{k,\ell,\gamma}(f; x)$ analogously to $e_{k,\ell,\gamma}(f; x)$, but where we take $s = 0$ and take the supremum over $\phi \in C^{k+\ell+1}(N_{1/10}^+)$ instead² of $C^{k+\ell}(N_1^+)$. Given $r > 0$, set

$$\Omega_r^- := N_r^- \Omega. \quad (3.3)$$

We define

$$e_{k,\ell,\gamma}(f) := \sup_{x \in \Omega_1^-} e_{k,\ell,\gamma}(f; x), \quad e_{k,\ell}(f) = \sup_{\gamma \in \mathcal{V}_{k+\ell}^\ell} e_{k,\ell,\gamma}(f). \quad (3.4)$$

Finally, we define $\|f\|_k$ and $\|f\|'_k$ by

$$\|f\|_k := \max_{0 \leq \ell \leq k} e_{k,\ell}(f), \quad \|f\|'_k := \max_{0 \leq \ell \leq k-1} \sup_{\gamma \in \mathcal{V}_{k+\ell+1}^\ell, x \in \Omega_{1/2}^-} e'_{k,\ell,\gamma}(f; x). \quad (3.5)$$

Note that the (semi-)norm $\|f\|'_k$ is weaker than $\|f\|_k$ since we are using more regular test functions and vector fields, and we are testing fewer derivatives of f .

It will also be convenient to introduce the following slightly stronger version of the norm $\|\cdot\|_k$; cf. Remark 3.11. Let $C^{0,1}(N_1^+)$ denote the space of Lipschitz functions on N^+ . We define coefficients $e_{1,0}^*(f)$ and $e_{1,1}^*(f)$, similarly to the coefficients $e_{1,0}$ and $e_{1,1}$ respectively in (3.2) and (3.4), but where, in both coefficients, the supremum is taken over all test functions $\phi \in C^{0,1}(N_1^+)$ with $\|\phi\|_{C^{0,1}} \leq 1$, instead of $C^1(N_1^+)$ and $C^2(N_1^+)$. Using these definitions, we introduce the following seminorm on $C_c^2(X)$:

$$\|f\|_1^* = e_{1,0}^*(f) + e_{1,1}^*(f). \quad (3.6)$$

Remark 3.1. Since the suprema in the definition of $\|\cdot\|_k$ are restricted to points on Ω_1^- , $\|\cdot\|_k$ defines a seminorm on $C_c^{k+1}(X)^M$. Moreover, since Ω_1^- is invariant by g_t for all $t \geq 0$, the kernel of this seminorm, denoted W_k , is invariant by \mathcal{L}_t . The seminorm $\|\cdot\|_k$ induces a norm on the quotient $C_c^{k+1}(X)^M / W_k$, which we continue to denote $\|\cdot\|_k$. A similar remark applies to $\|\cdot\|_1^*$.

Definition 3.2. We denote by \mathcal{B}_k the Banach space given by the completion of the quotient $C_c^{k+1}(X)^M / W_k$ with respect to the norm $\|\cdot\|_k$, where $C_c^{k+1}(X)^M$ denotes the subspace consisting of M -invariant functions. Similarly, we denote by \mathcal{B}_\star the completion of the quotient space $C_c^2(X)^M$ by the kernel of the seminorm $\|\cdot\|_1^*$ with respect to the induced norm on the quotient.

The following bound on the essential spectral radius of the resolvent is the main result of [Kha23].

Theorem 3.3 ([Kha23, Theorem 6.4]). *For all $z \in \mathbb{C}$ with $\operatorname{Re}(z) > 0$, the operator $R(z)$ defined in (3.1) extends to a bounded operator on \mathcal{B}_\star with spectral radius at most $1/\operatorname{Re}(z)$, and essential spectral radius $1/(\operatorname{Re}(z) + \sigma_0)$, for some $\sigma_0 > 0$ depending only on the critical exponent of Γ and the ranks of its cusps.*

Remark 3.4. Theorem 3.3 was obtained in [Kha23] for the norms $\|\cdot\|_k$, $k \geq 1$, however the proof extends readily to the norm $\|\cdot\|_1^*$ taking $\|\cdot\|'_1$ as its associated weak norm.

We also recall the following lemma on uniform boundedness of the semigroup $\{\mathcal{L}_t : t \geq 0\}$ on \mathcal{B}_k .

Lemma 3.5 ([Kha23, Lemma 7.1]). *For every $k, \ell \in \mathbb{N} \cup \{0\}$, $\gamma \in \mathcal{V}_{k+\ell}^\ell$, $t \geq 0$, and $x \in \Omega_1^-$,*

$$e_{k,\ell,\gamma}(\mathcal{L}_t f; x) \leq e^{-\varepsilon(\gamma)t} e_{k,\ell,\gamma}(f)(e^{-\beta t} + 1/V(x)),$$

where $\varepsilon(\gamma) \geq 0$ is the number of stable derivatives determined by γ . In particular, $\varepsilon(\gamma) = 0$ if and only if $\ell = 0$ or all components of γ point in the flow direction.

²The restriction on the supports allows us to handle non-smooth conditional measures; cf. proof of [Kha23, Prop. 6.6].

3.2. Fractal Mollifiers. In this section, we introduce certain mollification operators on smooth functions on X . These operators have the advantage that, roughly speaking, their Lipschitz norms are dominated by the norms introduced in (3.5). This property is very convenient in the estimates carried out in Section 4. The idea of using mollifiers to handle analogous steps is due to [BL12].

Fix a non-negative C^∞ bump function ψ supported inside $N_{1/2}^+$ and having value identically 1 on $N_{1/4}^+$. We also choose ψ to be symmetric and $\text{Ad}(M)$ -invariant, i.e.

$$\psi(n) = \psi(n^{-1}), \quad \psi(mnm^{-1}) = \psi(n), \quad \forall n \in N^+, m \in M. \quad (3.7)$$

Given $\varepsilon > 0$, define $\mathbb{M}_\varepsilon : C(X) \rightarrow C(X)$ be the operator defined by

$$\mathbb{M}_\varepsilon(f)(x) = \int \frac{\psi_\varepsilon(n)}{\int \psi_\varepsilon d\mu_x^u} f(nx) d\mu_x^u(n), \quad \psi_\varepsilon(n) = \psi(\text{Ad}(g_{-\log \varepsilon})(n)). \quad (3.8)$$

Note that ψ_ε is supported inside $N_{\varepsilon/2}^+$.

Remark 3.6. The condition that $\psi_\varepsilon(\text{id}) = \psi(\text{id}) = 1$ implies that for $x \in X$ with $x^+ \in \Lambda_\Gamma$,

$$\mu_x^u(\psi_\varepsilon) > 0, \quad \forall \varepsilon > 0. \quad (3.9)$$

In particular, since the conditional measures μ_x^u are supported on points nx with $(nx)^+ \in \Lambda_\Gamma$, the mollifier $\mathbb{M}_\varepsilon(f)$ is a well-defined function on all of X . That $\mathbb{M}_\varepsilon(f)$ is continuous follows by continuity of the map $x \mapsto \mu_x^u$ in the weak-* topology; cf. [Rob03, Lemme 1.16].

Remark 3.7. We note that $\text{Ad}(M)$ -invariance of ψ_ε and the conditional measures μ_x^u (cf. (2.3)) implies that $\mathbb{M}_\varepsilon(f)$ is M -invariant whenever f is.

The first result asserts that $\mathbb{M}_\varepsilon(f)$ is a good approximation of f .

Proposition 3.8. *For all $0 < \varepsilon \leq 1/10$, and $t \geq 1$, we have*

$$e_{1,0}^\star(\mathcal{L}_t(f - \mathbb{M}_\varepsilon(f))) \ll (\varepsilon + 1)e^{-t}e_{1,0}^\star(f).$$

In light of this statement, we will in fact only use \mathbb{M}_ε with $\varepsilon = 1/10$. However, for clarity, we state and prove the remaining results for a general value of ε .

The following results estimate the regularity of mollifiers. Recall the constant $\Delta_+ \geq 0$ in (2.8). The first result is an estimate of L^∞ type.

Proposition 3.9. *For all $0 < \varepsilon \leq 1$ and $x \in N_1^-\Omega$, we have*

$$|\mathbb{M}_\varepsilon(f)(x)| \ll \varepsilon^{-\Delta_+-1}e_{1,0}^\star(f)V(x).$$

Finally, we need the following Lipschitz estimate on mollifiers along the stable direction. Recall the stable parabolic group $P^- = N^-AM$ parametrizing the weak stable manifolds of g_t .

Proposition 3.10. *For all $0 < \varepsilon \leq 1/10$, $p^- \in P^-$, and $x \in X$ so that x belongs to $N_{3/4}^-\Omega$ and p^- is of the form u^-g_tm for $u^- \in N_{1/10}^-$, $|t| \leq 1/10$ and $m \in M$, we have that*

$$|\mathbb{M}_\varepsilon(f)(p^-x) - \mathbb{M}_\varepsilon(f)(x)| \ll \text{dist}(p^-, \text{id})\varepsilon^{-\Delta_+-2} \cdot \|f\|_1^\star V(x).$$

Remark 3.11. The purpose of introducing the norms $\|\cdot\|_1^\star$ is to simplify arguments related to the regularity of the function $n \mapsto \psi_\varepsilon(n)/\mu_{nx}^u(\psi_\varepsilon)$ in the proofs of the above results.

Propositions 3.8–3.10 are standard in the case of smooth mollifiers, but some care is required in our case due to the fractal nature and (possible) non-compactness of Ω . This is in part the reason for the non-standard shape of the chosen mollifier. Nonetheless, we omit the straightforward proofs of the above results, and refer the interested reader to the arXiv version of the article for details.

4. SPECTRAL GAP FOR RESOLVENTS WITH LARGE IMAGINARY PARTS

In this section, we establish the key estimate in the proof of Theorems 1.1 and 1.2. Theorem 3.3 implies the half plane $\{\operatorname{Re}(z) > -\sigma_0\}$, for a suitable $\sigma_0 > 0$, contains at most countably many isolated eigenvalues for the generator of \mathcal{L}_t . To show exponential mixing, it remains to rule out accumulation of these eigenvalues on the imaginary axis as their imaginary parts tend to ∞ .

Remark 4.1. Throughout the rest of this section, if X has cusps, we require the Margulis function $V = V_\beta$ in the definition of all the norms we use to have

$$\beta = \Delta/4 \quad (4.1)$$

in the notation of Theorem 2.3. In particular, the contraction estimate in Theorem 2.3 holds with V^p in place of V for all $1 \leq p \leq 2$. Recall that the constant Δ is given in (2.8).

We define for $B \neq 0$ an equivalent norm to $\|\cdot\|_1^*$ defined in (3.6) as follows:

$$\|f\|_{1,B}^* := e_{1,0}^*(f) + \frac{e_{1,1}^*(f)}{B}. \quad (4.2)$$

The following result is one of the main technical contributions of this article.

Theorem 4.2. *There exist constants $b_\star \geq 1$, and $\varkappa, a_\star, \sigma_\star > 0$, such that the following holds. For all $z = a_\star + ib \in \mathbb{C}$ with $|b| \geq b_\star$ and for $m = \lceil \log |b| \rceil$, we have that*

$$e_{1,0}^*(R(z)^m f) \leq C_\Gamma \frac{\|f\|_{1,B}^*}{(a_\star + \sigma_\star)^m},$$

where $C_\Gamma \geq 1$ is a constant depending only on the fundamental group Γ and $B = |b|^{1+\varkappa}$.

Remark 4.3. The constants \varkappa, a_\star , and σ_\star depend only on non-concentration parameters of the Patterson-Sullivan measure near proper subvarieties of the boundary at infinity; cf. Definition 6.1 for the precise definition of non-concentration and Corollary 7.3 where this non-concentration is established. This non-concentration property is used to apply the results of Section 6 in the proof of Prop. 4.14 and Theorem 4.17, which are the key steps in the proof of Theorem 4.2.

4.1. Sketch of the proof. We begin with a rough sketch of the proof of Theorem 4.2. To describe the main new idea based on additive combinatorics concisely, we will ignore many technical difficulties, including those posed by the presence of cusps.

Fix $z = a + ib$, where $a > 0$ is suitably small and $|b|$ is sufficiently large. Let $m = \lceil \log |b| \rceil$. The integrals we wish to estimate take the form

$$\int_0^\infty \frac{t^{m-1}}{(m-1)!} e^{-zt} \int_{N_1^+} \phi(n) \mathcal{L}_t f(nx) d\mu_x^u dt.$$

Using convergence of the integral defining $R(z)^m$, we will trivially estimate over the parts of the integral where t is very large and relatively small. The bulk of the work lies in finding $\sigma > 0$ so that the following bound holds in the range $T \asymp \log |b|$:

$$\left| \int_T^{T+1} e^{-zt} \int_{N_1^+} \phi(n) f(g_t nx) d\mu_x^u dt \right| \ll e^{-aT} \|f\| |b|^{-\sigma}.$$

The key to the proof is to exploit the oscillations of the phase function e^{ibt} and (average) Fourier decay properties of the measures μ_x^u obtained in Section 6. To do this, we put the integral in a form where this phase is integrated against μ_x^u . This is achieved using non-joint integrability of the stable and unstable foliations (i.e. that N^+ and N^- do not commute). To this end, we partition the

space into flow boxes B_ρ , apply the geodesic flow by amount T , and group pieces of the expanded unstable manifold according to which flow box they land in to get:

$$e^{-\delta T} \sum_{\text{flow boxes } B_\rho} \sum_{\substack{\text{conn. comps. of} \\ g_T N_1^+ x \cap B_\rho}} \int_0^1 e^{-zt} \int_{\text{conn. component}(x_{\rho,\ell})} \phi \circ g_{-T} \cdot f(g_t n x_{\rho,\ell}) d\mu_{x_{\rho,\ell}}^u(n) dt.$$

Here, the points $\{x_{\rho,\ell} : \ell\}$ are transverse intersection points of the expanded unstable manifolds $g_T N_1^+ x$ with a fixed transversal to the unstable foliation inside the flow box B_ρ .

Since the derivatives of $\phi \circ g_{-T}$ along N^+ are $O(e^{-T})$, these functions are nearly constant along each connected component and we can ignore them in the sequel.

Fix a box B_ρ and a reference point $y_\rho \in B_\rho$. We view the above integrals as taking place on the *weak unstable* manifold of $x_{\rho,\ell}$ for each ℓ . We change variables using local strong stable holonomy so that all the integrals are taking place along the local weak unstable manifold of y_ρ :

$$\sum_\ell \int_0^1 \int_{N_1^+} e^{-ib(t-\tau_\ell(n))} d\mu_{y_\rho}^u dt,$$

where we ignored the Jacobian of the change of variables for simplicity.

The functions $\tau_\ell(n)$ are known as the *temporal distance functions* in the literature. Roughly, they are defined dynamically as follows: for every $(t, n) \in (0, 1) \times N_1^+$, there exists a unique pair $(\tau_\ell(n), u_\ell(t, n)) \in (0, 1) \times N_1^+$ such that the point $g_{t-\tau_\ell(n)} u_\ell(t, n) \cdot y_\rho$ is on the same strong stable leaf of $g_t n x$. Now, non-joint integrability of the strong stable and unstable foliations imply that the derivative of $\tau_\ell(n)$ is uniformly bounded away from 0. Variants of this property are crucial in carrying out analysis of oscillatory integrals.

To proceed, we partition the domain of integration into small balls $\{A_j\}$ so we approximate τ_ℓ by its linearization. Fix one such ball and apply a suitable amount of geodesic flow to scale this ball to be a ball of radius 1 to get

$$\sum_\ell \int_{N_1^+} e^{-ib\langle v_{\ell,j}, n \rangle} d\mu_{z_j}^u(n), \quad (4.3)$$

where, roughly speaking, $v_{\ell,j}$ is the derivative of τ_ℓ at the center of the ball A_j .

Up to this point, the argument is very similar to that appearing in Liverani's original work [Liv04] and its subsequent generalizations, e.g. [BL12, BDL18, GLP13, GPL22]. The crucial difference comes at the next step. In these previous works, the measure $\mu_{z_i}^u$ was always absolutely continuous to Lebesgue, and proof proceeds by integration by parts.

In our case, these measures have fractional Hausdorff dimension in general. The novelty of our approach is to note that the sum in (4.3), when properly normalized, is an *average* over Fourier coefficients of the measure $\mu_{z_i}^u$. This makes it amenable to our flattening results (Cor. 6.4) which establish verifiable criteria under which measures enjoy polynomial Fourier decay outside of an arbitrarily sparse set of frequencies.

4.2. Proof of Theorem 4.2. The remainder of this section is dedicated to the proof of Theorem 4.2. Let $a \in (0, 1]$ to be chosen sufficiently small; cf. (4.60). We assume that $z = a + ib$ with $b > 0$, the other case being identical.

Time partition. Let $p : \mathbb{R} \rightarrow [0, 1]$ be a smooth bump function supported in $(-1, 1)$ such that

$$\sum_{j \in \mathbb{Z}} p(t - j) = 1, \quad \forall t \in \mathbb{R}. \quad (4.4)$$

Let $T_0 > 0$ be a parameter to be chosen large depending only on Γ and let

$$m = \lceil \log b \rceil. \quad (4.5)$$

By induction on $n \in \mathbb{N}$, we have for all $z \in \mathbb{C}$ with $\operatorname{Re}(z) > 0$ that

$$R(z)^n = \int_0^\infty \frac{t^{n-1}}{(n-1)!} e^{-zt} \mathcal{L}_t dt, \quad \left| \int_0^\infty \frac{t^{n-1}}{(n-1)!} e^{-zt} dt \right| \leq 1/\operatorname{Re}(z)^n. \quad (4.6)$$

Changing variables, we obtain

$$\begin{aligned} R(z)^m &= \int_0^\infty \frac{t^{m-1} e^{-zt}}{(m-1)!} \mathcal{L}_t dt \\ &= \int_0^\infty \frac{t^{m-1} e^{-zt}}{(m-1)!} p(t/T_0) \mathcal{L}_t dt + \sum_{j=0}^\infty \frac{((j+2)T_0)^{m-1} e^{-zjT_0}}{(m-1)!} \int_{\mathbb{R}} p_j(t) e^{-zt} \mathcal{L}_{t+jT_0} dt, \end{aligned} \quad (4.7)$$

where we define p_j as follows:

$$p_j(t) := \left(\frac{jT_0 + t}{(j+2)T_0} \right)^{m-1} p\left(\frac{t - T_0}{T_0} \right). \quad (4.8)$$

Note that p_j is supported in the interval $(0, 2T_0)$ for all $j \geq 0$.

Contribution of very small and very large times. We will estimate the contribution of each term in the sum over j in (4.7) individually. We will restrict our attention to values of j of size $\asymp \log b$. We begin by estimating the first term in (4.7) trivially. Since $p(t/T_0)$ is supported in $(-T_0, T_0)$ and $a \leq 1$, by taking b large enough and using the triangle inequality for the seminorm $e_{1,0}^*$ and Lemma 3.5, we obtain

$$e_{1,0}^* \left(\int_0^\infty \frac{t^{m-1} e^{-zt}}{(m-1)!} p(t/T_0) \mathcal{L}_t f dt \right) \ll e_{1,0}^*(f) T_0^{m-1} / (m-1)! \ll \frac{e_{1,0}^*(f)}{(a+1)^m}. \quad (4.9)$$

Next, we let

$$\eta_1 = 4/3, \quad \eta_2 = 2/a. \quad (4.10)$$

Note that since $a \leq 1$, we have $\eta_1 < \eta_2$. We wish to find a trivial bound on the terms corresponding to $j \notin [\eta_2, \eta_2]m/T_0$. First, we have the following bound on the sum of the terms $j \leq \eta_1 m/T_0$.

$$\sum_{j: jT_0 < \eta_1 m - 2T_0} \frac{((j+2)T_0)^{m-1} e^{-ajT_0}}{(m-1)!} \int_{\mathbb{R}} p_j(t) e^{-at} e_{1,0}^*(\mathcal{L}_{t+jT_0} f) dt \ll e_{1,0}^*(f) \int_0^{\eta_1 m} \frac{t^{m-1} e^{-at}}{(m-1)!} dt. \quad (4.11)$$

Similarly, we have the following bound on the tail of the sum:

$$\sum_{j: jT_0 > \eta_2 m} \frac{((j+2)T_0)^{m-1} e^{-ajT_0}}{(m-1)!} \int_{\mathbb{R}} p_j(t) e^{-at} e_{1,0}^*(\mathcal{L}_{t+jT_0} f) dt \ll e_{1,0}^*(f) \int_{\eta_2 m}^\infty \frac{t^{m-1} e^{-at}}{(m-1)!} dt \quad (4.12)$$

The following lemma provides the desired bound on the integrals appearing in the above bounds.

Lemma 4.4. *Suppose that $a\eta > 1$. Then, there exists $\theta \in (0, 1)$ such that*

$$\int_{\eta m}^\infty \frac{t^{m-1} e^{-at}}{(m-1)!} dt \ll_{a,\eta} (\theta/a)^m.$$

On the other hand, if $a\eta < 1/e$, then there exists $\theta \in (0, 1)$ such that

$$\int_0^{\eta m} \frac{t^{m-1} e^{-at}}{(m-1)!} dt \ll (\theta/a)^m.$$

In both cases, we may take $\theta = a\eta e^{1-a\eta}$.

Proof. Integration by parts and induction on m yield

$$\int_{\eta m}^{\infty} \frac{t^{m-1} e^{-at}}{(m-1)!} dt = \frac{e^{-a\eta m}}{a^m} \sum_{k=0}^{m-1} \frac{(a\eta m)^k}{k!} = \frac{e^{-a\eta m} (a\eta m)^m}{a^m m!} \sum_{k=0}^{m-1} \frac{m \cdots (k+1)}{(a\eta m)^{m-k}}. \quad (4.13)$$

Note that the k^{th} term of the latter sum is at most $(a\eta)^{-m+k}$. Moreover, from Stirling's formula, we have that $m! \gg m^{m+1/2} e^{-m}$. Hence, when $a\eta > 1$, we get

$$\int_{\eta m}^{\infty} \frac{t^{m-1} e^{-at}}{(m-1)!} dt \ll \frac{e^{(1-a\eta)m} (a\eta)^m}{a^m}.$$

Taking $\theta = a\eta e^{1-a\eta}$ and noting that xe^{1-x} is strictly less than 1 for all $x \geq 0$ with $x \neq 1$, concludes the proof of the first assertion. For the second assertion, assume that $a\eta e < 1$. Then, combining (4.6) with (4.13), we get

$$\int_0^{\eta m} \frac{t^{m-1} e^{-at}}{(m-1)!} dt = \frac{e^{-a\eta m}}{a^m} \sum_{k=m}^{\infty} \frac{(a\eta m)^k}{k!}.$$

Stirling's formula shows that the k^{th} -term of the above sum is $O((a\eta e)^k)$, for all $k \geq m$. Since $a\eta e < 1$, we get that the integral is $\ll (\theta/a)^m$ for $\theta = a\eta e^{1-a\eta}$. \square

In view of (4.9), (4.11), (4.12), and Lemma 4.4, in what follows, we restrict our attention to the terms where j satisfies $-2T_0 + 4m/3 \leq jT_0 \leq 2m/a$. We shall assume b is large enough so that $4m/3 - 2T_0 \geq 2m/3$. In particular, we estimate the terms satisfying

$$2m/3 \leq jT_0 \leq 2m/a. \quad (4.14)$$

Finally, in view of (4.6), we have that

$$\sum_{j: jT_0 \geq 2m/3} \frac{((j+2)T_0)^{m-1} e^{-ajT_0}}{(m-1)!} \leq e^{2aT_0} (1 + 6/m)^m \int_0^{\infty} \frac{t^{m-1} e^{-at}}{(m-1)!} dt \ll \frac{e^{2aT_0}}{a^m}, \quad (4.15)$$

where we used the bound $(1 + 6/m)^m \ll 1$ for all large enough m .

Contribution of points in the cusp. We estimate the contribution of each term in the sum over j in (4.7) individually. We begin by reducing to the case where the basepoint has bounded height. Let $\alpha \geq 0$ be a small parameter to be chosen at the end of the argument and satisfies

$$\alpha \leq a/40. \quad (4.16)$$

Let $x \in N_1^- \Omega$ be arbitrary. Suppose that $V(x) > e^{\beta\alpha jT_0}$. Then, Lemma 3.5 implies that

$$e_{1,0}^*(\mathcal{L}_{t+jT_0} f; x) \ll_{\beta} e^{-\beta\alpha jT_0} e_{1,0}^*(f).$$

In light of (4.15), summing the above errors over j , we obtain an error term of the form

$$e^{2aT_0} e_{1,0}^*(f) \frac{1}{(a + \beta\alpha)^m}. \quad (4.17)$$

Thus, we may assume for the remainder of the section that

$$V(x) \leq e^{\beta\alpha jT_0/2}. \quad (4.18)$$

Approximation with mollifiers. To begin our estimates, fix a suitable test function ϕ for $e_{1,0}^*$. In particular, ϕ has $C^{0,1}(N^+)$ norm at most 1. The integrals we wish to estimate take the form

$$\begin{aligned} \int_{N_1^+} \phi(n) \int_{\mathbb{R}} p_j(t) e^{-zt} \mathcal{L}_{t+jT_0}(f)(g_s nx) dt d\mu_x^u(n) \\ = \int_{\mathbb{R}} e^{-zt} \int_{N_1^+} p_j(t) \phi(n) f(g_{s+t+jT_0} nx) d\mu_x^u(n) dt, \end{aligned} \quad (4.19)$$

for all $s \in [0, 1]$. We again only provide the estimate in the case $s = 0$ to simplify notation, the general case being essentially identical.

Recall that p_j is supported in the interval $(0, 2T_0)$. In particular, the extra t in \mathcal{L}_{t+jT_0} could be rather large, which will ruin certain trivial estimates later. To remedy this, recall the partition of unity of \mathbb{R} given in (4.4) and set

$$p_{j,w}(t) := p_j(t+w)p(t), \quad \forall w \in \mathbb{Z}. \quad (4.20)$$

Using a change of variable, we obtain

$$(4.19) = \sum_{w \in \mathbb{Z}} e^{-zw} \int_{\mathbb{R}} e^{-zt} \int_{N_1^+} p_{j,w}(t) \phi(n) f(g_{t+w+jT_0} nx) d\mu_x^u(n) dt. \quad (4.21)$$

Note the above sum is supported on $0 \leq w \ll T_0$, and the support of each integral in t is now $(-1, 1)$. For the remainder of the section, we fix some $w \in \mathbb{Z}$ in that support.

Let $\mathbb{M} := \mathbb{M}_{1/10}$, where for $\varepsilon > 0$, \mathbb{M}_ε denotes the mollifier defined in Section 3.2. To simplify notation, we set

$$g_j^w := g_{w+jT_0}, \quad F := \mathbb{M}(f). \quad (4.22)$$

Since $\phi \in C^{0,1}(N_1^+)$ with $\|\phi\|_{C^{0,1}} \leq 1$, it follows by Proposition 3.8 that

$$\left| \int_{N_1^+} \phi(n) \mathcal{L}_t(f - F)(g_j^w nx) d\mu_x^u(n) \right| \ll e^{-(t+w+jT_0)} e_{1,0}^*(f) V(x) \mu_x^u(N_1^+).$$

Arguing as in (4.17), summing the above errors over j , we get an error term of the form

$$O_{T_0} \left(\frac{e_{1,0}^*(f) V(x) \mu_x^u(N_1^+)}{(a+1)^m} \right). \quad (4.23)$$

Hence, we may replace f with F in (4.21). We will frequently use the following observation. Writing $F = F - f + f$ and using Proposition 3.8, we have that

$$e_{1,0}^*(F) \ll e_{1,0}^*(f). \quad (4.24)$$

Partitions of unity and flow boxes. We begin by finding convenient partitions of the space by *flow boxes*, i.e., sets of the form $P_r^- N_s^+ \cdot x$ for $r, s > 0$ and $x \in G/\Gamma$ and such that the map $P_r^- N_s^+ \ni g \mapsto gx$ is injective. To this end, we have to restrict our attention to the part of the space where the injectivity radius is bounded away from 0. Define

$$K_j := \left\{ y \in X : V(y) \leq e^{(2\beta\alpha j + 3\beta)T_0} \right\}, \quad \iota_j := \min \{1/10, \text{inj}(K_j)\}, \quad \iota_b := b^{-2/3}. \quad (4.25)$$

We note that Proposition 2.4 implies that

$$\iota_j^{-1} \ll e^{(4\alpha j + 6)T_0}. \quad (4.26)$$

Remark 4.5. Since we are working in the regime where α is small and j is bounded linearly in $\log b$, cf. (4.14), (4.5), and (4.16), the bound (4.26) implies that ι_b is much smaller than ι_j in general.

The following lemma provides an efficient cover of $K_j \cap N_{1/2}^- \Omega$ by flow boxes which are very narrow in the unstable direction. This will be useful in the proof of Lemma 4.13 where we linearize the phase functions of the oscillatory integrals that arise over the course of the proof.

Lemma 4.6. *The collection of flow boxes $\{P_{\iota_j}^- N_{\iota_b}^+ \cdot x : x \in K_j \cap N_1^- \Omega\}$ admits a finite subcover \mathcal{B} of $N_{1/2}^- \Omega$ with uniformly bounded multiplicity; i.e. for all $x \in K_j \cap N_{1/2}^- \Omega$, $\sum_{B \in \mathcal{B}} \mathbb{1}_B(x) \ll 1$.*

Proof. Let \mathcal{Q} denote a cover of the unit neighborhood of K_j by flow boxes of the form $P_{\iota_j}^- N_{\iota_b}^+ \cdot x$, where ι_j is as in (4.25). With the help of the Vitali covering lemma, such cover can be chosen to have multiplicity $C_G \geq 1$, depending only on the dimension of G . We will build our collection of boxes \mathcal{B} by refining this cover as follows.

Let \mathcal{Q}^0 denote the subcollection of boxes $Q \in \mathcal{Q}$ such that Q intersects $K_j \cap N_{1/2}^- \Omega$ non-trivially. For each $Q \in \mathcal{Q}^0$, fix some $x_Q \in Q \cap N_{1/2}^- \Omega$. Then, we can find a finite set of points $\{u_i : i \in I_Q\} \subset N_{2\iota_j}^+$ such that the points $x_i := u_i x_Q$ belong to $N_1^- \Omega$ and so that the balls $N_{\iota_b}^+ \cdot x_i$ provide a cover of $N_{1/2}^- \Omega \cap N_{\iota_j}^+ \cdot x_Q$ with uniformly bounded multiplicity (i.e. with multiplicity that is independent of b and j). This is again possible thanks to the Vitali covering lemma. Now, define

$$\mathcal{B} := \{P_{\iota_j}^- N_{\iota_b}^+ \cdot u_i x_Q : i \in I_Q, Q \in \mathcal{Q}^0\}.$$

To bound the multiplicity of \mathcal{B} , let $x \in K_j \cap N_{1/2}^- \Omega$ be arbitrary, and note that

$$\sum_{B \in \mathcal{B}} \mathbb{1}_B(x) = \sum_{Q \in \mathcal{Q}^0} \sum_{i \in I_Q} \mathbb{1}_{P_{\iota_j}^- N_{\iota_b}^+ \cdot u_i x_Q}(x) \ll \sum_{Q \in \mathcal{Q}^0} \mathbb{1}_{\cup_{i \in I_Q} P_{\iota_j}^- N_{\iota_b}^+ \cdot u_i x_Q}(x).$$

Moreover, if $Q = P_{\iota_j}^- N_{\iota_b}^+ \cdot x'_Q$ for some x'_Q , then the union $\cup_{i \in I_Q} P_{\iota_j}^- N_{\iota_b}^+ \cdot u_i x_Q$ is contained inside $Q^+ := P_{\iota_j}^- N_{2\iota_j}^+ \cdot x'_Q$. Finally, bounded multiplicity of \mathcal{Q}^0 implies that $\sum_{Q \in \mathcal{Q}^0} \mathbb{1}_{Q^+}(x) \ll 1$. This concludes the proof. \square

Let \mathcal{B} be the finite cover provided by Lemma 4.6 and let \mathcal{P} denote a partition of unity subordinate to it. For each $\rho \in \mathcal{P}$, we denote by B_ρ the element of \mathcal{B} containing the support of ρ . In particular, such partition of unity can be chosen so that for all $\rho \in \mathcal{P}$, we have

$$\|\rho\|_{C^1} \ll b^{2/3}. \quad (4.27)$$

Over the course of the argument, we need to apply the geodesic flow to enlarge the width of the boxes B_ρ in the N^+ direction to be $\asymp 1$, for e.g. to apply Theorems 2.3 and 2.5. It will be important to ensure that these boxes meet the compact set K_j after flowing. To this end, we define the following subset of \mathcal{P} consisting of boxes which return to K_j at time $b^{2/3}$:

$$\mathcal{P}_b := \{\rho \in \mathcal{P} : g_{-\log \iota_b} B_\rho \cap K_j \neq \emptyset\}. \quad (4.28)$$

Note that for each $B \in \mathcal{B}$, $g_{-\log \iota_b} B$ has diameter $O(1)$.

Transversals. We fix a system of transversals $\{T_\rho\}$ to the strong unstable foliation inside the boxes B_ρ . Since B_ρ meets $N_{1/2}^- \Omega$ for all $\rho \in \mathcal{P}$, we take y_ρ in the intersection $B_\rho \cap N_{1/2}^- \Omega$. In this notation, we can find neighborhoods of identity $P_\rho^- \subset P^- = MAN^-$ and $N_\rho^+ \subset N^+$ such that

$$B_\rho = P_\rho^- N_\rho^+ \cdot y_\rho, \quad T_\rho = P_\rho^- \cdot y_\rho. \quad (4.29)$$

We also let M_ρ, A_ρ , and N_ρ^- be neighborhoods of identity in M, A , and N^- respectively so that $P_\rho^- = M_\rho A_\rho N_\rho^-$.

Localizing away from the cusp. Our next step is to restrict the support of the integral away from the cusp. Define the following smoothed cusp indicator function $\zeta_j : X \rightarrow [0, 1]$ by $\zeta_j(y) := 1 - \sum_{\rho \in \mathcal{P}} \rho(y)$. Let

$$\gamma = 1/2, \quad g^\gamma := g_{\gamma(w+jT_0)}. \quad (4.30)$$

It will be convenient to take T_0 large enough so that

$$(1 - \gamma)(w + jT_0) = \gamma(w + jT_0) \geq 4. \quad (4.31)$$

First, we note that Proposition 3.9 implies $|\mathcal{L}_t F(g_j^w nx)| \ll e_{1,0}^*(f) \mathcal{L}_t V(g_j^w nx)$. Hence, since $|\phi|$ is bounded by 1 and ζ_j is non-negative, we obtain

$$\left| \int_{N_1^+} \phi(n) \zeta_j(g^\gamma nx) \mathcal{L}_t F(g_j^w nx) d\mu_x^u \right| \ll e_{1,0}^*(f) \left| \int_{N_1^+} \zeta_j(g^\gamma nx) \mathcal{L}_t V(g_j^w nx) d\mu_x^u \right|.$$

To proceed, we show that the support of the above integral is in the cusp to apply Theorem 2.3.

Lemma 4.7. *For every $n \in \text{supp}(\mu_x^u) \cap N_1^+$, we have $\zeta_j(g^\gamma nx) > 0 \implies V(g^\gamma nx) > e^{2\beta\alpha j T_0}$.*

Proof. Let $n \in \text{supp}(\mu_x^u) \cap N_1^+$. Since $x \in N_1^- \Omega$, it follows by (2.6) and Remark 2.1 that $nx \in N_2^- \Omega$. Then, $g^\gamma nx \in N_r^- \Omega$, for $r = 2e^{-\gamma(w+jT_0)}$. By (4.31), we get that $g^\gamma nx \in N_{1/2}^- \Omega$. On the other hand, if $\zeta_j(g^\gamma nx) > 0$, then $g^\gamma nx \notin K_j \cap N_{1/2}^- \Omega$. The lemma now follows by definition of K_j in (4.25). \square

Lemma 4.7 and the Cauchy-Schwarz inequality thus yield

$$\left| \int_{N_1^+} \zeta_j(g^\gamma nx) \mathcal{L}_t V(g_j^w nx) d\mu_x^u \right|^2 \leq \mu_x^u \left(n \in N_1^+ : V(g^\gamma nx) > e^{2\beta\alpha j T_0} \right) \times \int_{N_1^+} \mathcal{L}_t V^2(g_j^w nx) d\mu_x^u.$$

By Theorem 2.3 and Chebyshev's inequality, we have that the set on the right side has measure $O(e^{-2\beta\alpha j T_0} V(x) \mu_x^u(N_1^+))$. Moreover, recall that we are assuming that V^2 satisfies the Margulis inequality in Theorem 2.3; cf. Remark 4.1. Hence, applying Theorem 2.3 once more shows that the integral on the right side is at most $O(V^2(x) \mu_x^u(N_1^+))$. These bounds together yield

$$\left| \int_{N_1^+} \phi(n) \zeta_j(g^\gamma nx) \mathcal{L}_t F(g_j^w nx) d\mu_x^u \right| \ll e_{1,0}^*(f) \mu_x^u(N_1^+) V^{3/2}(x) e^{-\beta\alpha j T_0}.$$

Using the bound on $V(x)$ in (4.18), we get

$$\begin{aligned} & \int_{N_1^+} \phi(n) \mathcal{L}_t F(g_j^w nx) d\mu_x^u(n) \\ &= \sum_{\rho \in \mathcal{P}} \int_{N_1^+} \phi(n) \rho(g^\gamma nx) \mathcal{L}_t F(g_j^w nx) d\mu_x^u + O\left(e_{1,0}^*(f) \mu_x^u(N_1^+) V(x) e^{-3\beta\alpha j T_0/4}\right). \end{aligned} \quad (4.32)$$

Using (4.15) to sum the above errors over j , we obtain an error term of the form

$$O\left(\frac{e_{1,0}^*(f) \mu_x^u(N_1^+) V(x)}{(a + 3\beta\alpha/4)^m}\right). \quad (4.33)$$

Saturation and localization to flow boxes. Next, we partition the integral over N_1^+ into pieces according to the flow box they land in under flowing by g^γ . To simplify notation, we write

$$x_j := g^\gamma x. \quad (4.34)$$

We denote by $N_1^+(j)$ a neighborhood of N_1^+ defined by the property that the intersection

$$B_\rho \cap (\text{Ad}(g^\gamma)(N_1^+(j)) \cdot x_j)$$

consists entirely of full local strong unstable leaves in B_ρ . We note that since $\text{Ad}(g^\gamma)$ expands N^+ and B_ρ has radius < 1 , $N_1^+(j)$ is contained inside N_2^+ . Since ϕ is supported inside N_1^+ , we have

$$\chi_{N_1^+}(n) \phi(n) = \chi_{N_1^+(j)}(n) \phi(n), \quad \forall n \in N^+. \quad (4.35)$$

For simplicity, we set

$$\varphi_j(n) := \phi(\text{Ad}(g^\gamma)^{-1}n), \quad \mathcal{A}_j := \text{Ad}(g^\gamma)(N_1^+(j)).$$

For $\rho \in \mathcal{P}$, we let $\tilde{W}_{\rho,j}$ denote the collection of connected components of the set

$$\{n \in \mathcal{A}_j : nx_j \in B_\rho\}.$$

To simplify notation, let

$$F_\gamma := \mathcal{L}_{(1-\gamma)(w+jT_0)}(F). \quad (4.36)$$

In view of (4.35), changing variables using (2.3) yields

$$\sum_{\rho \in \mathcal{P}} \int_{N_1^+} \phi(n) \rho(g^\gamma n x) \mathcal{L}_t F(g_j^w n x) d\mu_x^u = e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}, W \in \tilde{\mathcal{W}}_{\rho,j}} \int_{n \in W} \varphi_j(n) \rho(n x_j) F_\gamma(g_t n x_j) d\mu_{x_j}^u. \quad (4.37)$$

Contribution of non-recurrent orbits. In this subsection, we wish to restrict our attention to those $n \in N_1^+$ for which the orbit $(g_t n x)$ spends most of its time away from the cusp.

To this end, let $\varepsilon_1 = \beta\alpha/2(1+2\alpha)$ and apply Theorem 2.5 with $\varepsilon = \varepsilon_1$ to find $H = H(\varepsilon_1, T_0) \geq 1$ such that the conclusion of the theorem holds. Let χ_H denote the indicator function of the set $\{x : V(x) > H\}$ and define

$$\tilde{\mathcal{E}}_j = \left\{ n \in N_1^+ : \int_0^{(1+2\alpha)\gamma(w+jT_0)} \chi_H(g_t n x) dt > 2\alpha\gamma(w+jT_0) \right\}.$$

We wish to define a saturated version of the set $\tilde{\mathcal{E}}_j$, which we denote by \mathcal{E}_j . The goal of doing so is to discard all the disks $W \in \tilde{\mathcal{W}}_{\rho,j}$ with the property that it contains the image of a point in $\tilde{\mathcal{E}}_j$ under g^γ . In particular, \mathcal{E}_j is contained inside the $O(e^{-\gamma(w+jT_0)\iota_b})$ -neighborhood of $\tilde{\mathcal{E}}_j$. This is made precise in the following lemma.

Lemma 4.8. *Let ρ and $W \in \tilde{\mathcal{W}}_{\rho,j}$ be arbitrary. Then, for all $n_1, n_2 \in N_1^+$ with $g^\gamma n_i x, i = 1, 2$, we have that $V(g_t n_1 x) \asymp V(g_t n_2 x)$, uniformly over all ρ, W and $0 \leq t \leq (1+2\alpha)\gamma(w+jT_0)$.*

Proof. Since B_ρ has width ι_b along N^+ , we have $d_{N^+}(n_1, n_2) \ll e^{-\gamma(w+jT_0)\iota_b}$. Hence, letting $n^t = g_t n_1 n_2^{-1} g_{-t}$, we get $d_{N^+}(n^t, \text{id}) \ll e^{2\alpha\gamma(w+jT_0)\iota_b}$, for all $t \leq (1+2\alpha)\gamma(w+jT_0)$. By (4.14) and (4.16), $d_{N^+}(n^t, \text{id}) \ll_{T_0} b^{-1/2} \ll 1$. The lemma follows from Prop. 2.4 since $g_t n_1 x = n^t g_t n_2 x$. \square

Given $n \in N_1^+$, let $\rho(n)$ be a flow box index satisfying $g^\gamma n x \in B_{\rho(n)}$. We also let $W(n) \in \tilde{\mathcal{W}}_{\rho(n),j}$ be such that $g^\gamma n x \in W(n) \subset B_{\rho(n)}$. With this notation, we define \mathcal{E}_j as follows:

$$\mathcal{E}_j := \left\{ n \in N_1^+ : \text{there is } n' \in \tilde{\mathcal{E}}_j \text{ such that } W(n) = W(n') \right\}. \quad (4.38)$$

By Lemma 4.8, there is a uniform constant $C \geq 1$ such that

$$\mathcal{E}_j \subseteq \left\{ n \in N_1^+ : \int_0^{(1+2\alpha)\gamma(w+jT_0)} \chi_{H/C}(g_t n x) dt > 2\alpha\gamma(w+jT_0) \right\}.$$

Hence, Theorem 2.5 shows that $\mu_x^u(\mathcal{E}_j)$ is $O(e^{-(\beta\theta-\varepsilon_1)(1+2\alpha)\gamma(w+jT_0)} V(x) \mu_x^u(N_1^+))$, where $\theta = 2\alpha/(1+2\alpha)$. Here, we used Prop. 2.4 to deduce this continuous time version of Theorem 2.5 from its discrete time formulation. Moreover, recalling that $\gamma = 1/2$, we obtain by (4.18) that $V(x) \leq e^{\beta\alpha\gamma j T_0}$. Since $\varepsilon_1 = \beta\alpha/2(1+2\alpha)$, these bounds thus yield

$$\mu_x^u(\mathcal{E}_j) \ll e^{-\beta\alpha\gamma(w+jT_0)/2} \mu_x^u(N_1^+). \quad (4.39)$$

Next, we let $\mathcal{W}_{\rho,j}$ denote the connected components that avoid \mathcal{E}_j . More precisely, let

$$\mathcal{W}_{\rho,j} := \left\{ W \in \tilde{\mathcal{W}}_{\rho,j} : W \neq W(n) \text{ for any } n \in \mathcal{E}_j \right\}. \quad (4.40)$$

We now restrict the sum in (4.37) to the subsets $\mathcal{W}_{\rho,j}$. Reversing the change of variables in (4.37) and using the fact that our test functions have C^0 -norm at most 1, we get

$$(4.37) = e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}, W \in \mathcal{W}_{\rho,j}} \int_{n \in W} \varphi_j(n) \rho(nx_j) F_\gamma(g_t nx_j) d\mu_{x_j}^u + O \left(\int_{\mathcal{E}_j} |\mathcal{L}_t F(g_j^w nx)| d\mu_x^u \right). \quad (4.41)$$

To estimate the integral on the right side, we argue as before using Proposition 3.9, to get that $|\mathcal{L}_t F(g_j^w nx)|$ is at most $O(e_{1,0}^*(f) \mathcal{L}_t V(g_j^w nx))$. Then, Cauchy-Schwarz gives

$$\left| \int_{\mathcal{E}_j} \mathcal{L}_t V(g_j^w nx) d\mu_x^u \right|^2 \leq \mu_x^u(\mathcal{E}_j) \times \int_{N_1^+} \mathcal{L}_t V^2(g_j^w nx) d\mu_x^u.$$

Applying Theorem 2.3 on integrability of V^2 and (4.39), we obtain

$$\int_{\mathcal{E}_j} |\mathcal{L}_t F(g_j^w nx)| d\mu_x^u \ll e^{-\beta\alpha\gamma(w+jT_0)/4} e_{1,0}^*(f) V(x) \mu_x^u(N_1^+).$$

Our next step is to restrict the sum in (4.41) to the recurrent boxes \mathcal{P}_b defined in (4.28) using a similar argument. Let

$$\tilde{\mathcal{E}}_b = \left\{ n \in N_1^+ : g^\gamma nx \notin \bigcup_{\rho \in \mathcal{P}_b} B_\rho \right\}, \quad (4.42)$$

and define its saturation \mathcal{E}_b analogously to (4.38). Then, by definition of \mathcal{P}_b and a similar argument to Lemma 4.8, we can find a uniform constant $C \geq 1$ so that \mathcal{E}_b is contained in the set of $n \in N_1^+$ so that $V(g_{-\log \iota_b} g^\gamma nx) > e^{2\beta\alpha j T_0}/C$. Theorem 2.3 then gives that \mathcal{E}_b has measure $O(e^{-2\beta\alpha j T_0} V(x) \mu_x^u(N_1^+))$. Hence, splitting the sum in (4.41) into sum over \mathcal{P}_b and $\mathcal{P} \setminus \mathcal{P}_b$, and recalling that $w \leq 2T_0$, we get

$$(4.37) = e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}_b, W \in \mathcal{W}_{\rho,j}} \int_W \varphi_j(n) \rho(nx_j) F_\gamma(g_t nx_j) d\mu_{x_j}^u + O(e^{-\beta\alpha\gamma(w+jT_0)/4} V(x) \mu_x^u(N_1^+)). \quad (4.43)$$

As in (4.15), summing over j , we obtain an error term of the form

$$O \left(\frac{e_{1,0}^*(f)}{(a + \beta\alpha/4)^m} \right). \quad (4.44)$$

The remainder of the section, is dedicated to estimating the sum on the right side of (4.43).

Centering the integrals. It will be convenient to center all the integrals in (4.37) so that their basepoints belong to the transversals T_ρ of the respective flow box B_ρ ; cf. (4.29).

Let $I_{\rho,j}$ denote an index set for $\mathcal{W}_{\rho,j}$. For $W \in \mathcal{W}_{\rho,j}$ with index $\ell \in I_{\rho,j}$, let $n_{\rho,\ell} \in W$, $m_{\rho,\ell} \in M_\rho$, $n_{\rho,\ell}^- \in N_\rho^-$, and $t_{\rho,\ell} \in (-\iota_j, \iota_j)$ be such that

$$x_{\rho,\ell} := m_{\rho,\ell} g_{-t_{\rho,\ell}} n_{\rho,\ell} \cdot x_j = n_{\rho,\ell}^- \cdot y_\rho \in T_\rho. \quad (4.45)$$

Arguing as in the proof of Lemma 4.7, since x belongs to $N_1^- \Omega$, we have that

$$x_{\rho,\ell} \in N_1^- \Omega. \quad (4.46)$$

Moreover, if we let $u_\ell = \text{Ad}((g^\gamma)^{-1})(n_{\rho,\ell}) \in N_1^+(j)$, then in light of the fact that the components $W \in \mathcal{W}_{\rho,j}$ correspond to recurrent orbits, cf. (4.40) for a precise definition, we may and will assume that there is $s_{\rho,\ell} > 0$ such that

$$\gamma(w + jT_0) \leq s_{\rho,\ell} \leq (1 + 2\alpha)\gamma(w + jT_0), \quad V(g_{s_{\rho,\ell}} u_\ell x) \ll_{T_0} 1. \quad (4.47)$$

Regularity of test functions. For each such ℓ and W , let $W_\ell = \text{Ad}(m_{\rho,\ell} g_{t_{\rho,\ell}})(W n_{\rho,\ell}^{-1})$ and

$$\tilde{\phi}_{\rho,\ell}(t, n) := p_{j,w}(t - t_{\rho,\ell}) \cdot e^{zt_{\rho,\ell}} \cdot \phi(\text{Ad}(m_{\rho,\ell} g_{-t_{\rho,\ell}})^{-1}(n n_{\rho,\ell})) \cdot \rho(g_{t_{\rho,\ell}} n x_{\rho,\ell}). \quad (4.48)$$

Note that $\tilde{\phi}_{\rho,\ell}$ has bounded support in the t direction and (4.27) implies

$$\|\tilde{\phi}_{\rho,\ell}\|_{C^0(\mathbb{R} \times N^+)} \ll 1, \quad \|\tilde{\phi}_{\rho,\ell}(t, \cdot)\|_{C^{0,1}(N^+)} \ll \iota_b^{-1}, \quad (4.49)$$

for all $t \in \mathbb{R}$. Moreover, recalling (4.8), we see that

$$\|\tilde{\phi}_{\rho,\ell}\|_{C^{0,1}(\mathbb{R} \times N^+)} \ll \iota_b^{-1} m. \quad (4.50)$$

Integrating the main term in (4.43) in t , and changing variables using (2.3), we get

$$\begin{aligned} e^{-\delta\gamma(w+jT_0)} \int_{\mathbb{R}} e^{-zt} p_{j,w}(t) \sum_{\rho \in \mathcal{P}_b, W \in \mathcal{W}_{\rho,j}} \int_{n \in W} \varphi_j(n) \rho(n x_j) F_\gamma(g_t n x_j) d\mu_{x_j}^u(n) dt \\ = e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}_b} \sum_{\ell \in I_{\rho,j}} \int_{\mathbb{R}} e^{-zt} \int_{n \in W_\ell} \tilde{\phi}_{\rho,\ell}(t, n) F_\gamma(g_{t+t_{\rho,\ell}} n x_{\rho,\ell}) d\mu_{x_{\rho,\ell}}^u(n) dt, \end{aligned} \quad (4.51)$$

where we also used M -invariance of F_γ ; cf. Remark 3.7.

Mass estimates. We record here certain counting estimates which will allow us to sum error terms in later estimates over \mathcal{P} . Note that by definition of $N_1^+(j)$, we have $\bigcup_{\rho \in \mathcal{P}, W \in \mathcal{W}_{\rho,j}} W \subseteq \mathcal{A}_j$. Thus, using the log-Lipschitz and contraction properties of V , it follows that

$$\begin{aligned} \sum_{\rho \in \mathcal{P}, \ell \in I_{\rho,j}} \mu_{x_{\rho,\ell}}^u(W_\ell) V(x_{\rho,\ell}) &\ll \int_{\mathcal{A}_j} V(n x_j) d\mu_{x_j}^u(n) \\ &= e^{\delta\gamma(w+jT_0)} \int_{N_1^+(j)} V(g_j^w n x) d\mu_x^u(n) \ll e^{\delta\gamma(w+jT_0)} \mu_x^u(N_1^+) V(x), \end{aligned} \quad (4.52)$$

where we used that $|t_{\rho,\ell}| < 1$ and the last inequality follows by Proposition 2.2 since $N_1^+(j) \subseteq N_2^+$. We also used the uniformly bounded multiplicity of the partition of unity \mathcal{P} .

We also need the following weighted number of flow boxes parametrized by \mathcal{P} .

Lemma 4.9. *Recall that $\iota_b = b^{-2/3}$. Then, we have*

$$\sum_{\rho \in \mathcal{P}} \mu_{y_\rho}^u(N_{\iota_b}^+) \ll e^{O_\beta(\alpha_j T_0)}.$$

Proof. Recall the Bowen-Margulis-Sullivan measure defined below (2.1), and its conditional measures μ_\bullet^s along orbits of N^- defined analogously to (2.2). Recall further that each B_ρ is of the form $P_\rho^- N_\rho^+ \cdot y_\rho$, where P_ρ^- and N_ρ^+ are identity neighborhoods of radius $\asymp \iota_j$ and $\asymp \iota_b$ respectively. Bounded multiplicity of \mathcal{P} implies that $\sum_\rho m^{\text{BMS}}(B_\rho) \ll 1$. Hence, the local product structure of m^{BMS} implies that

$$m^{\text{BMS}}(B_\rho) \asymp \iota_j^{\dim M + 1} \mu_{y_\rho}^u(N_{\iota_b}^+) \mu_{y_\rho}^s(N_{\iota_j}^-),$$

where M is the centralizer of the geodesic flow inside the maximal compact group K , and μ_\bullet^s are the conditional measures along N^- -orbits defined similarly to (2.2). This estimate implicitly uses the uniform doubling property from Prop. 2.2 to assert that the ratio of the measures of all local strong (un)stable disks inside B_ρ is uniformly $O(1)$. Finally, by definition of $\mu_{y_\rho}^s$, we have that $\mu_{y_\rho}^s(N_{\iota_j}^-) \gg e^{-\delta \text{dist}(o, y_\rho)}$. By Prop. 2.4, we have that $e^{\text{dist}(y_\rho, o)} \ll V(y_\rho)^{O_\beta(1)} \ll e^{O_\beta(\alpha_j T_0)}$, since y_ρ belongs to the unit neighborhood of the set K_j defined in (4.25). The lemma follows by combining the above estimates with (4.26). \square

Remark 4.10. The proof of Lemma 4.9 shows that the sum in question is $O_\Gamma(1)$ when Γ is convex cocompact. The point of the above lemma is that this sum has, in general, much fewer terms than the sum in (4.52).

Stable holonomy. Fix some $\rho \in \mathcal{P}_b$. Recall the points $y_\rho \in T_\rho$ and $n_{\rho,\ell}^- \in N_\rho^-$ satisfying (4.45). The product map $M \times N^- \times A \times N^+ \rightarrow G$ is a diffeomorphism on a ball of radius 1 around identity; cf. Section 2.5. Hence, given $\ell \in I_{\rho,j}$, we can define maps \tilde{u}_ℓ , $\tilde{\tau}_\ell$, m_ℓ and \tilde{u}_ℓ^- from W_ℓ to N^+ , \mathbb{R} , M and N^- respectively by the following formula

$$g_{t+t_{\rho,\ell}} n n_{\rho,\ell}^- = g_{t+t_{\rho,\ell}} m_\ell(n) \tilde{u}_\ell^-(n) g_{\tilde{\tau}_\ell(n)} \tilde{u}_\ell(n) = m_\ell(n) \tilde{u}_\ell^-(t, n) g_{t+t_{\rho,\ell}+\tilde{\tau}_\ell(n)} \tilde{u}_\ell(n), \quad (4.53)$$

where we set $\tilde{u}_\ell^-(t, n) = \text{Ad}(g_{t+t_{\rho,\ell}})(\tilde{u}_\ell^-(n))$. We define the following change of variable map:

$$\Phi_\ell : \mathbb{R} \times W_\ell \rightarrow \mathbb{R} \times N^+, \quad \Phi_\ell(t, n) = (t + \tilde{\tau}_\ell(n), \tilde{u}_\ell(n)). \quad (4.54)$$

We suppress the dependence on ρ and j to ease notation. Then, Φ_ℓ induces a map between the weak unstable manifolds of $x_{\rho,\ell}$ and y_ρ , also denoted Φ_ℓ , and defined by

$$\Phi_\ell(g_t n x_{\rho,\ell}) = g_{t+\tilde{\tau}_\ell(n)} \tilde{u}_\ell(n) y_\rho.$$

In particular, this induced map coincides with the local strong stable holonomy map inside B_ρ .

Note that we can find a neighborhood $W_\rho \subset N^+$ of identity of radius $\asymp \iota_b$ such that

$$\Phi_\ell(\mathbb{R} \times W_\ell) \subseteq \mathbb{R} \times W_\rho, \quad (4.55)$$

for all $\ell \in I_{\rho,j}$. Moreover, we may assume that b is large enough (and hence ι_b is small enough), depending only on G , so that all the maps Φ_ℓ in (4.54) are invertible on $\mathbb{R} \times W_\rho$. Hence, we can define the following:

$$\begin{aligned} \tau_\ell(n) &= \tilde{\tau}_\ell(\tilde{u}_\ell^{-1}(n)) + t_{\rho,\ell} \in \mathbb{R}, & u_\ell^-(t, n) &= \tilde{u}_\ell^-(t - \tau_\ell(n), \tilde{u}_\ell^{-1}(n)) \in N^-, \\ \phi_{\rho,\ell}(t, n) &= e^{-a(t-\tau_\ell(n))} \times J\Phi_\ell(n) \times \tilde{\phi}_{\rho,\ell}(t - \tau_\ell(n), \tilde{u}_\ell^{-1}(n)), \end{aligned} \quad (4.56)$$

and $J\Phi_\ell$ denotes the Jacobian of the change of variable Φ_ℓ ; cf. (2.7).

Changing variables and using M -invariance of F_γ , we obtain

$$(4.51) = e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}_b} \sum_{\ell \in I_{\rho,j}} \int_{\mathbb{R}} \int_{W_\rho} e^{-ib(t-\tau_\ell(n))} \phi_{\rho,\ell}(t, n) F_\gamma(u_\ell^-(t, n) g_t n y_\rho) d\mu_{y_\rho}^u(n) dt. \quad (4.57)$$

Stable derivatives. Our next step is to remove F_γ from the sum over ℓ in (4.57). Due to non-joint integrability of the stable and unstable foliations, our estimate involves a derivative of f in the flow direction. In particular, since the norm of flow derivatives are scaled by a factor of $1/B$, cf. (4.2), this step is the most “expensive” estimate in our argument. In essence, the reason we only flow by $(w + jT_0)/2$ is to compensate for the loss in this step.

Recall the definition of F_γ in (4.36). Since y_ρ belongs to $N_{1/2}^- \Omega$ and $u_\ell^-(t, n)$ belongs to a neighborhood of identity in N^- of radius $O(\iota_j)$, uniformly over (t, n) in the support of our integrals, Proposition 3.10 yields

$$|F_\gamma(u_\ell^-(t, n) g_t n y_\rho) - F_\gamma(g_t n y_\rho)| \ll e^{-(1-\gamma)(w+jT_0)} \|f\|_1^* V(y_\rho), \quad (4.58)$$

where we implicitly used the fact that $W_\rho \subset N_1^+$ and $|t| \leq 1$ so that $V(g_t n y_\rho) \ll V(y_\rho)$. Indeed, the additional gain is due to the fact that g_s contracts N^- by at least e^{-s} for all $s \geq 0$.

To sum the above errors over ℓ and ρ , we wish to use (4.52). We first note that Propositions 2.2 and 2.4 allow us to use closeness of y_ρ and $x_{\rho,\ell}$ along with regularity of holonomy to deduce that

$$V(y_\rho) \mu_{y_\rho}^u(W_\rho) \asymp V(x_{\rho,\ell}) \mu_{x_{\rho,\ell}}^u(W_\ell). \quad (4.59)$$

Here, we also use the fact that both $x_{\rho,\ell}$ and y_ρ belong to $N_1^-\Omega$; cf. (4.46). Hence, we can use (4.52) to estimate the sum of the errors in (4.58) yielding

$$(4.57) = e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}_b} \sum_{\ell \in I_{\rho,j}} \int_{\mathbb{R}} \int_{W_\rho} \left(\sum_{\ell \in I_{\rho,j}} e^{-ib(t-\tau_\ell(n))} \phi_{\rho,\ell}(t,n) \right) F_\gamma(g_t n y_\rho) d\mu_{y_\rho}^u dt \\ + O \left(e^{-(1-\gamma)(w+jT_0)} \|f\|_1^* \mu_x^u(N_1^+) V(x) \right),$$

where we used that the above integrands have uniformly bounded support in the \mathbb{R} direction, independently of ℓ (and ρ). Indeed, this boundedness follows from that of the partition of unity p_j ; cf. (4.8). We also used (4.49) to bound the C^0 norm of $\phi_{\rho,\ell}$. Summing over j and w using (4.15), and recalling that $\gamma = 1/2$, we obtain

$$O_{T_0} \left(\frac{\|f\|_1^* \mu_x^u(N_1^+) V(x)}{(a+1/2)^m} \right).$$

Recall the norm $\|\cdot\|_{1,B}^*$ defined in (4.2) and note that $\|\cdot\|_1^* \leq B \|\cdot\|_{1,B}^*$. Choosing a and $\varkappa > 0$ small enough, we can ensure that $e^{1+\varkappa}/(a+1/2)$ is at most $1/(a+1/6)$. With this choice, taking $B = b^{1+\varkappa}$ yields an error term of the form:

$$O \left(\frac{\|f\|_{1,B}^* \mu_x^u(N_1^+) V(x)}{(a+1/6)^m} \right). \quad (4.60)$$

Mollifiers and Cauchy-Schwarz. We are left with estimating integrals of the form:

$$\int_{\mathbb{R} \times W_\rho} \Psi_\rho(t,n) F_\gamma(g_t n y_\rho) d\mu_{y_\rho}^u dt, \quad \Psi_\rho(t,n) := \sum_{\ell \in I_{\rho,j}} e^{-ib(t-\tau_\ell(n))} \phi_{\rho,\ell}(t,n). \quad (4.61)$$

We begin by giving an apriori bound on Ψ_ρ . Denote by $J_\rho \subset \mathbb{R}$ the bounded support of the integrand in t coordinate of the above integrals. Note that (4.49) and the fact that $|t| \ll 1$ imply

$$\|\phi_{\rho,\ell}\|_{L^\infty(J_\rho \times W_\rho)} \ll 1, \quad \|\Psi_\rho\|_{L^\infty(J_\rho \times W_\rho)} \ll \#I_{\rho,j}. \quad (4.62)$$

To simplify notation, we let

$$r = (1-\gamma)(w+jT_0).$$

Note that we have that $y_\rho \in N_1^-\Omega$, $|J_\rho| \ll 1$, and $r \geq 1$. Hence, Proposition 3.9, along with (4.24), the definition of F_γ in (4.36) and the Cauchy-Schwarz inequality, yield

$$\left| \int_{\mathbb{R} \times W_\rho} \Psi_\rho(t,n) F_\gamma(g_t n y_\rho) d\mu_{y_\rho}^u dt \right|^2 \ll e_{1,0}^*(f)^2 \int_{J_\rho \times W_\rho} |\Psi_\rho(t,n)|^2 d\mu_{y_\rho}^u dt \int_{W_\rho} V^2(g_r n y_\rho) d\mu_{y_\rho}^u.$$

The following lemma estimates the integral of V^2 on the right side of the above inequality.

Lemma 4.11. *We have the bound $\int_{W_\rho} V^2(g_r n y_\rho) d\mu_{y_\rho}^u \ll_{T_0} e^{4\beta\alpha j T_0} \mu_{y_\rho}^u(W_\rho)$.*

Proof. Recall that W_ρ has radius $\asymp \iota_b = b^{-2/3}$, and hence the expanded disk $W_\rho^b = \text{Ad}(g_{-\log \iota_b})(W_\rho)$ has radius $\asymp 1$. We also recall from (4.14) that $r \geq -\log \iota_b$. We also that ρ is an element of \mathcal{P}_b defined in (4.28), so that $V^2(g_{-\log \iota_b} y_\rho) \ll_{T_0} e^{4\beta\alpha j T_0}$. Let $r_1 = r + \log \iota_b \geq 0$ and $y_\rho^b = g_{-\log b} y_\rho$. Changing variables using (2.3), and using Remark 4.1 and the Margulis inequality for V^2 in Theorem 2.3, we deduce the lemma from the following estimate

$$\int_{W_\rho} V^2(g_r n y_\rho) d\mu_{y_\rho}^u = \iota_b^\delta \int_{W_\rho^b} V^2(g_{r_1} n y_\rho^b) d\mu_{y_\rho^b}^u \ll \iota_b^\delta V^2(y_\rho^b) \mu_{y_\rho^b}^u(W_\rho^b) = V^2(y_\rho^b) \mu_{y_\rho}^u(W_\rho).$$

□

The above lemma hence yields the bound

$$\left| \int_{\mathbb{R} \times W_\rho} \Psi_\rho(t, n) F_\gamma(g_t n y_\rho) d\mu_{y_\rho}^u dt \right|^2 \ll_{T_0} e_{1,0}^*(f)^2 e^{4\beta\alpha j T_0} \mu_{y_\rho}^u(W_\rho) \int_{J_\rho \times W_\rho} |\Psi_\rho(t, n)|^2 d\mu_{y_\rho}^u dt. \quad (4.63)$$

Cusp-adapted partitions. To estimate the right side of (4.63), it will be convenient to linearize the phase functions τ_k . For this purpose, we need to pick a cover of W_ρ by balls with radius determined by a certain return time of their centers to a given compact set.

Proposition 4.12. *There exists $\beta_0 \asymp \beta$ such that the following holds. For all $b \geq 1$ and $\rho \in \mathcal{P}_b$, there exist a cover $\{A_i : i\}$ of W_ρ and a set $\mathcal{R}_\rho \subseteq W_\rho$ with $\mu_{y_\rho}^u(W_\rho \setminus \mathcal{R}_\rho) \ll_{T_0} b^{-\beta_0} e^{2\alpha\beta j T_0} \mu_{y_\rho}^u(W_\rho)$ such that for all i with $A_i \cap \mathcal{R}_\rho \neq \emptyset$, we have*

- (1) A_i has the form $A_i = N_{r_i}^+ \cdot u_i$ for some $r_i > 0$ and $u_i \in W_\rho$.
- (2) If $t_i = -\log r_i$, then $V(g_{t_i} u y_\rho) \ll_\beta 1$ for all $u \in A_i$.
- (3) $b^{-8/10} \ll r_i \ll b^{-7/10}$.
- (4) $\sum_i \mu_{y_\rho}^u(A_i) \ll \mu_{y_\rho}^u(W_\rho)$.

Proof. Let $y_\rho^b = g_{-\log \iota_b} y_\rho$, where $\iota_b = b^{-2/3}$. Then, since $\rho \in \mathcal{P}_b$, by (4.28), $V(y_\rho^b) \ll_{T_0} e^{2\beta\alpha j T_0}$. As in Lemma 4.11, the disk $W_\rho^b = \text{Ad}(g_{-\log \iota_b})(W_\rho)$ has radius $\asymp 1$.

Let $r_0 \geq 1$ be the constant provided by Theorem 2.5 applied with $\varepsilon = \beta/40$. Let $m_0 = \lceil r_0^{-1} \log b \rceil$ and let $H = e^{3\beta r_0}$ be the height provided by Theorem 2.5. Let χ_H denote the indicator function of the set of points of height at least H , i.e. the set $\{y : V(y) > H\}$. Then, Theorem 2.5 yields

$$\mu_{y_\rho^b}^u \left(n \in W_\rho^b : \sum_{1 \leq \ell \leq m_0} \chi_H(g_{\ell r_0} n y_\rho^b) > m_0/20 \right) \ll_\beta b^{-\beta/40 r_0} V(y_\rho^b) \mu_{y_\rho^b}^u(W_\rho^b).$$

Denote the set on the left side in the above estimate by \mathcal{E}_ρ^b . Let $c = 7/10 - 2/3$ and $d = 8/10 - 2/3$. We claim that, if b is large enough, then for every $n \in W_\rho^b \setminus \mathcal{E}_\rho^b$, we can find $\eta \in [c, d]$ such that $V(g_{\eta \log b} n y_\rho^b) \leq H$. Indeed, suppose not. Then, it follows that

$$\sum_{1 \leq \ell \leq m_0} \chi_H(g_{\ell r_0} n x) \geq \frac{\log b}{10 r_0} - 1 \geq m_0/10 - 2.$$

This contradicts the fact that $n \notin \mathcal{E}_\rho^b$ when b is large enough.

Let $\mathcal{R}_\rho^b := \text{supp}(\mu_{y_\rho^b}^u) \cap W_\rho^b \setminus \mathcal{E}_\rho^b$, and define \mathcal{R}_ρ to be its preimage in W_ρ . More precisely, $\mathcal{R}_\rho = \text{Ad}(g_{\log \iota_b})(\mathcal{R}_\rho^b) \subseteq W_\rho$. Define a function $\varsigma : \mathcal{R}_\rho \rightarrow [7/10, 8/10]$ by setting $\varsigma(n)$ to be the least value of $\eta \in [7/10, 8/10]$ such that $V(g_{\eta \log b} n y_\rho) \leq H$. Consider the cover $\{\tilde{A}_u : u \in \mathcal{R}_\rho\}$, where each A_u is the ball around each u of radius $b^{-\varsigma(u)}$. Using the Vitali covering lemma and the uniform doubling in Prop. 2.2, we can find a finite subcover $\{A_{u_i} : i\}$ such that $\sum_i \mu_{y_\rho}^u(A_{u_i}) \ll \mu_{y_\rho}^u(W_\rho)$. This completes the proof by taking $\beta_0 = \beta/40 r_0$, $A_i := A_{u_i}$, and $r_i = 5b^{-\varsigma(u_i)}$. \square

Let $\{A_i\}$ be the cover provided by Proposition 4.12. Combining this result with (4.62), we obtain

$$\int_{J_\rho \times W_\rho} |\Psi_\rho(t, n)|^2 d\mu_{y_\rho}^u dt \leq \sum_i \int_{J_\rho \times A_i} |\Psi_\rho(t, n)|^2 d\mu_{y_\rho}^u dt + O_{T_0} \left(b^{-\beta_0} \# I_{\rho, j}^2 e^{2\beta\alpha j T_0} \mu_{y_\rho}^u(W_\rho) \right). \quad (4.64)$$

Linearizing the phase. We now turn to estimating the sum of oscillatory integrals in (4.64). For $k, \ell \in I_{\rho,j}$, we let

$$\psi_{k,\ell}(t, n) := \phi_{\rho,k}(t, n) \overline{\phi_{\rho,\ell}(t, n)}.$$

Expanding the square, we get

$$\sum_i \int_{J_\rho \times A_i} |\Psi_\rho(t, n)|^2 d\mu_{y_\rho}^u dt = \sum_i \sum_{k, \ell \in I_{\rho,j}} \int_{J_\rho \times A_i} e^{-ib(\tau_k(n) - \tau_\ell(n))} \psi_{k,\ell}(t, n) d\mu_{y_\rho}^u dt. \quad (4.65)$$

Using (2.3), we change variables in the integrals using the maps taking each A_i onto N_1^+ . More precisely, recall that A_i is a ball of radius r_i around $u_i \in W_\rho$. Letting

$$t_i = -\log r_i, \quad y_\rho^i = g_{t_i} u_i y_\rho, \quad \tau_k^i = \tau_k(\text{Ad}(g_{-t_i})(n) u_i), \quad \psi_{k,\ell}^i(t, n) = \psi_{k,\ell}(t, \text{Ad}(g_{-t_i})(n) u_i), \quad (4.66)$$

we can bound the above sum as follows:

$$(4.65) \leq \sum_i e^{-\delta t_i} \sum_{k, \ell \in I_{\rho,j}} \left| \int_{J_\rho \times N_1^+} e^{-ib(\tau_k^i(n) - \tau_\ell^i(n))} \psi_{k,\ell}^i(t, n) d\mu_{y_\rho^i}^u dt \right|. \quad (4.67)$$

We note that the radius r_i of A_i satisfies

$$b^{-8/10} \ll e^{-t_i} = r_i \ll b^{-7/10}. \quad (4.68)$$

We also recall from Proposition 4.12 that r_i was chosen so that

$$V(y_\rho^i) \ll 1, \quad \forall i. \quad (4.69)$$

This is important for the proof of Theorem 4.17 below.

Next, we use the coordinate parametrization of N^+ by its Lie algebra $\mathfrak{n}^+ := \text{Lie}(N^+)$ via the exponential map. We suppress composition with \exp from our notation for simplicity and continue to denote by $\mu_{y_\rho^i}^u$ and N_1^+ their preimage to \mathfrak{n}^+ under \exp .

Recall from Section 2.4 the parametrization of N^- by its Lie algebra $\mathfrak{n}^- = \mathfrak{n}_\alpha^- \oplus \mathfrak{n}_{2\alpha}^-$ via the exponential map and similarly for N^+ . Let $w_i = (v_i, r_i) \in \mathfrak{n}_\alpha^+ \times \mathfrak{n}_{2\alpha}^+$ be such that $u_i = \exp(w_i)$, where u_i is the center of the ball A_i . Recall the notation for transverse intersection points $n_{\rho,k}^-$ in (4.45). For each $k \in I_{\rho,j}$, write

$$n_{\rho,k}^- = \exp(u_k + s_k)$$

with $u_k \in \mathfrak{n}_\alpha^-$ and $s_k \in \mathfrak{n}_{2\alpha}^-$. With this notation, we have the following formula for the temporal functions τ_k . The proof of this lemma is given in Section 5.

Lemma 4.13. *For every i , there exists a bilinear form $\langle \cdot, \cdot \rangle : \mathfrak{n}^- \times \mathfrak{n}_\alpha^+ \rightarrow \mathbb{R}$ such that the following holds. For every $k \in I_{\rho,j}$, there is a constant $c_k^i \in \mathbb{R}$ such that for all $n = \exp(v, r) \in N_1^+$ with $v \in \mathfrak{n}_\alpha^+$ and $r \in \mathfrak{n}_{2\alpha}^+$, we have that*

$$\tau_k^i(n) - \tau_\ell^i(n) = c_{k,\ell}^i + e^{-t_i} \langle u_k - u_\ell + s_k - s_\ell, v \rangle + O(b^{-4/3}).$$

Moreover, for every $(u, s) \in \mathfrak{n}_\alpha^- \times \mathfrak{n}_{2\alpha}^-$, the linear functional $\langle u + s, \cdot \rangle : \mathfrak{n}_\alpha^+ \rightarrow \mathbb{R}$ satisfies

$$\|\langle u + s, \cdot \rangle\| \gg \|u\|,$$

where $\|\langle u + s, \cdot \rangle\| := \sup_{\|v\|=1} |\langle u + s, v \rangle|$.

We apply Lemma 4.13 to linearize the phase and amplitude functions in (4.67). Let

$$w_{k,\ell}^i := e^{-t_i} (u_k - u_\ell + s_k - s_\ell). \quad (4.70)$$

Note that $\text{Ad}(g_{-t_i})$ contracts N^+ by at least $e^{-t_i} \ll b^{-7/10}$; cf. (4.68). Hence, in light of (4.50), the Lipschitz norm of $\psi_{k,\ell}^i$ along N^+ is $O(b^{2/3-7/10})$. Moreover, linearizing the phase function introduces an error $O(b^{1-4/3})$. Hence, recalling that $|J_\rho| \ll 1$, we get

$$(4.67) \ll \sum_i e^{-\delta t_i} \sum_{k,\ell \in I_{\rho,j}} \left| \int_{N_1^+} e^{-ib\langle w_{k,\ell}^i, v \rangle} d\mu_{y_\rho^i}^u \right| + b^{-3/100} \mu_{y_\rho}^u(W_\rho) \# I_{\rho,j}^2, \quad (4.71)$$

where we used the estimate $\sum_i \mu_{y_\rho^i}^u(A_i) \ll \mu_{y_\rho}^u(W_\rho)$.

Excluding close pairs of unstable manifolds. Consider the following partition of $I_{\rho,j}^2$:

$$C_{\rho,j} = \left\{ (k,\ell) \in I_{\rho,j}^2 : \|u_k - u_\ell\| \leq b^{-1/10} \right\}, \quad S_{\rho,j} = I_{\rho,j}^2 \setminus C_{\rho,j}. \quad (4.72)$$

Then, $C_{\rho,j}$ parametrizes pairs of unstable manifolds which are too close along the $\mathfrak{n}_{2\alpha}^-$ direction in the stable foliation. In particular, since $\mathfrak{n}_{2\alpha}^- = \{0\}$ when X is real hyperbolic, $C_{\rho,j}$ simply parametrizes pairs of unstable manifolds which are too close along the stable foliation in this case. With this notation, the sum on the right side of (4.71) can be estimated as follows:

$$\sum_i e^{-\delta t_i} \sum_{k,\ell \in I_{\rho,j}} \left| \int_{N_1^+} e^{-ib\langle w_{k,\ell}^i, v \rangle} d\mu_{y_\rho^i}^u \right| \ll \#C_{\rho,j} \mu_{y_\rho}^u(W_\rho) + \sum_i e^{-\delta t_i} \sum_{(k,\ell) \in S_{\rho,j}} \left| \int_{N_1^+} e^{-ib\langle w_{k,\ell}^i, v \rangle} d\mu_{y_\rho^i}^u \right|. \quad (4.73)$$

We estimate the first term in (4.73) via the following proposition, proved in Section 7.3.

Proposition 4.14. *Assume that the parameter α is chosen sufficiently small. Then, there exists a constant $\kappa_0 > 0$ such that for all $\ell \in I_{\rho,j}$,*

$$\# \{k \in I_{\rho,j} : (k,\ell) \in C_{\rho,j}\} \ll_{T_0} (b^{-\kappa_0/10} + e^{-\kappa_0\gamma(w+jT_0)}) e^{\delta\gamma(w+jT_0)}.$$

We may take $\kappa_0 = \kappa/2$, where κ is the constant provided by Theorem 6.23.

Remark 4.15. When X is non-real hyperbolic, Prop. 4.14 requires a polynomial decay estimate for PS measures near certain proper subspaces of the boundary, Theorem 6.23. When X additionally has cusps, the latter result in turn requires the full strength of the L^2 -flattening results in Section 6. These estimates are not needed in the real hyperbolic case.

In what follows, we shall assume that α is chosen small enough so that Prop. 4.14 holds. Summarizing our estimates in (4.64), (4.71), (4.73), and Proposition 4.14, we have shown that

$$\begin{aligned} & \int_{J_\rho \times W_\rho} |\Psi_\rho(t, n)|^2 d\mu_{y_\rho}^u dt \\ & \ll \sum_i e^{-\delta t_i} \sum_{(k,\ell) \in S_{\rho,j}} \left| \int_{N_1^+} e^{-ib\langle w_{k,\ell}^i, v \rangle} d\mu_{y_\rho^i}^u \right| \\ & + \left((b^{-\beta_0} e^{2\beta_0\alpha j T_0} + b^{-3/100}) \# I_{\rho,j} + (b^{-\kappa_0/10} + e^{-\kappa_0\gamma(w+jT_0)}) e^{\delta\gamma(w+jT_0)} \right) \times \# I_{\rho,j} \mu_{y_\rho}^u(W_\rho). \end{aligned} \quad (4.74)$$

4.3. The role of additive combinatorics. To proceed, we wish to make use of the oscillations due to the large frequencies $bw_{k,\ell}^i$ to obtain cancellations. First, we note that Lemma 4.13 and the separation between pairs of unstable manifolds with indices in $S_{\rho,j}$ imply that the frequencies $bw_{k,\ell}^i$ have large size. More precisely, the linear functionals $\langle w_{k,\ell}^i, \cdot \rangle : \mathfrak{n}_\alpha^+ \rightarrow \mathbb{R}$ satisfy

$$b^{-9/10} \ll \|\langle w_{k,\ell}^i, \cdot \rangle\| \ll b^{-7/10}. \quad (4.75)$$

Let $\pi : \mathfrak{n}^+ \rightarrow \mathfrak{n}_\alpha^+$ denote the projection parallel to $\mathfrak{n}_{2\alpha}^+$ and note that the integrands on the right side of (4.74) depend only on the \mathfrak{n}_α^+ component of the variable. To simplify notation, we let³

$$\nu_i := \pi_* \mu_{y_\rho^i}^u \Big|_{N_1^+}. \quad (4.76)$$

Remark 4.16. It is worth emphasizing that the linearization provided by Lemma 4.13 only depends on the unstable directions with weakest expansion under the flow. The reason we do so is that our metric on \mathfrak{n}^+ is not invariant by addition when X is not real hyperbolic (it is invariant by the nilpotent group operations), but our non-concentration estimates for the measures μ_\bullet^u only hold for this metric. This in particular means the results of Section 6 do not apply to these measures in this case, which is the reason we work with projections. It is possible to develop the theory in Section 6 for measures and convolutions on nilpotent groups such as N^+ to avoid working with projections, however we believe the approach we adopt here is more amenable to generalizations beyond the algebraic setting of this article.

For $w \in \mathfrak{n}^-$, let

$$\hat{\nu}_i(w) := \int_{\mathfrak{n}_\alpha^+} e^{-i\langle w, v \rangle} d\nu_i(v). \quad (4.77)$$

Note that the total mass of ν_i , denoted $|\nu_i|$, is $\mu_{y_\rho^i}^u(N_1^+)$. Let $\lambda > 0$ be a small parameter to be chosen using Theorem 4.17 below. Define the following set of frequencies where $\hat{\nu}_i$ is large:

$$B(i, k, \lambda) := \left\{ \ell \in I_{\rho, j} : (k, \ell) \in S_{\rho, j} \text{ and } |\hat{\nu}_i(bw_{k, \ell}^i)| > b^{-\lambda} |\nu_i| \right\}. \quad (4.78)$$

Then, splitting the sum over frequencies according to the size of the Fourier transform $\hat{\nu}_i$ and reversing our change variables to go back to integrating over A_i , we obtain

$$\sum_i e^{-\delta t_i} \sum_{(k, \ell) \in S_{\rho, j}} \int_{\mathfrak{n}_\alpha^+} e^{-ib\langle w_{k, \ell}^i, v \rangle} d\nu_i(v) \ll \left(\max_{i, k} \#B(i, k, \lambda) + b^{-\lambda} \#I_{\rho, j} \right) \#I_{\rho, j} \mu_{y_\rho}^u(W_\rho), \quad (4.79)$$

where we again used the estimate $\sum_i \mu_{y_\rho^i}^u(A_i) \ll \mu_{y_\rho}^u(W_\rho)$. The following key counting estimate for $B(i, k, \lambda)$ is deduced from Corollary 6.4. Its proof is given in Section 7.4.

Theorem 4.17. *For every $\varepsilon > 0$, there exists $\lambda > 0$ such that for all i and k , we have*

$$\#B(i, k, \lambda) \ll_\varepsilon b^\varepsilon \left(b^{-\kappa_0/10} + e^{-\kappa_0\gamma(w+jT_0)} \right) e^{\delta\gamma(w+jT_0)},$$

where $\kappa_0 > 0$ is the constant provided by Proposition 4.14.

Combining estimates on oscillatory integrals. Let β_0 and $\kappa_0 > 0$ be as in Propositions 4.12 and 4.14 respectively. In what follows, we assume ε is chosen smaller than $\kappa_0/100$ and that $\lambda \leq \min \{\beta_0, 3/100, \kappa_0/20\}$. Let

$$Q = (b^{-\kappa_0/20} + b^\varepsilon e^{-\kappa_0\gamma(w+jT_0)}) e^{\delta\gamma(w+jT_0)}.$$

Theorem 4.17, combined with (4.63), (4.74) and (4.79), yields:

$$\int_{\mathbb{R} \times W_\rho} \Psi_\rho(t, n) F_\gamma(g_t n y_\rho) d\mu_{y_\rho}^u dt \ll e_{1,0}^*(f) e^{3\beta\alpha j T_0} \mu_{y_\rho}^u(W_\rho) \times \left(b^{-\lambda/2} \#I_{\rho, j} + \sqrt{\#I_{\rho, j} \times Q} \right), \quad (4.80)$$

where we used the elementary inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for any $x, y \geq 0$.

³Note that π is the identity map in the real hyperbolic case.

Our next goal is to estimate the sum of the above bound over ρ . Recall that $\mu_{y_\rho}^u(W_\rho) \asymp \mu_{x_{\rho,\ell}}^u(W_\ell) \asymp \mu_{y_\rho}^u(N_{\ell_b}^+)$ for all $\ell \in I_{\rho,j}$ by Prop. 2.2. Hence, the Cauchy-Schwarz inequality yields

$$\begin{aligned} \sum_{\rho \in \mathcal{P}_b} \mu_{y_\rho}^u(W_\rho) \sqrt{\#I_{\rho,j}} &\ll \left(\sum_{\rho \in \mathcal{P}_b} \mu_{y_\rho}^u(W_\rho) \times \sum_{\rho \in \mathcal{P}_b, \ell \in I_{\rho,j}} \mu_{x_{\rho,\ell}}^u(W_\ell) \right)^{1/2} \\ &\ll e^{O_\beta(\alpha j T_0)} \times \mu_x^u(N_1^+)^{1/2} e^{\delta(\gamma(w+jT_0)/2)}, \end{aligned}$$

where the second inequality follows by Lemma 4.9 and (4.52). By definition of μ_x^u and Prop. 2.4, we have that $\mu_x^u(N_1^+) \gg e^{-\delta \text{dist}(x,o)} \gg V(x)^{O_\beta(1)}$. Hence, we get

$$\sum_{\rho \in \mathcal{P}_b} \mu_{y_\rho}^u(W_\rho) \sqrt{\#I_{\rho,j}} \ll \mu_x^u(N_1^+) \times e^{\delta(\gamma(w+jT_0)/2 + O_\beta(\alpha j T_0))}, \quad (4.81)$$

We also note that a similar argument to (4.52) yields $\sum_{\rho \in \mathcal{P}_b} \mu_{y_\rho}^u(W_\rho) \#I_{\rho,j} \ll e^{\delta\gamma(w+jT_0)} \mu_x^u(N_1^+)$.

Recall that $\lambda/2 \leq \kappa_0/40$. It follows that upon combining the above estimate with (4.80) and (4.81), we obtain the following bound on the sum of the integrals in (4.80):

$$\begin{aligned} e^{-\delta\gamma(w+jT_0)} \sum_{\rho \in \mathcal{P}_b} \int_{\mathbb{R} \times W_\rho} \Psi_\rho(t, n) F_\gamma(g_t n y_\rho) d\mu_{y_\rho}^u dt \\ \ll e_{1,0}^*(f) \mu_x^u(N_1^+) \times e^{O_\beta(\alpha j T_0/2)} \times \left(b^{-\lambda/2} + b^{\varepsilon/2} e^{-\kappa_0 \gamma(w+jT_0)/2} \right), \end{aligned}$$

where we again used the inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$.

Using (4.15) to sum the above error terms over j and w , we obtain

$$O_{T_0, \varepsilon} \left(e_{1,0}^*(f) \mu_x^u(N_1^+) \times \left[\frac{b^{-\lambda/2}}{(a - O_\beta(\alpha))^m} + \frac{b^{\varepsilon/2}}{(a + \kappa_0 \gamma/2)^m} \right] \right). \quad (4.82)$$

4.4. Parameter selection and conclusion of the proof. In this subsection, we finish the proof of Theorem 4.2 assuming Lemma 4.13, Proposition 4.14, and Theorem 4.17.

First, we simplify the bound (4.82). Recall that λ is chosen according to Theorem 4.17 and hence its size depends on ε , however κ_0 is given by Proposition 4.14 and is independent of ε . Moreover, $\gamma = 1/2$, λ and κ_0 are independent of a and α , and we are free to choose the parameter α as small as needed. We also recall that $m = \lceil \log b \rceil$; cf. (4.5). Hence, we may choose a , and ε small enough relative to $\kappa\gamma$ to ensure that $e^{\varepsilon/2} \leq (a + \kappa\gamma/2)/(a + \kappa\gamma/3)$. Using the bound $e^{-\lambda/2} \leq 1/(1 + \lambda/2)$ and taking α small enough, depending on a, β and λ , we get

$$\frac{e^{-\lambda/2}}{a - O_\beta(\alpha)} \leq \frac{1}{a + a\lambda/4}.$$

Hence, taking a small enough so that $a\lambda/4 \leq \kappa\gamma/3$, the error term in (4.82) becomes

$$O \left(\frac{e_{1,0}^*(f) V(x) \mu_x^u(N_1^+)}{(a + a\lambda/4)^m} \right), \quad (4.83)$$

where we used the inequality $V(x) \gg 1$.

Collecting the error terms in Lemma 4.4, (4.9), (4.17), (4.23), (4.33), (4.44), (4.60), and (4.83), and letting $\sigma_\star > 0$ be the minimum of all the gains in these error terms, we obtain

$$e_{1,0}^*(R(z)^m f) \ll \frac{\|f\|_{1,B}^\star}{(a + \sigma_\star)^m}.$$

Letting C_Γ denote the implied constant, this estimate concludes the proof of Theorem 4.2.

5. THE TEMPORAL FUNCTION AND PROOF OF LEMMA 4.13

In this section, we give an explicit formula for the temporal functions $\tau_{k,\ell}$ appearing in Section 4 and prove Lemma 4.13. Our argument is Lie theoretic. We refer the reader to [Kna02, Chapter 1] for background on the material used in this section. Similar results are known more generally outside of the homogeneous setting by more dynamical/geometric arguments building on work of Katok and Burns [Kat94].

5.1. Proximal representations and temporal functions. Let $\rho : G \rightarrow H := \mathrm{SL}_n(\mathbb{R})$ be a proximal irreducible representation of G , i.e., ρ is irreducible and the top eigenspace of $\rho(g_1)$ is one-dimensional. The existence of such a representation is guaranteed by [Tit71]. In what follows, we suppress ρ from the notation and view G as a subgroup of H and view elements of the Lie algebra of G as (traceless) matrices in $\mathfrak{h} = \mathrm{Lie}(H)$.

Let $\{e_i\}$ denote the standard basis of \mathbb{R}^n . Without loss of generality, we assume that e_1 is a top eigenvector for g_1 and denote by $e^\lambda > 1$ the corresponding top eigenvalue. Up to a change of basis, we shall further assume that N^+ (resp. N^-) consists of upper (resp. lower) triangular matrices.

Given a matrix $h \in \mathrm{SL}_n(\mathbb{R})$, we let $\pi_0(h)$ denote its top left entry. In particular, $\pi_0(h) = 1$ for all $h \in N^+ \cup N^-$. Moreover, since ρ is proximal, M acts trivially on the top eigenspace of g_1 , where we recall that M denotes the centralizer of the geodesic flow inside the maximal compact subgroup of G . It follows that $\pi_0(m) = 1$ for all $m \in M$. Finally, we have the simple formula

$$t = \lambda^{-1} \log \pi_0(g_t).$$

These observations will allow us to compute the functions τ_k^i using elementary matrix calculations.

Let $X \in \mathfrak{n}^-$ and $Y \in \mathfrak{n}^+$ be sufficiently close to 0. As the product map $M \times A \times N^+ \times N^- \rightarrow G$ is a diffeomorphism near identity, there exist unique $\tau_X(Y) \in \mathbb{R}$ and $\phi_X(Y) \in \mathfrak{n}^+$ such that

$$\exp(Y) \exp(X) \in N^- M g_{\tau_X(Y)} \exp(\phi_X(Y)), \quad (5.1)$$

where $\exp(Z) = \sum_{n \geq 0} \frac{Z^n}{n!}$. To compute $\tau_X(Y)$, for a matrix M , let $r_1(M)$ and $c_1(M)$ denote its top row and first column respectively. Then, (5.1) shows that $\pi_0(\exp(Y) \exp(X)) = 1 + r_1(Y) \cdot c_1(\exp(X)) + O(\|Y\|^2)$. Hence,

$$\tau_X(Y) = \lambda^{-1} \log(1 + r_1(Y) \cdot c_1(\exp(X)) + O(\|Y\|^2)). \quad (5.2)$$

5.2. Proof of the first assertion of Lemma 4.13. Fix $k \in I_{\rho,j}$ and recall the elements $n_{\rho,k}^- \in N^-$ which were defined by the displacement of the points $x_{\rho,k}$ from y_ρ along N^- inside the flow box B_ρ ; cf. (4.45). We also recall the elements $u_k \in \mathfrak{n}_\alpha^-$ and $s_k \in \mathfrak{n}_{2\alpha}^-$ chosen so that $n_{\rho,k}^- = \exp(u_k + s_k)$. In what follows, we set $X_k := u_k + s_k$. Given $Y \in \mathfrak{n}^+$, we write Y_α and $Y_{2\alpha}$ for its \mathfrak{n}_α^+ and $\mathfrak{n}_{2\alpha}^+$ components respectively.

Recall the vectors $w_i = v_i + r_i \in \mathfrak{n}^+$ defined above Lemma 4.13, where v_i and r_i denoted the \mathfrak{n}_α^+ and $\mathfrak{n}_{2\alpha}^+$ components of w_i respectively. We also recall the return times t_i in (4.66). For $Y \in \mathfrak{n}^+$, let $Y^i = \log(\exp(\mathrm{Ad}(g_{-t_i})(Y)) \exp(w_i)) \in \mathfrak{n}^+$. In particular, Y^i takes the form

$$Y^i = w_i + e^{-t_i} Y_\alpha + e^{-t_i} [Y_\alpha, v_i]/2 + e^{-2t_i} Y_{2\alpha}.$$

In this notation, we have by definition of the functions τ_k (cf. (4.56)) and τ_k^i (cf. (4.66)) that $\tau_k^i(Y) = \tau_k(Y^i)$. Moreover, it follows from the definition of τ_k and (4.53) that

$$g_{t_{\rho,k}} \exp(\phi_{X_k}^{-1}(Y^i)) \exp(X_k) \in N^- M g_{\tau_k(Y^i)} \exp(Y^i).$$

Rearranging this identity and using the fact that g_t normalizes $N^- M = M N^-$, we obtain

$$\exp(Y^i) \exp(-X_k) \in N^- M g_{t_{\rho,k} - \tau_k(Y^i)} \exp(\phi_{X_k}^{-1}(Y^i)),$$

where we used the fact that $\exp(X_k)^{-1} = \exp(-X_k)$. We thus obtain the formula

$$\tau_k(Y^i) = t_{\rho,k} - \tau_{-X_k}(Y^i), \quad (5.3)$$

where the notation is as in (5.1) above.

Define a bilinear form $\langle \cdot, \cdot \rangle : \mathfrak{n}^- \times \mathfrak{n}_\alpha^+ \rightarrow \mathbb{R}$ by

$$\langle X, Y_\alpha \rangle := \lambda^{-1} r_1(Y_\alpha) \cdot c_1(X). \quad (5.4)$$

Recall that our flow boxes B_ρ have radius $\leq b^{-2/3}$ in the unstable direction; cf. (4.28). In particular, $\|w_i\| \ll b^{-2/3}$. Moreover, by (4.68), we have that $e^{-t_i} \ll b^{-7/10}$. Hence, we get that $\|Y^i\| \ll b^{-2/3}$. Note further that since $e^{-t_i} v_i = O(b^{-4/3})$ by (4.68), we also have that $Y^i = w_i + e^{-t_i} Y_\alpha + O(b^{-4/3})$. Finally, note by Lemma 5.1(1) below that for all $n \geq 2$ and $X \in \mathfrak{n}^+$, $r_1(Y_\alpha) \cdot c_1(X^n) = 0$. In particular, since Y_α is strictly upper triangular, $r_1(Y_\alpha) \cdot c_1(\exp(X_k)) = r_1(Y_\alpha) \cdot c_1(X_k)$.

Let $d_k^i := 1 + r_1(w_i) \cdot c_1(X_k)$. Then, using the estimate $\log(1+x) = x + O(|x|^2)$ for x near 0, the above discussion yields

$$\tau_k(Y^i) = t_{\rho,k} - \lambda^{-1} d_k^i + e^{-t_i} \langle X_k, Y_\alpha \rangle + O(b^{-4/3}).$$

Thus, taking $c_{k,\ell}^i = t_{\rho,k} - \lambda^{-1} d_k^i - t_{\rho,\ell} + \lambda^{-1} d_\ell^i$, we obtain

$$\tau_k^i(Y) - \tau_\ell^i(Y) = c_{k,\ell}^i + e^{-t_i} \langle X_k - X_\ell, Y_\alpha \rangle + O(b^{-4/3}),$$

which completes the proof of the first assertion of Lemma 4.13.

5.3. Norm of the bilinear form. The following lemma is needed to prove the second assertion of Lemma 4.13 in the next subsection. Recall that \mathfrak{h} denotes the Lie algebra of $\mathrm{SL}_n(\mathbb{R})$. Let \mathfrak{h}_β denote the $\mathrm{Ad}(g_t)$ -eigenspace with eigenvalue $e^{\beta t}$. In particular, $\mathfrak{n}_{2\alpha}^- \subseteq \mathfrak{h}_{-2\alpha}$.

Lemma 5.1. (1) For every $Y \in \mathfrak{h}_\alpha$ and $Z \in \mathfrak{h}_{-k\alpha}$, $k \geq 2$, we have $r_1(Y) \cdot c_1(Z) = 0$.

(2) The restriction of the linear map $X \mapsto c_1(X)$ to \mathfrak{n}_α^- is injective.

Proof. Let $\theta : \mathrm{Lie}(G) \rightarrow \mathrm{Lie}(G)$ denote a Cartan involution preserving $\mathrm{Lie}(K)$ and acting by $-\mathrm{id}$ on $\mathrm{Lie}(A)$. In particular, θ sends \mathfrak{n}^+ onto \mathfrak{n}^- while respecting their decompositions into $\mathrm{Ad}(g_t)$ -eigenspaces. By [Mos55], we may assume that θ is the restriction to $\mathrm{Lie}(G)$ of the map on \mathfrak{h} that sends each matrix to its negative transpose. Finally, let $B(\cdot, \cdot)$ denote the Killing form on $\mathrm{Lie}(G)$ and recall that the quadratic form $-B(\cdot, \theta(\cdot))$ is positive definite.

To show Item (1), note that, since Z is strictly lower triangular and Y is strictly upper triangular, we have $\pi_0([Y, Z]) = r_1(Y) \cdot c_1(Z)$, where we recall that $\pi_0(\cdot)$ is the top left entry. On the other hand, $[Y, Z]$ is strictly lower triangular since $k \geq 2$ and $[\mathfrak{h}_\alpha, \mathfrak{h}_{-k\alpha}] \subseteq \mathfrak{h}_{-(k-1)\alpha}$. The claim follows.

For Item (2), fix an arbitrary non-zero element $X \in \mathfrak{n}_\alpha^-$ and let $Z = [X, \theta(X)]$. Note that, since θ is a Lie algebra homomorphism, then $\theta(Z) = -Z$. In particular, Z belongs to $\mathrm{Lie}(A)$. Moreover, we claim that $Z \neq 0$, and, hence, $\pi_0(Z) \neq 0$. Indeed, let $\omega \in \mathrm{Lie}(A)$ be such that $[\omega, \cdot]$ fixes \mathfrak{n}_α^+ pointwise. Then, by properties of the Killing form, we obtain

$$B(Z, \omega) = B(X, [\theta(X), \omega]) = -B(X, \theta(X)) \neq 0,$$

since $X \neq 0$ and $-B(\cdot, \theta(\cdot))$ is positive definite. In particular, arguing as in the first part, we have that $\|c_1(X)\|^2 = |\pi_0(Z)| \neq 0$, concluding the proof. \square

5.4. Proof of the second assertion of Lemma 4.13. Fix $(u, s) \in \mathfrak{n}_\alpha^- \times \mathfrak{n}_{2\alpha}^-$. We show that $\sup |\langle u + s, Y \rangle| \gg \|u\|$, where the supremum ranges over all $Y \in \mathfrak{n}_\alpha^+$ with $\|Y\| = 1$. Let $Y = \theta(u)/\|u\|$. Then, by Lemma 5.1(1), we have that

$$\langle u + s, Y \rangle = \lambda^{-1} r_1(Y) \cdot c_1(u).$$

Hence, since $\theta(u)$ is the negative transpose of u , $\langle u + s, Y \rangle = -\lambda^{-1} \|c_1(u)\|^2 / \|u\|$. It follows by Lemma 5.1(2) and equivalence of norms that $\|u\| \asymp \|c_1(u)\|$, concluding the proof of Lemma 4.13.

6. DIMENSION INCREASE UNDER ITERATED CONVOLUTIONS

The goal of this section is to prove that measures that do not concentrate near proper affine subspaces in \mathbb{R}^d become smoother under iterated self-convolutions in the sense of quantitative increase in their L^2 -dimension; cf. Theorem 6.3 below. This result immediately implies Theorem 1.5. As a corollary, we deduce that the Fourier transforms of such measures enjoy polynomial decay outside of a very sparse set of frequencies; cf. Corollary 6.4. In fact, we prove that such results hold for certain *projections* of non-concentrated measures.

Corollary 6.4 provides the key ingredient in the proof of Theorem 4.17 where it is applied to (projections of) conditional measures of the BMS measure. Moreover, the proof of Proposition 4.14 in the case of cusped non-real hyperbolic manifolds requires a polynomial non-concentration estimate near hyperplanes which we deduce from Theorem 6.3; cf. Theorem 6.23.

6.1. General setting. Throughout this section, N denotes a connected nilpotent group, equipped with a right-invariant metric. Affine subspaces of N are defined analogously to Definition 7.1. Given $\rho > 0$, $W \subset N$ and $x \in N$, we write $W^{(\rho)}$ and $B(x, \rho)$ for ρ -neighborhood of W and ρ -ball around x respectively. We fix a surjective homomorphism $\pi : N \rightarrow \mathbb{R}^d$ from N onto \mathbb{R}^d . We assume that N is equipped with a compatible 1-parameter group of dilation automorphisms, which we denote g_t , such that for all $x \in N$, $t \in \mathbb{R}$ and $r > 0$,

$$\pi(g_t \cdot x) = e^t \pi(x), \quad \text{and} \quad \text{dist}(g_t \cdot x, \text{id}) = e^t \text{dist}(x, \text{id}). \quad (6.1)$$

We further fix $\tilde{\Lambda}$ to be a lattice in N (i.e. a discrete cocompact subgroup) and let $D \subset N$ be a fundamental domain for $\tilde{\Lambda}$ containing the identity element. By scaling $\tilde{\Lambda}$ using g_{-t} if necessary, we shall assume without loss of generality that $D \subseteq B(\text{id}, 1)$. Up to composing π with a linear change of basis, we shall assume that

$$\pi(\tilde{\Lambda}) = \mathbb{Z}^d, \quad \pi(D) = [0, 1)^d.$$

Finally, we fix a norm on \mathbb{R}^d and assume that its induced metric is compatible with the metric on N in the sense that there is a uniform $c \geq 1$ such that for all sets E and $\rho > 0$, we have

$$\pi^{-1}(E^{(\rho)}) \subseteq \pi^{-1}(E)^{(c\rho)}. \quad (6.2)$$

Examples. (1) In our application in this article, N will be N^\pm , equipped with the Cygan metric from Section 2.4, and π will be the projection onto the abelianization $N/[N, N] \cong \mathbb{R}^d$ (which is the identity map in the real hyperbolic case where N is abelian).

(2) Another interesting example that falls under our setting is $N = \mathbb{R}^D$ for some $D \geq d$, π is a standard projection, and g_t is a diagonal matrix, where one equips N with a suitable analog of the Cygan metric which satisfies the above scaling properties. In particular, we anticipate that the results of this section will have applications towards the study of fractal geometric properties of self-affine measures and their projections.

6.2. Non-uniform affine non-concentration. We begin by introducing our non-concentration hypothesis, which allows for exceptional sets of points and scales where concentration may happen.

Definition 6.1. Let positive functions λ , φ , and C on $(0, 1]$ be given such that $\varphi(x) \xrightarrow{x \rightarrow 0} 0$. We say a Borel measure μ on N is (λ, φ, C) -*affinely non-concentrated at almost every scale* (or (λ, φ, C) -ANC for short) if the following holds. For every $0 < \varepsilon, \theta \leq 1$, $k \in \mathbb{N}$, and $r \geq C(\theta)$:

- (1) There is an exceptional set $\mathcal{E} = \mathcal{E}(k, \varepsilon, \theta, r) \subset B(\text{id}, 2)$ with $\mu(\mathcal{E}) \leq C(\theta) 2^{-\lambda(\theta)k} \mu(B(\text{id}, 2))$.
- (2) For every $x \in B(\text{id}, 2) \cap \text{supp}(\mu) \setminus \mathcal{E}$, there is a set of good scales $\mathcal{N}(x) \subseteq [0, k] \cap \mathbb{N}$ with $\#\mathcal{N}(x) \geq (1 - \theta)k$.

- (3) For every $x \in B(\text{id}, 2) \cap \text{supp}(\mu) \setminus \mathcal{E}$, every affine subspace $W < N$, every $\ell \in \mathcal{N}(x)$ and $\rho \asymp 2^{-r\ell}$, we have

$$\mu(W^{(\varepsilon\rho)} \cap B(x, \rho)) \leq (\varphi(\theta) + C(\theta)\varphi(\varepsilon))\mu(B(x, \rho)). \quad (6.3)$$

We say μ is ANC at almost every scale when the parameters are understood from context.

In words, this definition says that μ exhibits strong non-concentration near proper subspaces at nearly all scales outside of a small exceptional set, however the size of the exceptional set is allowed to depend on the strength and frequency of non-concentration.

Remark 6.2. (1) Note that we do not require μ to be a probability measure or compactly supported. Instead, we require that non-concentration holds uniformly over all balls centered in a given ball around identity (outside of some exceptional set). This flexibility allows us to avoid edge effects in verifying (6.3) for the restrictions of the PS conditional measures μ_x^u to bounded balls.

- (2) For purposes of following the arguments in this section, there is no harm in considering the example $\lambda(x) = \beta x$ for some $\beta > 0$ and the stronger bound

$$\mu(W^{(\varepsilon\rho)} \cap B(x, 2^{-r\ell})) \leq C(\theta)\varphi(\varepsilon)\mu(B(x, \rho)),$$

in place of (6.3). In fact, the measures μ_x^u are shown to satisfy this bound in Corollary 7.3.

6.3. The L^2 -flattening theorem. For $k \in \mathbb{N}$, let $\Lambda_k := 2^{-k}\mathbb{Z}^d$, and let \mathcal{D}_k be the dyadic partition of \mathbb{R}^d given by translates of $2^{-k}[0, 1]^d$ by Λ_k . For $x \in \mathbb{R}^d$, we denote by $\mathcal{D}_k(x)$ the unique element of \mathcal{D}_k containing x . For a Borel probability measure ν , we define $\nu_k \in \text{Prob}(\Lambda_k)$ to be the scale- k discretization of ν , i.e.

$$\nu_k = \sum_{\lambda \in \Lambda_k} \nu(\mathcal{D}_k(\lambda))\delta_\lambda. \quad (6.4)$$

For any $\mu \in \text{Prob}(\Lambda_k)$ and $0 < q < \infty$, we set $\|\mu\|_q^q := \sum_{\lambda \in \Lambda_k} \mu(\lambda)^q$. The *convolution* $\mu * \nu$ of two probability measures μ and ν on \mathbb{R}^d is defined by

$$\mu * \nu(A) = \int \int 1_A(x + y) d\mu(x) d\nu(y),$$

for all Borel sets $A \subseteq \mathbb{R}^d$. Recall the setup in Section 6.1.

Theorem 6.3. *Let λ , φ and C be given. For every $\varepsilon > 0$, there exist $n, k_1 \in \mathbb{N}$ such that the following holds. Let $\tilde{\mu}$ be a (λ, φ, C) -ANC Borel measure on the group N . Let μ be the projection to \mathbb{R}^d of $\tilde{\mu}|_{B(\text{id}, 1)}$ under π , normalized to be a probability measure. Then, for every $k \geq k_0$, we have*

$$\|\mu_k^{*n}\|_2^2 \ll_{\varepsilon, d, n} 2^{-(d-\varepsilon)k},$$

with implicit constant depending only on d and the non-concentration parameters of $\tilde{\mu}$. In particular, for all $P \in \mathcal{D}_k$, we have

$$\mu_k^{*2n}(P) \ll_{\varepsilon, d, n} 2^{-(d-\varepsilon)k}.$$

The following is a more precise version of Corollary 1.7.

Corollary 6.4. *Let $\tilde{\mu}$ and μ be as in Theorem 6.3. Then, for every $\varepsilon > 0$, there is $\delta > 0$, depending only on the non-concentration parameters of $\tilde{\mu}$, such that for every $T \geq 1$, the set*

$$\left\{ w \in \mathbb{R}^d : \|w\| \leq T \text{ and } |\hat{\mu}(w)| \geq T^{-\delta} \right\}$$

can be covered by $O_\varepsilon(T^\varepsilon)$ balls of radius 1, where $\hat{\mu}$ denotes the Fourier transform of μ . The implicit constant depends only on ε , the diameter of the support of μ and the non-concentration parameters.

Remark 6.5. (1) As noted in [BY], the proof of Theorem 6.3 and Corollary 6.4 goes through under the weaker hypothesis replacing the ball $B(x, \rho)$ in (6.3) with the larger ball $B(x, c\rho)$, for some fixed $c \geq 1$. Indeed, the proof relies on the discretized form of (6.3) in Lemma 6.8, where such weaker inequality naturally appears. This weaker hypothesis is very useful however for applying the above results to non-doubling measures; cf. [BY].

(2) Note that Def. 6.1 requires ANC to hold at points in $B(\text{id}, 2)$, while Theorem 6.3 and Corollary 6.4 concern the restriction of the measure to $B(\text{id}, 1)$. The same arguments work for any two fixed nested balls, after suitably enlarging the implicit constants. In particular, such requirements can be vacuously satisfied if $\tilde{\mu}$ is compactly supported. This flexibility however allows us to avoid certain edge effects when working with restrictions of $\tilde{\mu}$ to a ball.

6.4. Preliminary lemmas on discretized measures. The first lemma asserts that convolution and discretization essentially commute. This justifies the statement of Theorem 6.3.

Lemma 6.6. *Let μ and ν be Borel probability measures on \mathbb{R}^d . Then, for all $q > 1$ and $k \in \mathbb{N}$, we have $\|(\mu * \nu)_k\|_q \asymp_{q,d} \|\mu_k * \nu_k\|_q$.*

Proof. This lemma is a direct consequence of the fact that a ball of radius ρ with $2^{-k-1} < \rho \leq 2^{-k}$, $k \in \mathbb{Z}$, can be covered with $O_d(1)$ elements of \mathcal{D}_k ; cf. [Shm19, Lemma 4.3] for a detailed proof in the case $d = 1$, which readily generalizes to higher dimensions. \square

For each $k \in \mathbb{N}$, set

$$\tilde{\Lambda}_k := g_{-k \log 2}(\tilde{\Lambda}).$$

In particular, $\pi(\tilde{\Lambda}_k) = \Lambda_k = 2^{-k}\mathbb{Z}^d$. Denote by \mathcal{D}_k the partition of N consisting of translates of $g_{-k \log 2}(D)$ by $\tilde{\Lambda}_k$. Using the lattices $\tilde{\Lambda}_k$, we define the scale- k discretization $\tilde{\mu}_k$ of a Borel measure $\tilde{\mu}$ on N analogously to (6.4). Given k and a set $E \subset \mathbb{R}^d$, we let

$$\Lambda_k(E) = \{w \in \Lambda_k : \mathcal{D}_k(w) \cap E \neq \emptyset\}.$$

We define $\tilde{\Lambda}_k(E)$ analogously for $E \subset N$. For a set $E \subseteq \mathbb{R}^d$, we define its scale- k smoothing by

$$E_k := \bigsqcup_{v \in \Lambda_k(E)} \mathcal{D}_k(v). \quad (6.5)$$

Discretizations of subsets of N are defined analogously. The following lemma relates the discretization of a measure $\tilde{\mu}$ on N to the discretization of its projection on \mathbb{R}^d .

Lemma 6.7. *Let $\tilde{\mu}$ and μ be as in Theorem 6.3 and $k \geq 0$. Let $E \subseteq \mathbb{R}^d$ be a Borel set and $F = \pi^{-1}(E_k) \cap B(\text{id}, 1)$. Then, $\tilde{\mu}_k(F_k) \geq \mu_k(E)\tilde{\mu}(B(\text{id}, 1))$.*

Proof. Note that for any measurable set G and measure ν , we have the following properties by the definition of discretizations: $G \subseteq G_k$ and $\nu(G) \leq \nu(G_k) = \nu_k(G_k)$. Let $C = \tilde{\mu}(B(\text{id}, 1))$. Then, applying the previous observation to μ and $\tilde{\mu}$, we get

$$\mu_k(E) \leq \mu(E_k) = C^{-1}\tilde{\mu}(F) \leq C^{-1}\tilde{\mu}(F_k) = C^{-1}\tilde{\mu}_k(F_k).$$

\square

The next lemma shows that affine non-concentration passes to discretizations. In what follows, in light of (6.1), we may and will assume that the constant $c \geq 1$ in (6.2) is chosen large enough so that the following diameter bound holds:

$$\text{diam}(\mathcal{D}_j(v)) \leq c2^{-j}, \quad \forall j \in \mathbb{N}, v \in \Lambda_j \cup \tilde{\Lambda}_j. \quad (6.6)$$

Lemma 6.8. *Let $\tilde{\mu}$ be as in Theorem 6.3. Let $\theta \in (0, 1)$ and a sufficiently large natural number $r \geq C(\theta)$ be given. Then, for all $\varepsilon \geq 2^{-r}$ and sufficiently large $k \in \mathbb{N}$, the scale- kr discretized measure μ_{kr} is affinely non-concentrated in the following sense.*

Let $\mathcal{E} = \mathcal{E}(k, \varepsilon, \theta, r) \subset N$ denote the exceptional set for $\tilde{\mu}$ provided by Definition 6.1.

- (1) *There is an exceptional set $\mathcal{E}_{\text{dis}} = \mathcal{E}_{\text{dis}}(k, \varepsilon, \theta, r)$ with $\tilde{\mu}_{kr}(\mathcal{E}_{\text{dis}}) \leq 2C(\theta)2^{-\lambda(\theta)k}\tilde{\mu}(B(\text{id}, 2))$.*
- (2) *For every $w \in \tilde{\Lambda}_{kr}(B(\text{id}, 1)) \cap \text{supp}(\tilde{\mu}_{kr}) \setminus \mathcal{E}_{\text{dis}}$, there is a set of good scales $\mathcal{N}(w) \subseteq [0, k]$ with $\#\mathcal{N}(w) \geq (1 - \theta)k - O(1)$. Moreover, there is $x \in \mathcal{D}_{kr}(w) \cap \text{supp}(\tilde{\mu}) \setminus \mathcal{E}$ such that $\mathcal{N}(w) = \mathcal{N}(x) \cap [0, k - 1]$, where $\mathcal{N}(x)$ is as in Definition 6.1.*
- (3) *For every $w \in \tilde{\Lambda}_{kr}(B(\text{id}, 1)) \cap \text{supp}(\tilde{\mu}_{kr}) \setminus \mathcal{E}_{\text{dis}}$, every affine subspace $W < N$ and every $\ell \in \mathcal{N}(w)$, setting $\rho_\ell = 2^{-r\ell}$, we have*

$$\tilde{\mu}_{kr}(W^{(\varepsilon\rho_\ell)} \cap \mathcal{D}_{r\ell}(w)) \leq (\varphi(\theta) + C(\theta)\varphi(2\varepsilon/c))\tilde{\mu}(B(w, 3c\rho_\ell)), \quad (6.7)$$

where $c \geq 1$ is as in (6.6).

Proof. Define $\mathcal{E}_{\text{dis}} = \left\{ w \in \tilde{\Lambda}_{kr}(B(\text{id}, 1)) : \tilde{\mu}(\mathcal{D}_{kr}(w) \cap \mathcal{E}) > \tilde{\mu}(\mathcal{D}_{kr}(w))/2 \right\}$. Then, we get

$$\tilde{\mu}_{kr}(\mathcal{E}_{\text{dis}}) = \sum_{w \in \mathcal{E}_{\text{dis}}} \tilde{\mu}(\mathcal{D}_{kr}(w)) < 2 \sum_{w \in \mathcal{E}_{\text{dis}}} \tilde{\mu}(\mathcal{D}_{kr}(w) \cap \mathcal{E}) \leq 2\tilde{\mu}(\mathcal{E}) \leq 2C(\theta)2^{-\lambda(\theta)k}\tilde{\mu}(B(\text{id}, 2)).$$

For each $w \in \tilde{\Lambda}(B(\text{id}, 1)) \cap \text{supp}(\tilde{\mu}_{kr}) \setminus \mathcal{E}_{\text{dis}}$, fix an arbitrary $x \in \text{supp}(\tilde{\mu}) \cap \mathcal{D}_{kr}(w) \setminus \mathcal{E}$. For each such w , since $\mathcal{D}_{kr}(w)$ intersects $B(\text{id}, 1)$, $\mathcal{D}_{kr}(w)$ is contained in $B(\text{id}, 2)$ whenever k is large enough. In particular, $x \in B(\text{id}, 2)$. Set

$$\mathcal{N}(w) = \mathcal{N}(x) \cap [0, k - 2].$$

Then, $\#\mathcal{N}(w) \geq (1 - \theta)k - 2$. For w and x as above, let $\ell \in \mathcal{N}(w)$ and set $\rho_\ell = 2^{-r\ell}$. Then, given any proper affine subspace $W \subset N$, we have

$$\tilde{\mu}_{kr}(W^{(\varepsilon\rho_\ell)} \cap \mathcal{D}_{r\ell}(w)) = \sum_{v \in \tilde{\Lambda}_{kr} \cap W^{(\varepsilon\rho_\ell)} \cap \mathcal{D}_{r\ell}(w)} \tilde{\mu}(\mathcal{D}_{kr}(v)).$$

Next, by (6.6), the cell $\mathcal{D}_{kr}(v)$ has diameter $\leq c2^{-rk}$. In particular, if r is large enough relative to c , since $\varepsilon \geq 2^{-r}$ and $\ell \leq k - 2$, we have that $\mathcal{D}_{kr}(v)$ is contained inside $W^{(2\varepsilon\rho_\ell)}$ for all $v \in W^{(\varepsilon\rho_\ell)}$. Similarly, we have that $\mathcal{D}_{r\ell}(w)$ is contained inside $B(x, 2c\rho_\ell)$. It follows that

$$v \in \tilde{\Lambda}_{kr} \cap W^{(\varepsilon\rho_\ell)} \cap \mathcal{D}_{r\ell}(w) \implies \mathcal{D}_{kr}(v) \subset W^{(2\varepsilon\rho_\ell)} \cap B(x, 2c\rho_\ell).$$

We thus get that

$$\tilde{\mu}_{kr}(W^{(\varepsilon\rho_\ell)} \cap \mathcal{D}_{r\ell}(w)) \leq \tilde{\mu}(W^{(2\varepsilon\rho_\ell)} \cap B(x, 2c\rho_\ell)).$$

Hence, since $\tilde{\mu}$ is affinely non-concentrated, and $x \in B(\text{id}, 2)$ and $\ell \in \mathcal{N}(x)$, we obtain

$$\tilde{\mu}_{kr}(W^{(\varepsilon\rho_\ell)} \cap \mathcal{D}_{r\ell}(w)) \leq (\varphi(\theta) + C(\theta)\varphi(2\varepsilon/c))\tilde{\mu}(B(x, 2c\rho_\ell)).$$

Finally, we observe that since $x \in \mathcal{D}_{kr}(w)$, the ball $B(x, 2c\rho_\ell)$ is contained in $B(w, 3c\rho_\ell)$. Together with the above estimate, this yields (6.7) and concludes the proof. \square

We end this section with the following useful lemma regarding intersection multiplicities.

Lemma 6.9. *Let $C \geq 1$ be given. Then, for all $\ell \in \mathbb{N}$, the balls $\{B(v, C\rho_\ell) : v \in \tilde{\Lambda}_\ell\}$ have intersection multiplicity $O_{N,C}(1)$.*

Proof. Fix some $x \in N$ and let $E(x)$ denote the set of $v \in \tilde{\Lambda}_\ell$ with $x \in B(v, C\rho_\ell)$. Then, $E(x) \subseteq \tilde{\Lambda}_\ell \cap B(v_0, 2C\rho_\ell)$, for any fixed $v_0 \in E(x)$. By right-invariance of the metric, $\#E(x)$ is at most $\#\tilde{\Lambda}_\ell \cap B(\text{id}, 2C\rho_\ell)$. Applying the scaling automorphism $g_{\ell \log 2}$ to the latter set, and using (6.1), we conclude that $\#E(x) \leq \#\tilde{\Lambda} \cap B(\text{id}, 2C)$. The lemma follows by discreteness of the lattice $\tilde{\Lambda}$. \square

6.5. Asymmetric Balog-Szemerédi-Gowers Lemma. The following is the asymmetric version of the Balog-Szemerédi-Gowers Lemma due to Tao and Vu, which is the first key ingredient in the proof of Theorem 6.3. For a finite set $A \subset \mathbb{R}^d$, $|A|$ denotes its cardinality.

Theorem 6.10 (Corollary 2.36, [TV06]). *Let $A, B \subset \mathbb{R}^d$ be finite sets such that $\|1_A * 1_B\|_2^2 \geq 2\alpha|A||B|^2$ and $|A| \leq L|B|$ for some $0 < \alpha \leq 1$ and $L \geq 1$. Let $\varepsilon' > 0$ be given. Then, there exist sets $A' \subseteq A$ and $B' \subseteq B$ such that*

- (1) A' and B' are sufficiently dense: $|A'| \gg_{\varepsilon'} \alpha^{O_{\varepsilon'}(1)} L^{-\varepsilon'} |A|$ and $|B'| \gg_{\varepsilon'} \alpha^{O_{\varepsilon'}(1)} L^{-\varepsilon'} |B|$.
- (2) A' is approximately invariant by B' : $|A' + B'| \ll_{\varepsilon'} \alpha^{-O_{\varepsilon'}(1)} L^{\varepsilon'} |A'|$.

Remark 6.11. The quoted result is stated in terms of the additive energy $E(A, B)$ in *loc. cit.*, which is nothing but $\|1_A * 1_B\|_2^2$.

6.6. Hochman's inverse theorem for entropy. In order to be able to bring our affine non-concentration hypothesis into play, we will need to convert the approximate additive invariance provided by the Balog-Szemerédi-Gowers Lemma into exact additive obstructions to flattening under convolution, i.e. affine subspaces. Our key tool for this step is Hochman's inverse entropy theorem for convolutions of measures. We need some notation before stating the result.

For a Borel probability measure ν on \mathbb{R}^d , the entropy $H_k(\nu)$ of ν at scale k is defined to be

$$H_k(\nu) := -\frac{1}{k} \sum_{P \in \mathcal{D}_k} \nu(P) \log_2 \nu(P).$$

By concavity of log and Jensen's inequality, we have the following elementary inequality

$$H_k(\nu) \leq \frac{\log_2 \# \{P \in \mathcal{D}_k : \nu(P) \neq 0\}}{k}. \quad (6.8)$$

It also follows from Jensen's inequality that the above inequality becomes equality if and only if ν gives equal weights to the elements P of \mathcal{D}_k with $\nu(P) \neq 0$.

Given a Borel probability measure ν on \mathbb{R}^d and $z \in \mathbb{R}^d$ with $\nu(\mathcal{D}_k(z)) > 0$, we define the component measure $\nu^{z,k}$ by

$$\int f d\nu^{z,k} := \frac{1}{\nu(\mathcal{D}_k(z))} \int_{\mathcal{D}_k(z)} f(T(y)) d\nu(y),$$

where $T : \mathcal{D}_k(z) \rightarrow \mathcal{D}_k(0)$ is the affine map given by composing scaling by 2^k with translation by the element of Λ_k sending $\mathcal{D}_k(z)$ to $\mathcal{D}_k(0)$.

Given a Borel subset $\mathcal{P} \subseteq \text{Prob}(\mathbb{R}^d)$ and $k \in \mathbb{N}$, we define

$$\mathbb{P}_{0 \leq i \leq k}(\nu^{z,i} \in \mathcal{P}) := \frac{1}{k+1} \sum_{i=0}^k \int 1_{\mathcal{P}}(\nu^{z,i}) d\nu(z). \quad (6.9)$$

Given a linear subspace $0 \leq V \leq \mathbb{R}^d$, $\varepsilon > 0$ and a probability measure ν , we say that ν is (V, ε) -concentrated if there is a translate L of V such that $\nu(L^{(\varepsilon)}) > 1 - \varepsilon$. We say that ν is (V, ε, m) -saturated for a given $m \in \mathbb{N}$ if

$$H_m(\nu) \geq H_m(\pi_W \nu) + \dim V - \varepsilon, \quad (6.10)$$

where $W = V^\perp$ and $\pi_W \nu$ is the pushforward of ν under the orthogonal projection to W .

Theorem 6.12 (Theorem 2.8, [Hoc15]). *For every $\varepsilon, R > 0$ and $r \in \mathbb{N}$, there are $\sigma > 0$ and $m_0, k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ and Borel probability measures ν and μ on $[-R, R]^d$ satisfying*

$$H_{kr}(\mu * \nu) < H_{kr}(\nu) + \sigma,$$

there exists a sequence of subspaces $0 \leq V_0, \dots, V_k \leq \mathbb{R}^d$ such that

$$\mathbb{P}_{0 \leq i \leq k} \left(\begin{array}{l} \mu^{x,ir} \text{ is } (V_i, \varepsilon) - \text{concentrated and} \\ \nu^{x,ir} \text{ is } (V_i, \varepsilon, m_0) - \text{saturated} \end{array} \right) > 1 - \varepsilon.$$

Remark 6.13. Theorem 6.12 is stated in [Hoc15] in the case $r = 1$. However, the extension to general r is rather routine since it roughly corresponds to working in base 2^r in place of base 2.

6.7. Flattening of discretized measures. The following quantitative result is the main ingredient in the proof of Theorem 6.3.

Proposition 6.14. *Let positive functions λ, φ , and C on $(0, 1]$ be given. Then, for every $0 < \gamma < 1$, there exist $\eta > 0$, and $r \in \mathbb{N}$, depending on γ, λ, C , and φ , such that the following holds.*

For every $n \geq 0$, there exists $k_1 = k_1(\lambda, \varphi, C, \gamma) \in \mathbb{N}$, so that the following hold for all $k \geq k_1$ and any probability measure supported on $2^{-kr}\mathbb{Z}^d \cap B(0, 2^n)$ and satisfying

$$\|\nu\|_2^2 > 2^{-(1-\gamma)dkr}. \quad (6.11)$$

Let $\tilde{\mu}$ be a (λ, φ, C) -ANC Borel measure on N . Let μ be the projection to \mathbb{R}^d of $\tilde{\mu}|_{B(\text{id}, 2)}$ under π , normalized to be a probability measure. Then,

$$\|\mu_{kr} * \nu\|_2 \leq 2^{-\eta kr} \|\nu\|_2. \quad (6.12)$$

This proposition says that the convolution of an arbitrary measure ν with a non-concentrated measure causes ν to “spread out”, i.e. leads to a quantitative reduction in the ℓ^2 norm of ν , unless $\|\nu\|_2$ is already very close to 0.

6.7.1. From measures to sets. The remainder of this subsection is dedicated to the proof of Proposition 6.14. Let $\gamma > 0$ and $\eta > 0$ be small parameters and $r, k \in \mathbb{N}$ be large integers to be specified over the course of the proof. We frequently assume that γ is sufficiently small so that various properties hold and the values of η, r and k will depend only on γ and the non-concentration parameters. Suppose towards a contradiction that (6.11) holds but (6.12) fails.

We first translate the failure of (6.12) from measures to indicator functions of certain sets using standard arguments. This allows us to apply the Balog-Szemerédi-Gowers Lemma.

Lemma 6.15 (Lemma 3.3, [Shm19]). *For every $\eta > 0$ and $n \geq 0$, the following holds for all large enough ℓ . Let μ and ν be probability measures such that $\text{supp}(\mu) \subseteq \Lambda_\ell \cap [0, 1]^d$ and $\text{supp}(\nu) \subseteq \Lambda_\ell \cap B(0, 2^n)$. Assume that $\|\mu * \nu\|_2$ is at least $2^{-\eta\ell} \|\nu\|_2$. Then, there exist $j, j' \leq 4\eta\ell$ such that*

$$A := \left\{ x \in \Lambda_\ell : 2^{-j-1} \|\nu\|_2^2 < \nu(x) \leq 2^{-j} \|\nu\|_2^2 \right\}, \quad (6.13)$$

$$B := \left\{ x \in \Lambda_\ell : 2^{-j'-1-d\ell} < \mu(x) \leq 2^{-j'-d\ell} \right\} \quad (6.14)$$

satisfy

- (1) $\|1_A * 1_B\|_2^2 \geq 2^{-4\eta\ell} |A| |B|^2$,
- (2) $\|\nu|_A\|_2 \geq 2^{-2\eta\ell} \|\nu\|_2$, and
- (3) $\mu(B) \geq 2^{-2\eta\ell}$.

In particular, there exists a subset $A_0 \subseteq A$ such that

- (1) A_0 is contained in $w + [0, 1]^d$, for some $w \in \mathbb{Z}^d$.
- (2) $\|1_{A_0} * 1_B\|_2^2 \geq 2^{-5\eta\ell} |A| |B|^2$.

Proof. The properties of A and B were proved in [Shm19] for measures on \mathbb{R} and for $n = 0$, however the short argument, based on the pigeonhole principle, goes through for general d and n with minimal modifications.

To find A_0 with the claimed properties, let $A = \sqcup_{w \in \mathbb{Z}^d} A_w$ be the partition of A defined by $A_w = A \cap w + [0, 1)^d$. Since $A \subset B(0, 2^n)$, we have that $\#\{w : A_w \neq \emptyset\} \ll 2^{dn}$. By linearity of convolution, the triangle inequality and Cauchy-Schwarz, we get

$$2^{-4\eta\ell} |A||B|^2 \leq \|1_A * 1_B\|_2^2 \leq \left(\sum_w \|1_{A_w} * 1_B\|_2 \right)^2 \ll 2^{dn} \sum_w \|1_{A_w} * 1_B\|_2^2.$$

The corollary follows for all ℓ large enough, depending on d, n , and η , by taking $A_0 = A_w$, for w such that $\|1_{A_w} * 1_B\|_2^2$ is maximal. \square

6.7.2. From ℓ^2 -concentration to entropy concentration. Let $A_0 \subseteq A$ and B be as in Lemma 6.15, applied with $\ell = kr$ and $\mu = \mu_{kr}$. Taking η small enough, we get by (6.11), the definition of A , and Chebyshev's inequality that

$$|A_0| \leq |A| \leq 2^{4\eta kr + 1 + (1-\gamma)dkr} \leq 2^{(1-\gamma/2)dkr + 1}. \quad (6.15)$$

We now apply Theorem 6.10 with $A = A_0$, $\alpha = 2^{-5\eta kr - 1}$, $L = \max\{1, |A_0|/|B|\}$, and

$$0 < \varepsilon' < 1$$

a parameter to be chosen small enough depending on ε . Let $A' \subseteq A_0$ and $B' \subseteq B$ be the sets provided by Theorem 6.10.

Let ν' and μ' be the uniform probability measures supported on A' and B' respectively. Combining the above estimate with (6.8), we obtain

$$H_{kr}(\mu' * \nu') \leq \frac{\log_2 |A' + B'|}{kr} \leq \frac{\log_2 |A'|}{kr} + O_{\varepsilon'}(\eta) + \log_2 L^{\varepsilon'} / kr.$$

Since ν' is the uniform measure on A' , the remark following (6.8) thus implies that

$$H_{kr}(\mu' * \nu') \leq H_{kr}(\nu') + O_{\varepsilon'}(\eta) + \log_2 L^{\varepsilon'} / kr.$$

By (6.15), we have $\log_2 L^{\varepsilon'} \leq \varepsilon' \log_2 |A| \leq \varepsilon'((1-\gamma/2)dkr + 1)$.

Recall from Lemma 6.15 A_0 , and hence the support of ν' , is contained in a box of the form $w + [0, 1)^d$. Moreover, the above inequality remains unchanged by translating ν' . Hence, for the purposes of applying Theorem 6.12, we may without loss of generality assume in the sequel that

$$\text{supp}(\nu') \subset [0, 1)^d.$$

Let $R = O(1)$ be such that $[0, 1)^d$ is contained in the R -ball around the origin. Let $\sigma > 0$ and $k_0 \in \mathbb{N}$ be the parameters provided by Theorem 6.12 applied with this R and

$$\varepsilon = 2^{-r}.$$

We shall assume that k is chosen to be larger than k_0 . Hence, taking ε' small enough (depending on σ) and η small enough (depending on ε' and σ), we obtain

$$H_{kr}(\mu' * \nu') < H_{kr}(\nu') + \sigma. \quad (6.16)$$

We show that the conclusion of Theorem 6.12 is incompatible with the non-concentration properties of the measure μ . Let V_0, \dots, V_k be the subspaces provided by Theorem 6.12 and

$$\mathcal{S} = \left\{ 0 \leq i \leq k : V_i = \mathbb{R}^d \right\}.$$

We begin by showing that a significant proportion of the V_i 's are proper subspaces. Intuitively, being \mathbb{R}^d -saturated on most scales means the measure ν is close to being absolutely continuous to Lebesgue on \mathbb{R}^d in the sense that its ℓ^2 -norm would be very close to 2^{-dk} . This would contradict (6.11).

Lemma 6.16. *If ε is small enough and k is large enough depending on γ , then $\#\mathcal{S} < (1 - \gamma/10)k$.*

Proof. Let $\gamma_1 = \gamma/10$ and suppose that $\#\mathcal{S} \geq (1 - \gamma_1)k$. Then, Theorem 6.12 and the definition of saturation (cf. (6.10)) imply that

$$\frac{1}{k+1} \sum_{i=0}^k \int H_{m_0}((\nu')^{z, ir}) d\nu'(z) \geq (1 - \gamma_1)(1 - \varepsilon)(d - \varepsilon) = (1 - \gamma_1)d - O(\varepsilon).$$

By [Hoc14, Lemma 3.4]⁴, this yields the following estimate on $H_{kr}(\nu')$:

$$H_{kr}(\nu') \geq (1 - \gamma_1)d - O(\varepsilon) - O_r\left(\frac{m_0}{k}\right) \geq (1 - \gamma_1)d - O(\varepsilon),$$

where the second inequality holds whenever k is large enough depending on r and m_0 . Moreover, by the remark following (6.8), we have $H_{kr}(\nu') = \log_2 |A'|/kr \leq \log_2 |A|/kr$. Hence, we obtain that $|A| \geq 2^{((1-\gamma_1)d - O(\varepsilon))kr}$. This contradicts (6.15) when ε is small enough compared to γ . \square

6.7.3. Concentration of large sets at many scales. Roughly speaking, our strategy is as follows. Armed with Lemma 6.16, we show that the concentration provided by Theorem 6.12 holds on a set of relatively large measure and on a definite proportion of scales. On the other hand, the non-concentration property of μ and induction on scales shows that such set must have very small measure, yielding a contradiction.

Recall that $\pi : N \rightarrow \mathbb{R}^d$ is our fixed surjective homomorphism. Let $0 < \gamma_2 < \gamma/40$ be a small parameter to be chosen depending only on γ . Let \mathcal{E}_{dis} be the exceptional set provided by Lemma 6.8 for our choices of k, r , and with $\theta = \gamma_2$ and $2c\varepsilon$ in place of ε , where c is the constant in (6.2). We set $\gamma_3 = \gamma/40 - \gamma_2$.

Lemma 6.17. *Suppose ε is small enough depending on γ , ε' is small enough depending on ε , η is small enough depending on ε' , and k is large enough depending on all the previous parameters.*

Then, there exist a set $F \subseteq \tilde{\Lambda}_{kr}(B(\text{id}, 1)) \subset N$ and a 2-separated set of scales $\ell_1 < \ell_2 < \dots < \ell_m$, where $m = \lceil \gamma_3 k \rceil$ such that the following hold for every $1 \leq i \leq m$:

- (1) $\tilde{\mu}_{kr}(F) \geq 2^{-\sqrt{\varepsilon'}kr - k - 1}$.
- (2) For every $w \in \tilde{\Lambda}_{r\ell_i}(F)$, there exists an affine subspace \tilde{V}_w such that $F \cap \mathcal{D}_{r\ell_i}(w) \subseteq \tilde{V}_w^{(c\varepsilon\rho_{\ell_i})}$, where $\rho_i = 2^{-ir}$ and $c \geq 1$ is as in (6.2).
- (3) \tilde{V}_w is a proper affine subspace for every $w \in \tilde{\Lambda}_{r\ell_i}(F)$.
- (4) F is disjoint from the exceptional set for non-concentration, i.e. $F \cap \mathcal{E}_{\text{dis}} = \emptyset$.
- (5) ℓ_i is a good scale for non-concentration at every point in F , i.e. $\ell_i \in \mathcal{N}(x)$ for all $x \in F$.

Remark 6.18. The proof of Lemma 6.17 in fact shows that for each fixed scale ℓ_i , the projection of the spaces $\{\tilde{V}_w : w \in \tilde{\Lambda}_{r\ell_i}(F)\}$ to \mathbb{R}^d are all parallel to one another.

We begin by deriving a lower bound on the measure of B' with respect to our original discretized measure μ_{kr} (not μ'). Recall the parameter ε' chosen above (6.16).

Lemma 6.19. *If η is chosen sufficiently small depending on ε' , then for all sufficiently large k ,*

$$\mu_{kr}(B') \geq 2^{-2d\varepsilon'kr}.$$

Proof. Recall that the set B was defined in (6.13) and $B' \subseteq B$ is provided by Theorem 6.10 with $\alpha = 2^{-4\eta kr - 1}$ and $L = \max\{1, |A|/|B|\}$. We calculate using Lemma 6.15 and Theorem 6.10:

$$\mu_{kr}(B') = \sum_{u \in B'} \mu(u) \geq 2^{-j' - dkr - 1} |B'| \gg_{\varepsilon'} 2^{-j' - dkr} 2^{-O_{\varepsilon'}(\eta kr)} L^{-\varepsilon'} |B| \geq 2^{-j' - dkr} 2^{-O_{\varepsilon'}(\eta kr)} |B| |A|^{-\varepsilon'}.$$

By (6.15), we have that $|A|^{\varepsilon'} \ll 2^{d\varepsilon'kr}$. Moreover, Lemma 6.15 implies that

$$2^{-2\eta kr} \leq \mu_{kr}(B) \leq 2^{-j' - dkr} |B|.$$

⁴The cited result is stated for step-size $r = 1$, however its short proof extends to work for any r with minor changes.

The lemma then follows once η is chosen sufficiently small depending on ε' . \square

Next, we define the following set of scales where the concentration provided by Theorem 6.12 gives non-trivial information:

$$\mathcal{C} := \{0, \dots, k\} \setminus \mathcal{S} = \left\{0 \leq i \leq k : V_i \not\leq \mathbb{R}^d\right\}.$$

By Lemma 6.16, we know that

$$|\mathcal{C}| \geq \gamma_1 k, \quad \gamma_1 = \gamma/10. \quad (6.17)$$

Our next goal is to transfer the concentration information provided by Theorem 6.12 for μ' to the measure μ . To do so, we convert the probabilistic concentration provided in the theorem into geometric containment into subspace neighborhoods.

Recall that $\{\mathcal{D}_\ell : \ell \in \mathbb{N}\}$ is a refining sequence of dyadic partitions of \mathbb{R}^d and $\Lambda_\ell = 2^{-\ell}\mathbb{Z}^d$. For $i \in \mathcal{C}$ and $w \in \Lambda_{ir}$, let $z \in \mathcal{D}_{ir}(w)$ be such that for $V_w := V_i + z$, we have

$$\mu'(V_w^{(\varepsilon\rho_i)} \cap \mathcal{D}_{ir}(w)) \geq (1 - \varepsilon)\mu'(\mathcal{D}_{ir}(w)). \quad (6.18)$$

If no such z exists, we let $V_w = V_i + w$. Denote by Q_i the set of concentrated points at scale ir , i.e.,

$$Q_i = \bigcup_{w \in \Lambda_{ir}} V_w^{(\varepsilon\rho_i)} \cap \mathcal{D}_{ir}(w).$$

For every $w \in \tilde{\Lambda}_{ir}$, let \tilde{V}_w denote the affine space $\pi^{-1}(V_\pi(w))$, where $\pi : N \rightarrow \mathbb{R}^d$ is our fixed surjective homomorphism. For $x \in \mathbb{R}^d$, we set

$$\mathcal{C}(x) = \{i \in \mathcal{C} : x \in Q_i\}.$$

In particular, for $x \in B'$, $\mathcal{C}(x)$ consists of scales at which x witnesses the concentration of B' .

Lemma 6.20. *If ε is small enough and k is large enough, depending on γ , then the subset*

$$B'' = \{x \in B' : |\mathcal{C}(x)| \geq |\mathcal{C}|/2\} \quad (6.19)$$

satisfies $\mu_{kr}(B'') \geq 2^{-3d\varepsilon'kr}$.

Proof. Let $E = B' \setminus B''$. First, we give an upper bound on the measure of E with respect to μ' . Let $Q_i^c = \mathbb{R}^d \setminus Q_i$. Then, the concentration provided by Theorem 6.12 implies that

$$\mathbb{P}_{0 \leq i \leq k}((\mu')^{x, ir} \text{ is } (V_i, \varepsilon) - \text{concentrated}) > 1 - \varepsilon.$$

To unpack the above inequality, let us denote by $\Theta_{ir} \subseteq \Lambda_{ir}$ the subset consisting of those $w \in \Lambda_{ir}$ for which (6.18) holds. For every $x \in \mathbb{R}^d$ and $1 \leq i \leq k$, let $w(x) \in \Lambda_{ir}$ be such that $\mathcal{D}_{ir}(x) = w(x) + 2^{-ir}[0, 1]^d$. In this notation, the above inequality reads

$$\sum_{0 \leq i \leq k} \mu' \left(x \in \mathbb{R}^d : \mu'(V_{w(x)}^{(\varepsilon\rho_i)} \cap \mathcal{D}_{ir}(x)) \geq (1 - \varepsilon)\mu'(\mathcal{D}_{ir}(x)) \right) > (1 - \varepsilon)(k + 1)$$

Note that if $\mu'(V_{w(x)}^{(\varepsilon\rho_i)} \cap \mathcal{D}_{ir}(x)) \geq (1 - \varepsilon)\mu'(\mathcal{D}_{ir}(x))$ holds for some x and i , then the inequality holds for all $y \in \mathcal{D}_{ir}(x)$ in place of x . Hence, we get

$$\begin{aligned} (1 - \varepsilon)k &< \sum_{0 \leq i \leq k} \mu' \left(x \in \mathbb{R}^d : \mu'(V_{w(x)}^{(\varepsilon\rho_i)} \cap \mathcal{D}_{ir}(x)) \geq (1 - \varepsilon)\mu'(\mathcal{D}_{ir}(x)) \right) \\ &= \sum_{0 \leq i \leq k} \sum_{w \in \Theta_{ir}} \mu'(\mathcal{D}_{ir}(w)) \\ &\leq (1 - \varepsilon)^{-1} \sum_{0 \leq i \leq k} \sum_{w \in \Theta_{ir}} \mu'(\mathcal{D}_{ir}(w) \cap V_w^{(\varepsilon\rho_i)}) \leq (1 - \varepsilon)^{-1} \sum_{0 \leq i \leq k} \mu'(Q_i). \end{aligned}$$

It follows that $\int \sum_{i \in \mathcal{C}} \mathbb{1}_{Q_i^c}(x) d\mu'(x) < 2\varepsilon k$. On the other hand, we have by (6.17) that

$$\int \sum_{i \in \mathcal{C}} \mathbb{1}_{Q_i^c}(x) d\mu'(x) \geq \int_E \sum_{i \in \mathcal{C}} \mathbb{1}_{Q_i^c}(x) d\mu'(x) \geq |\mathcal{C}| \mu'(E)/2 \geq \gamma k \mu'(E)/20.$$

Recalling that μ' is the uniform measure on B' , we can assert that these inequalities imply that $|B''| \geq (1 - 40\varepsilon/\gamma)|B'|$. Hence, the assertion of the lemma follows from Lemma 6.15 by the same argument as in the proof of Lemma 6.19. \square

6.7.4. Lifting concentration to N and proof of Lemma 6.17. Note that the scales $\mathcal{C}(x)$ may vary with x . Similarly, the scales at which our affine non-concentration hypothesis holds also vary from point to point. To arrive at a contradiction, we partition B'' into sets where there is a fixed subset of scales of \mathcal{C} at which the aforementioned phenomena hold simultaneously and find an upper bound on the measure of each piece separately.

Let $0 < \gamma_2 < \gamma_1/4$ be a small parameter to be chosen depending only on γ . Let \mathcal{E}_{dis} be the exceptional set provided by Lemma 6.8 for our choices of k, r , and with $\theta = \gamma_2$ and 2ε in place of ε . By taking $r \geq C(\gamma_2)$ large enough, then Lemma 6.8 implies that $\tilde{\mu}_{kr}(\mathcal{E}_{\text{dis}}) \leq 2C(\gamma_2)2^{-\lambda(\gamma_2)k}$. Recall the definition of smoothed scale- k sets in (6.5) and let

$$B''' = \tilde{\Lambda}_{kr}(E) \setminus \mathcal{E}_{\text{dis}} \subseteq \tilde{\Lambda}_{kr}(B(\text{id}, 1)), \quad \text{where } E = \pi^{-1}(B''_{kr}) \cap B(\text{id}, 1).$$

Then, by Lemma 6.7, taking ε' small enough depending on r and $\lambda(\gamma_2)$, we can ensure that

$$\tilde{\mu}_{kr}(B''') \geq 2^{-3d\varepsilon'kr} - 2C(\gamma_2)2^{-\lambda(\gamma_2)k} \tilde{\mu}(B(\text{id}, 2)) \geq 2^{-k\sqrt{\varepsilon'}}, \quad (6.20)$$

for all large enough k . Recall the sets of good scales $\mathcal{N}(\cdot)$ provided by Lemma 6.8. By a slight abuse of notation, for $x \in B'''$, we let

$$\mathcal{C}(x) = \mathcal{C}(\pi(x)).$$

We define

$$\mathcal{G}(x) = \text{maximal 2-separated subset of } \mathcal{C}(x) \cap \mathcal{N}(x).$$

By (6.17) and the definition of B'' in (6.19), setting $\gamma_3 = \gamma_1/4 - \gamma_2$, we also have

$$|\mathcal{G}(x)| \geq ((\gamma_1/2 - \gamma_2)k - 2)/2 \geq \gamma_3 k, \quad \forall x \in B''',$$

where the second inequality holds whenever k is large enough.

Given $\varpi \subseteq \{0, \dots, k\}$, we let

$$B'''_{\varpi} := \{x \in B''' : \varpi \subseteq \mathcal{G}(x)\}.$$

Then, the sets $\{B'''_{\varpi} : |\varpi| = \lceil \gamma_3 k \rceil\}$ provide a cover of B''' . Hence, we have that

$$\tilde{\mu}_{kr}(B''') \leq \sum_{|\varpi| = \lceil \gamma_3 k \rceil} \tilde{\mu}_{kr}(B'''_{\varpi}). \quad (6.21)$$

Fix a set $\varpi \subset [0, k] \cap \mathbb{N}$ for which $\tilde{\mu}_{kr}(B'''_{\varpi})$ is maximal and let $F = B'''_{\varpi}$. Since the sum in (6.21) has most 2^{k+1} terms, (6.20) implies that $\tilde{\mu}_{kr}(F) \geq 2^{-\sqrt{\varepsilon'}k-k-1}$.

It remains to prove that the set of scales given by ϖ satisfy items (2) and (3) of the lemma. We need the following observation regarding compatibility of dyadic partitions under our projection.

Lemma 6.21. *Let $\ell \geq 0$, $\tilde{w} \in \tilde{\Lambda}_{\ell}(F)$ and $w = \pi(\tilde{w})$. Then, $\pi(F \cap \mathcal{D}_{\ell}(\tilde{w})) \subseteq B'' \cap \mathcal{D}_{\ell}(w)$.*

Proof. For all ℓ , we have $\mathcal{D}_{\ell}(\tilde{w}) = g_{-\ell \log 2}(D) \cdot \tilde{w}$ by definition, and hence by (6.1), we get

$$\pi(\mathcal{D}_{\ell}(\tilde{w})) = 2^{-\ell} \pi(D) + w = \mathcal{D}_{\ell}(w).$$

Next, we show that $\pi(F) \subseteq B''$. Let $U = B(\text{id}, 1)$ and fix some $x \in F$. Then, $x \in \tilde{\Lambda}_{kr}(\pi^{-1}(B''_{kr}) \cap U)$. Hence, the intersection $\mathcal{D}_{kr}(x) \cap \pi^{-1}(B''_{kr}) \cap U$ is non-empty. Let y be a point in this intersection,

so that $y \in B''_{kr} = \sqcup_{v \in B''} \mathcal{D}_{kr}(v)$, where we used that B'' is a subset of Λ_{kr} . Letting $v \in B''$ be such that $y \in \mathcal{D}_{kr}(v)$, it follows that $\pi(\mathcal{D}_{kr}(x))$ intersects $\mathcal{D}_{kr}(v)$. On the other hand, we have shown that $\pi(\mathcal{D}_{kr}(x)) = \mathcal{D}_{kr}(\pi(x))$. Since $\pi(x) \in \Lambda_{kr}$, it follows that $\pi(x) = v$, concluding the proof. \square

Let $\ell \in \varpi$ and $\tilde{w} \in \tilde{\Lambda}_{r\ell}(F)$. Let $x \in F \cap \mathcal{D}_{r\ell}(\tilde{w})$ and $w = \pi(\tilde{w})$. Then, $\pi(x) \in \mathcal{D}_{r\ell}(w) \cap B''$ by Lemma 6.21, and $\ell \in \mathcal{C}(\pi(x))$ by definition. Hence, by definition of B'' in (6.19), we have that $\pi(x) \in V_w^{(\varepsilon\rho_\ell)}$. It follows by (6.2) that $x \in \tilde{V}_{\tilde{w}}^{(c\varepsilon\rho_\ell)}$. Moreover, since $\ell \in \mathcal{C}(\pi(x))$, we have that V_w is a proper subspace, and hence so is $\tilde{V}_{\tilde{w}}$. This completes the proof of Lemma 6.17.

6.7.5. ANC implies a contradiction to Lemma 6.17. In this section, we complete the proof of Proposition 6.14 by showing that the ANC condition gives a contradiction to Lemma 6.17 via an induction on scales argument showing that the multiscale structure of F given in the lemma implies that it has very small measure.

Let F , ℓ_i , and \tilde{V}_w be as in Lemma 6.17. We recall that $\tilde{\Lambda}_\ell(F)$ denotes those elements $v \in \tilde{\Lambda}_\ell$ for which the corresponding cells $\mathcal{D}_\ell(v)$ intersect F non-trivially. As a first step, we have the following basic estimate that will allow us to proceed by induction on scales:

$$\tilde{\mu}_{kr}(F) \leq \sum_{v \in \tilde{\Lambda}_{r\ell_m}(F)} \tilde{\mu}_{kr}(\mathcal{D}_{r\ell_m}(v)) = \sum_{w \in \tilde{\Lambda}_{r\ell_{m-1}}(F)} \sum_{\substack{v \in \tilde{\Lambda}_{r\ell_m}(F) \\ \mathcal{D}_{r\ell_m}(v) \subset \mathcal{D}_{r\ell_{m-1}}(w)}} \tilde{\mu}_{kr}(\mathcal{D}_{r\ell_m}(v)). \quad (6.22)$$

Recall by (6.6) that the diameter of each element of \mathcal{D}_n is at most $c\rho_n$ for a fixed uniform constant $c \geq 1$. Moreover, since the ℓ_i 's are 2-separated and $\varepsilon = 2^{-r}$, we may assume that r is large enough, depending on c so that

$$c\rho_{\ell_j} \leq \varepsilon\rho_{\ell_i}/10, \quad \forall 1 \leq i < j \leq m. \quad (6.23)$$

Hence, if $\tilde{V}_w^{(c\varepsilon\rho_{\ell_i})}$ intersects a box $\mathcal{D}_{r\ell_{i+1}}(v)$ non-trivially, then we have

$$\mathcal{D}_{r\ell_{i+1}}(v) \subseteq \tilde{V}_w^{(2c\varepsilon\rho_{\ell_i})}. \quad (6.24)$$

This containment, along with item (2) of Lemma 6.17, imply that for every $1 \leq i < m$ and $w \in \tilde{\Lambda}_{r\ell_i}(F)$, we have that

$$\sum_{\substack{v \in \tilde{\Lambda}_{r\ell_{i+1}}(F) \\ \mathcal{D}_{r\ell_{i+1}}(v) \subset \mathcal{D}_{r\ell_i}(w)}} \tilde{\mu}_{kr}(\mathcal{D}_{r\ell_{i+1}}(v)) \leq \tilde{\mu}_{kr}(\tilde{V}_w^{(2c\varepsilon\rho_{\ell_i})} \cap \mathcal{D}_{r\ell_i}(w)).$$

Recall that $\tilde{\mu}$ satisfies the ANC condition in Def. 6.1. Hence, for all i and $w \in \tilde{\Lambda}_{r\ell_i}(F)$, items (3), (4) and (5) of Lemma 6.17, along with Lemma 6.8 imply that

$$\tilde{\mu}_{kr} \left(\tilde{V}_w^{(2c\varepsilon\rho_{\ell_i})} \cap \mathcal{D}_{r\ell_i}(w) \right) \leq \delta(\gamma_2, \varepsilon) \tilde{\mu} (B(w, 3c\rho_{\ell_i})), \quad (6.25)$$

where $\delta(\gamma_2, \varepsilon) = (\varphi(\gamma_2) + C(\gamma_2)\varphi(4\varepsilon))$. Note that the above inequality has the discretized measure $\tilde{\mu}_{kr}$ on the left side and has the original measure $\tilde{\mu}$ on the right side. Applying this estimate with $i = m - 1$ and combining it with (6.22), we obtain

$$\tilde{\mu}_{kr}(F) \leq \delta(\gamma_2, \varepsilon) \sum_{w \in \tilde{\Lambda}_{r\ell_{m-1}}(F)} \tilde{\mu} (B(w, 3c\rho_{\ell_{m-1}})). \quad (6.26)$$

Our next lemma will allow us to apply induction on the above estimate.

Lemma 6.22. *There exists a uniform constant $C_N \geq 1$, depending only on the metric on the nilpotent group N , so that for all $2 \leq i \leq m$, we have*

$$\sum_{w \in \tilde{\Lambda}_{r\ell_i}(F)} \tilde{\mu}(B(w, 4c\rho_{\ell_i})) \leq C_N \delta(\gamma_2, \varepsilon) \sum_{v \in \tilde{\Lambda}_{r\ell_{i-1}}(F)} \tilde{\mu}(B(v, 4c\rho_{\ell_{i-1}})).$$

Proof. We begin by noting the following equality that relates the scale ρ_{ℓ_i} to the scale $\rho_{\ell_{i-1}}$:

$$\sum_{w \in \tilde{\Lambda}_{r\ell_i}(F)} \tilde{\mu}(B(w, 4c\rho_{\ell_i})) = \sum_{v \in \tilde{\Lambda}_{r\ell_{i-1}}(F)} \sum_{\substack{w \in \tilde{\Lambda}_{r\ell_i}(F) \\ \mathcal{D}_{r\ell_i}(w) \subset \mathcal{D}_{r\ell_{i-1}}(v)}} \tilde{\mu}(B(w, 4c\rho_{\ell_i})).$$

Let w be such that $w \in \tilde{\Lambda}_{r\ell_i}(F)$ and $\mathcal{D}_{r\ell_i}(w) \subset \mathcal{D}_{r\ell_{i-1}}(v)$. Since $\mathcal{D}_{r\ell_i}(w)$ meets F , we have that w is at distance at most $c\rho_{\ell_i} \leq \varepsilon\rho_{\ell_{i-1}}/10$ from a point in F . We also have that $F \cap \mathcal{D}_{r\ell_{i-1}}(v)$ is contained inside the subspace neighborhood $\tilde{V}_v^{(c\varepsilon\rho_{\ell_{i-1}})}$ by our choices above. Hence, by (6.23), we obtain that the ball $B(w, 4c\rho_{\ell_i})$ is contained inside $\tilde{V}_v^{(2c\varepsilon\rho_{\ell_{i-1}})}$. Finally, similar considerations imply that $B(w, 4c\rho_{\ell_i})$ is contained in $B(v, 2c\rho_{\ell_{i-1}})$. Put together, we arrive at the following inclusion:

$$\bigcup_{\substack{w \in \tilde{\Lambda}_{r\ell_i}(F) \\ \mathcal{D}_{r\ell_i}(w) \subset \mathcal{D}_{r\ell_{i-1}}(v)}} B(w, 4c\rho_{\ell_i}) \subseteq \tilde{V}_v^{(2c\varepsilon\rho_{\ell_{i-1}})} \cap B(v, 2c\rho_{\ell_{i-1}}).$$

On the other hand, by Lemma 6.9, the balls on the left side of the above equation have intersection multiplicity $O_{N,c}(1)$. Letting C_N denote this bound on multiplicity, we thus obtain

$$\sum_{\substack{w \in \tilde{\Lambda}_{r\ell_i}(F) \\ \mathcal{D}_{r\ell_i}(w) \subset \mathcal{D}_{r\ell_{i-1}}(v)}} \tilde{\mu}(B(w, 4c\rho_{\ell_i})) \leq C_N \tilde{\mu} \left(\tilde{V}_v^{(2c\varepsilon\rho_{\ell_{i-1}})} \cap B(v, 2c\rho_{\ell_{i-1}}) \right).$$

To apply our non-concentration hypothesis, we wish to find a suitable point x in the support of $\tilde{\mu}$ which is sufficiently close to v . To this end, recall by Lemma 6.8(2) that there is $x \in \mathcal{D}_{r\ell_{i-1}}(v)$ such that x is outside the exceptional set for $\tilde{\mu}$ and such that the set of good scales $\mathcal{N}(x)$ contains the set $\mathcal{N}(v)$. In particular, since $\ell_{i-1} \in \mathcal{N}(v)$ by construction, we also have that $\ell_{i-1} \in \mathcal{N}(x)$. Finally, since $\mathcal{D}_{r\ell_{i-1}}(v)$ intersects the unit ball around identity, it is contained in $B(\text{id}, 2)$ whenever r is larger than an absolute constant. In particular, $x \in B(\text{id}, 2)$, and hence, we obtain by our non-concentration hypothesis that

$$\tilde{\mu} \left(\tilde{V}_v^{(2c\varepsilon\rho_{\ell_{i-1}})} \cap B(v, 2c\rho_{\ell_{i-1}}) \right) \leq \tilde{\mu} \left(\tilde{V}_v^{(3c\varepsilon\rho_{\ell_{i-1}})} \cap B(x, 3c\rho_{\ell_{i-1}}) \right) \leq \delta(\gamma_2, \varepsilon) \tilde{\mu}(B(x, 3c\rho_{\ell_{i-1}})).$$

Using that $x \in \mathcal{D}_{r\ell}(v)$ once more, we get that the right side of the above inequality is at most $\delta(\gamma_2, \varepsilon) \tilde{\mu}(B(v, 4c\rho_{\ell_{i-1}}))$. This completes the proof of the lemma. \square

Applying the above lemma $(m-2)$ -times to the right side of (6.26), we obtain

$$\tilde{\mu}_{kr}(F) \leq \delta(\gamma_2, \varepsilon) \sum_{w \in \tilde{\Lambda}_{r\ell_{m-1}}(F)} \tilde{\mu}(B(w, 4c\rho_{\ell_i})) \leq (C_N \delta(\gamma_2, \varepsilon))^{m-1} \sum_{v \in \tilde{\Lambda}_{r\ell_1}(F)} \tilde{\mu}(B(v, 4c\rho_{\ell_1})).$$

Since $F \subset B(\text{id}, 1)$, each of the balls $B(v, 4c\rho_{\ell_1})$ is contained in $B(\text{id}, 2)$ for all $v \in \tilde{\Lambda}_{r\ell_1}(F)$ whenever r is large enough. Moreover, by Lemma 6.9, those balls have uniformly bounded multiplicity. We thus obtain the bound $\tilde{\mu}_{kr}(F) \ll (C_N \delta(\gamma_2, \varepsilon))^{m-1}$.

On the other hand, by Lemma 6.17(1), we have the lower bound $\tilde{\mu}_{kr}(F) \geq 2^{-\sqrt{\varepsilon'}k-k-1}$. Moreover, by taking γ_2 small enough, and taking ε sufficiently small depending on γ_2 , we can ensure that

$C_N \delta(\gamma_2, \varepsilon)$ is at most 1. Taking k large enough so that $1/k \leq \gamma_3/2$, we arrive at the inequality

$$2^{-\sqrt{\varepsilon'}-1-1/k} \leq C_0 (C_N \delta(\gamma_2, \varepsilon))^{\gamma_3-1/k} \leq C_0 (C_N \delta(\gamma_2, \varepsilon))^{\gamma_3/2},$$

where $C_0 \geq 1$ is the implicit constant in the previous inequality. Recall that $\gamma_1 = \gamma/10$, γ_2 is to be chosen smaller than $\gamma_1/4$, and $\gamma_3 = \gamma_1/4 - \gamma_2$. Hence, by choosing γ_2 first to be sufficiently small relative to γ_1 , then choosing ε very small, depending on γ_2 , we can make the right side of the above inequality at most $1/2$. This gives a contradiction since $\varepsilon' < 1$.

6.8. Proof of Theorem 6.3. Let $\eta > 0$ and $r \in \mathbb{N}$ be the parameters provided by Proposition 6.14 applied with $\gamma = \varepsilon/d$. Let $n \in \mathbb{N}$ be the smallest integer such that $(n-1)\eta \geq (d-\varepsilon)/2$. Note that by Young's inequality, for all $a, b, k \in \mathbb{N}$, we have that

$$\|\mu_k^{*a} * \mu_k^{*b}\|_2 \leq \|\mu_k^{*a}\|_1 \|\mu_k^{*b}\|_2 = \|\mu_k^{*b}\|_2.$$

We first observe that this inequality implies that it suffices to prove the first assertion of the theorem for multiples of r . Indeed, given any probability measure ν , $k \in \mathbb{N}$, and $0 \leq s < r$, we have

$$\sum_{P \in \mathcal{D}_{kr+s}} \nu(P)^2 = \sum_{Q \in \mathcal{D}_{kr}} \sum_{P \in \mathcal{D}_{kr+s}, P \subseteq Q} \nu(P)^2 \leq 2^{dr} \sum_{Q \in \mathcal{D}_{kr}} \nu(Q)^2.$$

Let k_1 be the parameter provided by Prop. 6.14 applied with $\lceil \log n \rceil$, where n is as above. Let $k \geq k_1$ be given and suppose that

$$\|\mu_{kr}^{*\ell}\|_2^2 \leq 2^{-(d-\varepsilon)kr}, \quad (6.27)$$

for some $\ell \in \mathbb{N}$ with $1 \leq \ell \leq n$. By Lemma 6.6, we also have that $\|\mu_{kr}^{*n}\|_2^2 \asymp \|\mu_{kr}^{*(n-\ell)} * \mu_{kr}^{*\ell}\|_2^2$. Hence, since convolution with $\mu_{kr}^{*(n-\ell)}$ does not increase the ℓ^2 -norm, we get that $\|\mu_{kr}^{*n}\|_2^2 \ll 2^{-(d-\varepsilon)kr}$ as desired. Now, suppose that (6.27) fails for all $1 \leq \ell \leq n$. Then, applying Prop. 6.14 $(n-1)$ -times by induction, and using Lemma 6.6, we obtain for some uniform constant $c \geq 1$

$$\|\mu_{kr}^{*n}\|_2 \leq c \|\mu_{kr} * \mu_{kr}^{*(n-1)}\|_2 \leq c 2^{-\eta kr} \|\mu_{kr}^{*(n-1)}\|_2 \leq \dots \leq c^{n-1} 2^{-\eta(n-1)kr} \|\mu_{kr}\|_2 \leq c^{n-1} 2^{-\eta(n-1)kr},$$

where the second inequality follows since $\|\mu_{kr}\|_2 \leq 1$. This proves the first assertion by our choice of n . The (short) deduction of the second assertion can be found in [MS18, Lemma 5.2].

6.9. Proof of Theorem 1.5, Corollary 1.7, and Corollary 6.4 from Theorem 6.3. Note that being uniformly affinely non-concentration immediately implies that μ is affinely non-concentrated at almost every (in fact at every) scale with an empty exceptional set. Hence, the second assertion of Theorem 6.3 immediately implies that $\dim_\infty \mu^{*n}$ tends to d as $n \rightarrow \infty$. The same holds for $\dim_q \mu^{*n}$ due to the inequality $\dim_q \mu \geq \dim_\infty \mu$ for all $q > 1$. Finally, the first assertion of Theorem 1.5 follows readily from Proposition 6.14; cf. [RS20, Proof of Theorem 1.1] for details of this deduction.

Similarly, Corollary 1.7 is a special case of Corollary 6.4. Hence, it remains to deduce Corollary 6.4 from Theorem 6.3 via the well-known relationship between L^2 -dimension and Fourier transform. Namely, by [FNW02, Proof of Claim 2.8], we have⁵

$$\int_{\|\xi\| \leq 1/r} |\hat{\nu}(\xi)|^2 d\xi \ll_d r^{-2d} \int \nu(B(x, r))^2 dx. \quad (6.28)$$

for every $r > 0$ and any Borel probability measure ν on \mathbb{R}^d . Moreover, if $k \in \mathbb{N}$ is such that $2^{-(k+1)} < r \leq 2^{-k}$, then $B(x, r)$ can be covered by $O_d(1)$ elements of the partition \mathcal{D}_k . Hence,

$$\int \nu(B(x, 2^{-k}))^2 dx \ll_d 2^{-dk} \sum_{P \in \mathcal{D}_k} \nu(P)^2 = 2^{-dk} \|\nu_k\|_2^2. \quad (6.29)$$

⁵The reference [FNW02] proves this fact in the case $d = 1$, however the proof works equally well for \mathbb{R}^d for any d .

Now, let μ be a measure satisfying the hypotheses of Corollary 6.4 and let $\varepsilon > 0$ be arbitrary. By Theorem 6.3, there are natural numbers n, k_1 , depending only on ε and the non-concentration parameters of μ , such that for all $k \geq k_1$,

$$\|\mu_k^{*n}\|_2^2 \ll_{\varepsilon, d, n} 2^{-(d-\varepsilon/2)k}.$$

Given $T > 2^{k_1}$, let $r = 1/T$ and $k \in \mathbb{N}$ be such that $2^{-(k+1)} < r \leq 2^{-k}$. We can apply (6.28) and (6.29) with $\nu = \mu_k^{*n}$ to get that

$$\int_{\|\xi\| \leq T} |\hat{\mu}(\xi)|^{2n} d\xi \ll_{\mu, \varepsilon} T^{2d} 2^{-dk - (d-\varepsilon/2)k} \ll T^{\varepsilon/2}. \quad (6.30)$$

The conclusion of the corollary will now follow by Chebyshev's inequality and the fact that the Fourier transform is Lipschitz. Indeed, note that

$$|\hat{\mu}(\xi_1) - \hat{\mu}(\xi_2)| \ll_{\mu} \|\xi_1 - \xi_2\|,$$

where the implicit constant depends only on the radius of the smallest ball around the origin containing the support of μ .

In particular, given $\delta > 0$, if $|\hat{\mu}(\xi)| > T^{-\delta}$ for some ξ , then $|\hat{\mu}|$ is at least $T^{-\delta}/2$ on a ball of radius $T^{-2\delta}$, when T is large enough depending on δ and μ . It follows that

$$\begin{aligned} & \# \left\{ v \in \mathbb{Z}^d : \|v\| \leq T, \text{ and there exists } \xi \in v + [0, 1)^d \text{ such that } |\hat{\mu}(\xi)| > T^{-\delta} \right\} \times T^{-2\delta d} \\ & \ll \left| \left\{ \xi \in \mathbb{R}^d : \|\xi\| \leq 2T \text{ and } |\hat{\mu}(\xi)| > T^{-\delta}/2 \right\} \right|. \end{aligned}$$

Chebyshev's inequality applied to (6.30) implies that the right side of the above inequality is $O(T^{\varepsilon/2 + 2\delta n})$. Thus, the number of radius one balls needed to cover the set of frequencies ξ of norm at most T such that $|\hat{\mu}(\xi)| > T^{-\delta}$ is $O(T^{\varepsilon/2 + \delta(2n+2d)})$. Thus, taking $\delta = \varepsilon/2(2n+2d)$, we obtain the assertion of Corollary 6.4 as desired.

6.10. Polynomial affine non-concentration. In this section, we show that Theorem 6.3 implies quantitative non-concentration estimates near proper subspaces.

Theorem 6.23. *Let $\tilde{\mu}$ and μ be as in Theorem 6.3. Then, there exist $\kappa > 0$ and C , depending on the non-concentration parameters of $\tilde{\mu}$, such that for all $\varepsilon > 0$ and all proper affine subspaces $W < \mathbb{R}^d$, we have that $\mu(W^{(\varepsilon)}) \leq C\varepsilon^{\kappa}$.*

We first need the following useful observation which translates polynomial non-concentration for self-convolution into a similar estimate for the original measure.

Lemma 6.24. *Let ν be a Borel probability measure, $\varepsilon, \alpha, C > 0$ be arbitrary constants, and $W < \mathbb{R}^d$ be an affine hyperplane. Let $V = W + W$. Assume that $\nu^{*2}(V^{(\varepsilon)}) \leq C\varepsilon^{\alpha}$. Then, $\nu(W^{(\varepsilon/2)}) \leq C\varepsilon^{\alpha/2}$.*

Proof. Note that the definition of convolution implies

$$\nu^{*2}(V^{(\varepsilon)}) = \int \int \mathbb{1}_{V^{(\varepsilon)}}(x+y) d\nu(x) d\nu(y) = \int \nu(V^{(\varepsilon)} - x) d\nu(x).$$

Hence, by Chebyshev's inequality and our hypothesis on ν , the set

$$B = \left\{ x \in \mathbb{R}^d : \nu(V^{(\varepsilon)} - x) > \varepsilon^{\alpha/2} \right\}$$

has ν measure at most $C\varepsilon^{\alpha/2}$. Hence, the lemma follows if $W^{(\varepsilon/2)}$ is contained inside B . Otherwise, let $x \in W^{(\varepsilon/2)} \setminus B$ and observe that $W^{(\varepsilon/2)}$ is contained inside $V^{(\varepsilon)} - x$. However, the latter set has ν measure at most $\varepsilon^{\alpha/2}$ since $x \notin B$. Hence, the lemma follows in this case as well. \square

We are now ready for the proof of Theorem 6.23.

Proof of Theorem 6.23. By Theorem 6.3, we can find $n \in \mathbb{N}$ and $r_0 > 0$, depending only on the non-concentration parameters of $\tilde{\mu}$, such that

$$\mu^{*2^n}(B(x, r)) \ll_d 2^{2^n dm} r^{d-1/2}, \quad (6.31)$$

for all $0 < r \leq r_0$ and all $x \in \mathbb{R}^d$. Fix one such value of n once and for all. Let $\nu = \mu^{*2^n}$ and let $B \subset \mathbb{R}^d$ be a large ball containing the supports of μ^{*k} for all $0 \leq k \leq 2^n$. In light of Lemma 6.24, it will suffice to find $C \geq 1$ and $\alpha > 0$ so that $\nu(V^{(\varepsilon)}) \leq C\varepsilon^\alpha$ for all proper affine hyperplanes V .

Let $0 < \varepsilon \leq 1$ and a proper affine hyperplane $V \subset \mathbb{R}^d$ be arbitrary. Then, note that $V^{(\varepsilon)} \cap B$ can be covered by $O_{B,d}(\varepsilon^{-(d-1)})$ balls of radius ε with multiplicity depending only on d . Then, (6.31) implies that $\nu(V^{(\varepsilon)}) \leq C'\varepsilon^{1/2}$ for a suitable constant $C' = C'(m, n, d) \geq 1$. Since V was arbitrary, Lemma 6.24 and induction on n show that $\mu(W^{(\varepsilon/2^n)}) \leq C'\varepsilon^\kappa$ for $\kappa = 2^{-n-1}$ for all proper hyperplanes W . Since $\varepsilon > 0$ was arbitrary, this completes the proof by taking $C = C'2^{\kappa n}$. \square

7. NON-CONCENTRATION OF PATTERSON-SULLIVAN MEASURES

In this section, we verify the non-concentration hypothesis in Corollary 6.4 for the measures μ_x^u . This enables us to apply these results to prove Proposition 4.14 and Theorem 4.17 which are the remaining pieces in the proof of Theorem 4.2.

Definition 7.1. Let \mathcal{L}' denote the collection of all proper linear subspaces of the Lie algebra \mathfrak{n}^+ and denote by \mathcal{L} the set of all N^+ -translates of images of elements of \mathcal{L}' under the exponential map. Elements of \mathcal{L} will be referred to as **affine subspaces** of N^+ . For $\varepsilon > 0$ and $L \in \mathcal{L}$, let $L^{(\varepsilon)}$ be the ε -neighborhood of L .

Recall that we fixed a choice of a Margulis function V in Remark 4.1 and define

$$t(\varepsilon) := \sup_{x \in N_1^- \Omega} \sup_{L \in \mathcal{L}} \frac{\mu_x^u(N_1^+ \cap L^{(\varepsilon)})}{V(x)\mu_x^u(N_1^+)}. \quad (7.1)$$

We also recall that Γ is a geometrically finite subgroup of $G = \text{Isom}^+(\mathbb{H}_{\mathbb{K}}^d)$.

Theorem 7.2. *Assume that Γ is Zariski-dense inside G . We have that $t(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.*

As a consequence, we verify the hypotheses of Corollary 6.4.

Corollary 7.3. *For every $x \in N_1^- \Omega$, the measure μ_x^u is affinely non-concentrated at almost every scale in the sense of Definition 6.1 with parameters depending only on $V(x)$; cf. (7.4).*

7.1. Proof of Theorem 7.2. Our key tool is the following result which is a consequence of the ergodicity of the geodesic flow. The case of real hyperbolic spaces of this result was known earlier in [FS90] by different methods.

Proposition 7.4 ([ELO22, Corollary 9.4]). *For all $x \in X$ and $L \in \mathcal{L}$, $\mu_x^u(L) = 0$.*

Theorem 7.2 follows from the above result and a compactness argument. Indeed, fix an arbitrary $\eta > 0$ and note that for all x with $V(x) > 1/\eta$, the inner supremum in the definition of $t(\varepsilon)$ is bounded above by η , for any choice of $\varepsilon > 0$. We now show that $t(\varepsilon) < \eta$ for all sufficiently small ε by restricting our attention to the bounded set of $x \in N_1^- \Omega$ where $V(x) \leq 1/\eta$. Suppose not and let $x_n \in N_1^- \Omega$, $L_n \in \mathcal{L}$, $\varepsilon_n > 0$ be sequences such that $V(x_n) \leq 1/\eta$, $\varepsilon_n \rightarrow 0$, and

$$\liminf_{n \rightarrow \infty} \frac{\mu_{x_n}^u(N_1^+ \cap L_n^{(\varepsilon_n)})}{\mu_{x_n}^u(N_1^+)} > 0. \quad (7.2)$$

Passing to a subsequence if necessary, we may assume $x_n \rightarrow y \in N_1^- \Omega$ and L_n converges to some $P \in \mathcal{L}$ (in the Hausdorff topology on compact sets). On the other hand, when x_n is sufficiently close to y , we can change variables using (2.3), and (2.7) to get $\mu_{x_n}^u(N_1^+ \cap L_n^{(\varepsilon_n)}) = \int f_n J_n d\mu_y^u$,

where J_n is the Jacobian of the change of variables and f_n is the indicator function of the image of $N_1^+ \cap L_n^{(\varepsilon_n)}$ under this change of variables. By Proposition 7.4, since L_n converges to L , f_n converges to 0 pointwise μ_y^u -almost everywhere. Additionally, J_n converges to 1 everywhere since x_n converges to y . Finally, $\mu_{x_n}^u(N_1^+)$ remains bounded away from 0 since x_n remain within a bounded set for all n . This gives a contradiction to (7.2) and concludes the proof.

7.2. Non-concentration and proof of Corollary 7.3. In this section, we show that the conditional measures μ_\bullet^u are affinely non-concentrated in the sense of Def. 6.1. Our key tools are Theorem 2.5 and Theorem 7.2.

Let $0 < \theta, \varepsilon < 1$ be arbitrary. Let $H, r_0 = O_{\beta, \theta}(1)$ be the constants provided by Theorem 2.5 when applied with $\varepsilon = \beta\theta/2$ and let $r \geq r_0$. For $\ell \in \mathbb{N}$, let $t_\ell = r\ell \log 2$ and define

$$\mathcal{E} = \left\{ n \in N_1^+ : \sum_{1 \leq \ell \leq k} \chi_H(g_{t_\ell} n x) \geq \theta k \right\}.$$

Then, by Theorem 2.5, we have that $\mu_x^u(\mathcal{E} \cap N_1^+) \ll e^{-\beta\theta k/2} V(x) \mu_x^u(N_1^+)$.

It remains to show that our desired non-concentration holds outside of \mathcal{E} . For $n \in N_1^+$, define the set of scales $\mathcal{N}(n)$ as follows:

$$\mathcal{N}(n) = \{1 \leq \ell \leq k : V(g_{t_\ell} n x) \leq H\}.$$

Let $n \in N_1^+ \cap \text{supp}(\mu_x^u) \setminus \mathcal{E}$. By definition, we have $\#\mathcal{N}(n) \geq (1 - \theta)k$.

Let $\ell \in \mathcal{N}(n)$ and let $W < N^+$ be a proper affine subspace. Recall the function $t(\varepsilon)$ defined in (7.1). Let $\rho \asymp 2^{-r\ell}$. Let $z = g_{-\log \rho} n x$ and $W_n = \text{Ad}(g_{-\log \rho})(W n^{-1})$. Then, changing variables and using the definition of $t(\varepsilon)$ along with the fact that $V(z) \ll H$, we obtain

$$\mu_x^u(W^{(\varepsilon\rho)} \cap N_\rho^+ \cdot n) = \rho^\delta \mu_z^u(W_n^{(\varepsilon)} \cap N_1^+) \ll H t(\varepsilon) \times \rho^\delta \mu_z^u(N_1^+) = H t(\varepsilon) \times \mu_x^u(N_\rho^+ \cdot n), \quad (7.3)$$

where the last equality follows by reversing the change of variables since $\rho^\delta \mu_z^u(N_1^+) = \mu_x^u(N_\rho^+ \cdot n)$.

Let $C_1 \geq 1$ be the larger of the implicit constants in the bound on the measure of \mathcal{E} and in (7.3). These two estimates imply that μ_x^u satisfies Definition 6.1 by taking

$$C(\theta) := C_1 V(x) H, \quad \varphi(\varepsilon) := C_1 t(\varepsilon), \quad \lambda(\theta) := (\beta\theta \log 2)/2. \quad (7.4)$$

That $\varphi(\varepsilon)$ tends to 0 as $\varepsilon \rightarrow 0$ follows by Theorem 7.2.

7.3. Counting close frequencies and proof of Proposition 4.14. The idea of the proof is similar to that of [Liv04, Lemma 6.2], with the significant added difficulty being the non-concentration result for PS measures established in Theorem 6.23. We note however that the case of real hyperbolic manifolds is much simpler in that it does not require Theorem 6.23 and instead uses only the doubling result in Proposition 2.2.

Recall our definition of the transverse intersection points $x_{\rho, \ell}$ in (4.45) and of $N_1^+(j)$ in the paragraph above (4.35). For each $\ell \in I_{\rho, j}$, fix some $u_\ell \in N_1^+(j) \subseteq N_3^+$ such that

$$x_{\rho, \ell} = g^\gamma p_\ell^+ \cdot x = n_{\rho, \ell}^- \cdot y_\rho, \quad p_\ell^+ := m_{\rho, \ell} g_{t_{\rho, \ell}} u_\ell. \quad (7.5)$$

Here, we are using that the groups $A = \{g_t : t \in \mathbb{R}\}$ and M commute. Denote by P^+ the parabolic subgroup $N^+ A M$ of G . Since M is compact, $|t_{\rho, \ell}| < 1$, and $N_1^+(j)$ is contained in N_3^+ , there is a uniform constant $C > 0$ such that

$$\{p_\ell^+ : \ell \in I_{\rho, j}\} \subset P_C^+, \quad (7.6)$$

where P_C^+ denotes the ball of radius C around identity in P^+ .

Fix some $\ell_0 \in I_{\rho,j}$ and denote by $C_{\rho,j}(\ell_0)$ the set of indices $\ell \in I_{\rho,j}$ such that $(\ell_0, \ell) \in C_{\rho,j}$. To simplify notation, we set

$$\epsilon := b^{-1/10}, \quad t_\star := \gamma(w + jT_0).$$

Let $Z = \exp(\mathfrak{n}_{2\alpha}^-) \subset N^-$. In particular⁶, $Z = \{\text{id}\}$ is the trivial group in the real hyperbolic case. Recalling the definition of the Cygan metric in (2.5), the definition of $C_{\rho,j}$ implies that

$$d_{N^-}(n_{\rho,\ell}^-(n_{\rho,\ell_0}^-)^{-1}, Z) \leq b^{-1/10}.$$

Denote by $Z^{(\epsilon)}$ the ϵ -neighborhood of Z inside N^- . Let

$$\tilde{u}_\ell^- = n_{\rho,\ell}^-(n_{\rho,\ell_0}^-)^{-1} \in Z^{(\epsilon)} \cap N_{\ell_j}^-,$$

where we recall that the points $n_{\rho,\ell}^-$ belong to $N_{\ell_j/10}^-$ by definition of our flow boxes B_ρ ; cf. paragraph preceding (4.27). Note that

$$g^\gamma p_\ell^+ \cdot x = \tilde{u}_\ell^- \cdot g^\gamma p_{\ell_0}^+ \cdot x, \quad \forall \ell \in C_{\rho,j}(\ell_0).$$

In particular, for $u_\ell^- = \text{Ad}(g^\gamma)^{-1}(\tilde{u}_\ell^-)$, since $g^\gamma = g_{t_\star}$ (cf. (4.30)), we have that

$$p_\ell^+ x = u_\ell^- \cdot p_{\ell_0}^+ x \in (Z^{(\epsilon^{t_\star})} \cap N_{\epsilon^{t_\star} \ell_j}^-) \cdot p_{\ell_0}^+ x, \quad \forall \ell \in C_{\rho,j}(\ell_0). \quad (7.7)$$

Our counting estimate will follow by estimating from below the separation between the points $p_\ell^+ x$, combined with a measure estimate on the sets $(Z^{(\epsilon^{t_\star})} \cap N_{\epsilon^{t_\star} \ell_j}^-) \cdot p_{\ell_0}^+ x$.

To this end, recall the sublevel set K_j and the injectivity radius ι_j in (4.25). Recall also by (4.18) that x belongs to K_j . It follows that the injectivity radius at every point of the weak unstable ball $P_C^+ \cdot x$ is $\gg \iota_j$. This implies that there is a radius r_j with $\iota_j \ll r_j \leq \iota_j$ such that for every $\ell \in C_{\rho,j}(\ell_0)$, the map $n^- \mapsto n^- \cdot p_\ell^+ x$ is an embedding of $N_{r_j}^-$ into X .

Let $\{B_m\}$ be a cover of P_C^+ by $O(\iota_j^{-\dim P^+})$ balls of radius r_j . Then, similar injectivity radius considerations imply that, for every m , the disks

$$\left\{ N_{r_j}^- \cdot p_\ell^+ x : \ell \in C_{\rho,j}(\ell_0), p_\ell^+ \in B_m \right\}$$

are disjoint. Indeed, otherwise, we can find $n \in N_{2r_j}^-$ and $\ell_1, \ell_2 \in C_{\rho,j}$ with $p_{\ell_1}^+, p_{\ell_2}^+ \in B_m$ such that $np_{\ell_1}^+ x = p_{\ell_2}^+ x$. By choosing r_j sufficiently smaller than ι_j , this gives a contradiction to the fact that the injectivity radius at x is $\gg \iota_j$ since $(p_{\ell_2}^+)^{-1} np_{\ell_1}^+ = (p_{\ell_2}^+)^{-1} np_{\ell_2}^+ \cdot (p_{\ell_2}^+)^{-1} p_{\ell_1}^+$ is at distance $O(r_j)$ from identity. The above disjointness, together with (7.7), imply that the disks $\left\{ N_{r_j}^- \cdot u_\ell^- : \ell \in C_{\rho,j}(\ell_0), p_\ell^+ \in B_m \right\}$ form a disjoint collection of disks inside $Z^{(\epsilon^{t_\star} \epsilon + \iota_j)} \cap N_{(\epsilon^{t_\star} + 1)\iota_j}^-$. In particular, we get that

$$\# \left\{ \ell \in C_{\rho,j}(\ell_0) : p_\ell^+ \in B_m \right\} \leq \frac{\mu_{p_{\ell_0}^+ x}^s \left(Z^{(\epsilon^{t_\star} \epsilon + \iota_j)} \cap N_{(\epsilon^{t_\star} + 1)\iota_j}^- \right)}{\min_{\ell \in C_{\rho,j}(\ell_0)} \mu_{p_{\ell_0}^+ x}^s (N_{r_j}^- \cdot u_\ell^-)}, \quad (7.8)$$

where μ_\bullet^s are the conditional measures along N^- -orbits defined similarly to (2.2).

To obtain good bounds on the ratio in (7.8) for a given ℓ , it will be important to change the basepoint $p_{\ell_0}^+ x$ to another point of the form $g_s p_\ell^+ x$ with uniformly bounded height. Fix some arbitrary $\ell \in C_{\rho,j}(\ell_0)$ and recall (7.5) and (7.7). Let $s_{\rho,\ell} \in [t_\star, (1 + 2\alpha)t_\star]$ be the return time defined in (4.47) and set

$$y_\ell = g_{s_{\rho,\ell}} p_\ell^+ x.$$

⁶This is the reason Theorem 6.23 is not needed in this case.

Note that our choice of u_ℓ^- implies that

$$Z^{(e^{t_\star}\epsilon + \iota_j)} \cap N_{(e^{t_\star}+1)\iota_j}^- \subseteq \left(Z^{(2e^{t_\star}\epsilon + \iota_j)} \cap N_{2(e^{t_\star}+1)\iota_j}^- \right) \cdot u_\ell^-.$$

In particular, we can use the set on the right side to estimate the numerator of (7.8). Let

$$Q := Z^{(2e^{t_\star}\epsilon + \iota_j)} \cap N_{2(e^{t_\star}+1)\iota_j}^-, \quad Q' := \text{Ad}(g_{s_{\rho,\ell}})(Q).$$

Then, changing variables using (2.3), we have

$$\frac{\mu_{p_{\ell_0}^+ x}^s(Q \cdot u_\ell^-)}{\mu_{p_{\ell_0}^+ x}^s(N_{r_j}^- \cdot u_\ell^-)} = \frac{\mu_{p_{\ell}^+ x}^s(Q)}{\mu_{p_{\ell}^+ x}^s(N_{r_j}^-)} = \frac{\mu_{y_\ell}^s(Q')}{\mu_{y_\ell}^s(N_{e^{-s_{\rho,\ell}} r_j}^-)}.$$

Moreover, by Sullivan's shadow lemma [Sch04, Theorem 3.2], since $V(y_\ell) \ll_{T_0} 1$, we have that

$$\mu_{y_\ell}^s(N_{e^{-s_{\rho,\ell}} r_j}^-) \gg_{T_0} e^{-\delta s_{\rho,\ell} r_j^\delta} \gg e^{-\delta(1+2\alpha)t_\star} \iota_j^\delta. \quad (7.9)$$

Here, we used [Cor90, Theorem 2.2] to relate the disks $N_r^- \cdot y_\ell$ to their shadows on the boundary.

Lemma 7.5. *We have the bound $\mu_{y_\ell}^s(Q') \ll b^{-\kappa/10} + e^{-\kappa t_\star}$, where $\kappa > 0$ is a uniform constant provided by Theorem 6.23.*

Proof. We wish to apply Theorem 6.23 to the measure $\mu_{y_\ell}^s$. This result concerns decay of measures of the intersection of subspace neighborhoods with N_1^- . First, we show that $Q' \subseteq Z^{(\varrho)} \cap N_1^-$ for $\varrho = 2\epsilon + \iota_j e^{-t_\star}$. Indeed, let $r_1 = (2e^{t_\star}\epsilon + \iota_j) e^{-s_{\rho,\ell}}$ and $r_2 = 2(e^{t_\star} + 1)\iota_j e^{-s_{\rho,\ell}}$. Then, $Q' = Z^{(r_1)} \cap N_{r_2}^-$. Since $s_{\rho,\ell} \geq t_\star$ and $\iota_j \leq 1/10$ (cf. (4.25)), we have $r_2 \leq 1$. Similarly, $r_1 \leq \varrho$.

Next, we note that the non-concentration hypothesis of Theorem 6.23 is verified in Corollary 7.3. Moreover, the corollary also shows that the non-concentration parameters can be chosen uniform over all y_ℓ in view of the fact that $V(y_\ell) \ll 1$. Let μ be the projection of $\mu_{y_\ell}^s|_{N_1^-}$ to the abelianization $N^-/[N^-, N^-] \cong \mathfrak{n}_\alpha^-$, normalized to be a probability measure. Then, Theorem 6.23 provides constants $C_1, \kappa > 0$, independent of ℓ , so that $\mu(W^{(\varepsilon)}) \leq C_1 \varepsilon^\kappa$ for all $\varepsilon > 0$ and all proper affine subspaces W , where such subspaces are defined in Def. 7.1. The lemma now follows since Q' is contained in the ϱ -neighborhood of a translate of the preimage of 0 under this projection. \square

Recall that the cover $\{B_m\}$ has cardinality at most $O(\iota_j^{-\dim P^+})$. Also, recall by (4.26) that $\iota_j^{-1} \ll_{T_0} e^{4\alpha t_\star}$. Hence, (7.8), (7.9), and Lemma 7.5 imply that

$$\#C_{\rho,j}(\ell_0) \ll_{T_0} e^{O(\alpha t_\star)} (\epsilon^\kappa + e^{-\kappa t_\star}) e^{\delta t_\star}.$$

Finally, note that (4.14) provides the bound $e^{\alpha t_\star} \ll_{T_0} b^{2\alpha/a}$. Hence, if α is small enough, the above bounds imply that $\#C_{\rho,j}(\ell_0) \ll_{T_0} (\epsilon^{\kappa_0} + e^{-\kappa_0 t_\star}) e^{\delta t_\star}$, for $\kappa_0 = \kappa/2$. This concludes the proof.

7.4. Flattening and proof of Theorem 4.17. We wish to apply Corollary 6.4. Recall that ν_i has total mass $\mu_{y_p^i}^u(N_1^+)$ and let $\mu = \nu_i / \mu_{y_p^i}^u(N_1^+)$. In particular, μ is a probability measure supported on the unit ball in \mathfrak{n}_α^+ . We fix identifications $\mathfrak{n}_\alpha^+ \cong \mathbb{K}^p \cong \mathfrak{n}_\alpha^-$ for some $p \in \mathbb{N}$; cf. Section 2.4. Note further that the restriction of the metric in (2.5) to \mathfrak{n}_α^+ is Euclidean. In particular, we will fix a linear isomorphism of \mathfrak{n}_α^+ and \mathfrak{n}_α^- with \mathbb{R}^d , where $d = p \dim \mathbb{K}$.

By Corollary 7.3, the measure $\mu_{y_p^i}^u$ is affinely non-concentrated at almost all scales in the sense of Def. 6.1. Hence, Corollary 6.4 provides $\lambda > 0$ such that, for $T = b^{4/10}$, the set

$$\mathfrak{B}(\lambda) := \left\{ w \in \mathbb{R}^d : \|w\| \leq T \text{ and } |\hat{\mu}(w)| \geq T^{-\lambda} \right\}$$

can be covered by $O_\varepsilon(T^\varepsilon)$ balls of radius 1. The result will follow once we estimate the spacing between the functionals $\langle w_{k,\ell}^i, \cdot \rangle$.

To simplify notation, let $w_\ell := \langle w_{k,\ell}^i, \cdot \rangle$. By (4.75), when b is large enough, we have that $b\|w_\ell\| \leq T$. In particular, we can view the set $B(i, k, \lambda)$ as a subset of $\mathfrak{B}(\lambda)$ above using the map $\ell \mapsto bw_\ell$. By Lemma 4.13, the definition of $w_{k,\ell}^i$ in (4.70), and (4.68), we have that

$$\|w_{\ell_1} - w_{\ell_2}\| \gg b^{2/10} \|u_{\ell_2} - u_{\ell_1}\|.$$

In particular, by Proposition 4.14, any ball of radius 1 in \mathbb{R}^d contains at most

$$O_{T_0} \left((b^{-\kappa/10} + e^{-\kappa\gamma(w+jT_0)}) e^{\delta\gamma(w+jT_0)} \right)$$

of the vectors w_ℓ . This completes the proof of Theorem 4.17.

8. PROOF OF THEOREMS 1.1 AND 1.2

The goal of this section is to complete the proofs of Theorems 1.1 and 1.2 establishing exponential mixing of the geodesic flow. The key ingredients are Theorems 3.3 and 4.2. The deduction is through a form of the Paley-Wiener theorems adapted for this purpose obtained in [But16a, But16b].

8.1. Paley-Wiener theorems. Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space equipped with a weaker norm $\|\cdot\|_{\mathcal{A}}$. Let \mathcal{L}_t be a bounded one-parameter semigroup of operators on \mathcal{B} in the norm $\|\cdot\|_{\mathcal{B}}$. Denote by \mathfrak{X} the infinitesimal generator of \mathcal{L}_t and let $R(z)$, $\operatorname{Re}(z) > 0$, be its resolvent.

Theorem 8.1 ([But16a, But16b, Theorem 1]). *Assume that \mathcal{L}_t is strongly continuous⁷, and*

- (1) *\mathcal{L}_t is weakly-Lipschitz, i.e., for all $t \geq 0$ and $f \in \mathcal{B}$, $\|\mathcal{L}_t f - f\|_{\mathcal{A}} \ll t \|f\|_{\mathcal{B}}$.*
- (2) *there exists $\lambda > 0$ such that $\rho_{\text{ess}}(R(z)) \leq 1/(\operatorname{Re}(z) + \lambda)$ for all $z \in \mathbb{C}$ with $\operatorname{Re}(z) > 0$, where ρ_{ess} denotes the essential spectral radius.*
- (3) *there exist positive constants C, α, β and $0 < \gamma < \log(1 + \lambda/\alpha)$ such that, for all $z \in \mathbb{C}$ with $\operatorname{Re}(z) = \alpha$ and $|\operatorname{Im}(z)| \geq \beta$, we have $\|R(z)^q\|_{\mathcal{B}} \leq C/(\alpha + \lambda)^q$, where $q = \lceil \gamma \log |\operatorname{Im}(z)| \rceil$. Here, λ is the constant in Assumption (2).*

Then, there exists an operator valued function $t \mapsto \mathcal{P}_t$ taking values in the space of bounded operators on \mathcal{B} , and for $1 \leq j \leq N \in \mathbb{N}$, there exist $z_j \in \mathbb{C}$ with $-\lambda < \operatorname{Re}(z_j) \leq \beta$, a finite rank projector Π_j , and a nilpotent operator \mathcal{N}_j , so that the following hold:

- (1) *For all $1 \leq j, k \leq N$ and $t \geq 0$, we have*

$$\Pi_j \Pi_k = \delta_{jk} \Pi_j, \quad \Pi_j \mathcal{P}_t = \mathcal{P}_t \Pi_j = 0, \quad \Pi_j \mathcal{N}_j = \mathcal{N}_j \Pi_j = \mathcal{N}_j.$$

- (2) *For all $t \geq 0$, $\mathcal{L}_t = \mathcal{P}_t + \sum_{j=1}^N e^{tz_j} \exp(t\mathcal{N}_j) \Pi_j$.*

- (3) *For all f in the domain of \mathfrak{X} , $t \geq 0$ and $0 < \ell < \lambda$, $\|\mathcal{P}_t f\|_{\mathcal{A}} \ll_\ell e^{-\ell t} \|\mathfrak{X} f\|_{\mathcal{B}}$.*

8.2. Verification of the hypotheses of Theorem 8.1. Recall the Banach space \mathcal{B}_\star defined below (3.6) and the weak norm $\|\cdot\|'_1$ defined in (3.5). The link between the norms we introduced and decay of correlations is furnished in the following lemma.

Lemma 8.2 ([Kha23, Lemma 7.11]). *For all $f, \varphi \in C_c^2(X)^M$, we have that $\int f \cdot \varphi \, d\mu^{\text{BMS}} \ll_\varphi \|f\|'_1$, where the implied constant depends on $\|\varphi\|_{C^2}$ and the injectivity radius of its support.*

In particular, this lemma implies that decay of correlations (for mean 0 functions) would follow at once if we verify that $\|\mathcal{L}_t f\|'_1$ decays in t with a suitable rate. Theorem 8.1 shows that such decay follows from suitable spectral bounds on the resolvent. Hence, it remains to verify the hypotheses of Theorem 8.1. Recalling the definitions of the norms in Section 3.1, we take

$$\|\cdot\|_{\mathcal{A}} = \|\cdot\|'_1, \quad \|\cdot\|_{\mathcal{B}} = \|\cdot\|_1^\star$$

in the notation of Theorem 8.1. Strong continuity of \mathcal{L}_t is a consequence of the mean value theorem, cf. [Kha23, Cor. 7.2], while Theorem 3.3 verifies Assumption (2) of Theorem 8.1. The following lemma verifies Assumption (1).

⁷Cf. [But16a] for a result that does not require strong continuity.

Lemma 8.3. *For all $t \geq 0$, $\|\mathcal{L}_t f - f\|_1' \ll t \|f\|_1^*$.*

Proof. Recall that the norm $\|\cdot\|_1'$ only involves the coefficient $e'_{1,0}$; cf. (3.5). Let $x \in N_1^- \Omega$ and $t \geq 0$. Then, given any test function ϕ for $e'_{1,0}$, we have that

$$\int_{N_1^+} \phi(n)(f(g_t n x) - f(n x)) d\mu_x^u = \int_0^t \int_{N_1^+} \phi(n) L_\omega f(g_r n x) d\mu_x^u dr,$$

where L_ω is the derivative with respect to the vector field generating g_t . Lemma 3.5 thus implies

$$\left| \int_{N_1^+} \phi(n)(f(g_t n x) - f(n x)) d\mu_x^u \right| \leq V(x) \mu_x^u(N_1^+) \int_0^t e_{1,1}^*(\mathcal{L}_r f) dr \ll t V(x) \mu_x^u(N_1^+) e_{1,1}^*(f),$$

where $e_{1,1}^*$ is the coefficient defined above (3.6). This completes the proof. \square

Finally, the following corollary verifies Assumption 3 of Theorem 8.1.

Corollary 8.4. *In the notation of Theorem 4.2, there exist constants $c_\star, \lambda_\star > 0$, such that the following holds. For all $z = a_\star + ib \in \mathbb{C}$ and for $q = \lceil c_\star \log |b| \rceil$, we have the operator norm bound*

$$\|R(z)^q\|_1^* \leq (a_\star + \lambda_\star)^{-q},$$

whenever $|b| \geq b_\Gamma$, where $b_\Gamma \geq 1$ is a constant depending on Γ .

Proof. First, we verify the corollary for the norm $\|\cdot\|_{1,B}^*$. Let $e_{1,1,b}^*$ be the scaled seminorm $e_{1,1}^*/|b|^{1+\varkappa}$. The proof of [Kha23, Lemmas 7.5 and 7.6] shows that, for $z = a_\star + ib$ with $|b| \geq a_\star$,

$$e_{1,1,b}^*(R(z)^m f) \leq C_\Gamma \frac{\|f\|_{1,B}^* (a_\star + |z|)}{a_\star^m b^{1+\varkappa}} \leq \frac{3C_\Gamma \|f\|_{1,B}^*}{a_\star^m |b|^\varkappa},$$

for some constant $C_\Gamma \geq 1$ depending only on Γ , where we used the fact that $a_\star + |z| \leq 3|b|$.

Moreover, if $m = \lceil \log |b| \rceil \geq 3/2$, we have that $|b|^\varkappa \geq e^{\varkappa m/2} \geq (1 + \varkappa/2)^m$ and hence $a_\star^m |b|^\varkappa$ is at least $(a_\star + \varkappa/2)^m$. It follows that, for all $f \in \mathcal{B}_\star$, we have

$$e_{1,1,b}^*(R(z)^m f) \leq \frac{3C_\Gamma \|f\|_{1,B}^*}{(a_\star + \varkappa/2)^m}.$$

This estimate, combined with the estimate in Theorem 4.2 implies that whenever $|b| \geq b_\star$, we have $\|R(z)^m\|_{1,B}^*$ is $O_\Gamma((a_\star + \sigma_1)^{-m})$, where $\sigma_1 > 0$ is the minimum of σ_\star and $\varkappa/2$. In particular, if $|b|$ is large enough, depending on Γ , we can absorb the implied constant in the estimate above to obtain

$$\|R(z)^m\|_{1,B}^* \leq (a_\star + \sigma_1/2)^{-m}.$$

Let $p \in \mathbb{N}$ be a large integer to be chosen shortly. To obtain the claimed estimate for the norm $\|\cdot\|_1^*$, note that for any f in the Banach space \mathcal{B}_\star , since $\|\cdot\|_{1,B}^* \leq \|\cdot\|_1^* \leq B \|\cdot\|_B^* = |b|^{1+\varkappa} \|\cdot\|_{1,B}^*$, iterating the above estimate yields

$$\|R(z)^{2pm} f\|_1^* \leq B \|R(z)^{2pm} f\|_B^* \leq \frac{B \|R(z)^{pm} f\|_{1,B}^*}{(a_\star + \sigma_1/2)^{pm}} \leq \frac{B \|f\|_1^*}{(a_\star + \sigma_1/2)^{2pm}}.$$

Since $m = \lceil \log |b| \rceil$, choosing p large enough, depending only on a_\star and σ_1 , we can ensure that $B/(a_\star + \sigma_1/2)^{pm} \leq 1/a_\star^{pm}$. In particular, taking λ_\star to be the positive solution of the quadratic polynomial $x \mapsto x^2 + 2a_\star x - a_\star \sigma_1/2$, we obtain the desired estimate with $c_\star = 2p$. \square

8.3. Proof of Theorems 1.1 and 1.2. Let \mathfrak{X} denote the generator of the semigroup \mathcal{L}_t acting on \mathcal{B}_* (which exists by [Kha23, Cor. 7.2]). In light of the above results and Theorem 8.1, we obtain the following decomposition of the transfer operator \mathcal{L}_t : $\mathcal{L}_t = \mathcal{P}_t + \sum_{i=1}^N e^{t\lambda_i} e^{t\mathcal{N}_i} \Pi_i$, where $\mathcal{P}_t, \mathcal{N}_i, \lambda_i$ and Π_i are as in Theorem 8.1. Moreover, for a suitable $\sigma > 0$, depending only on λ_* in Corollary 8.4 and on σ_0 given by Theorem 3.3, we have that $\|\mathcal{P}_t f\|_1' \ll e^{-\sigma t} \|\mathfrak{X}f\|_1^*$, for all $t \geq 0$ and $f \in \mathcal{B}_*$. Finally, it follows by [Kha23, Lemma 7.12] that the only eigenvalue λ_i lying on the imaginary axis is 0 and that its associated nilpotent operator \mathcal{N}_i vanishes. This concludes the proof.

REFERENCES

- [AGY06] Artur Avila, Sébastien Gouëzel, and Jean-Christophe Yoccoz, *Exponential mixing for the Teichmüller flow*, Publ. Math. Inst. Hautes Études Sci. (2006), no. 104, 143–211.
- [AM16] Vitor Araújo and Ian Melbourne, *Exponential decay of correlations for nonuniformly hyperbolic flows with a $C^{1+\alpha}$ stable foliation, including the classical Lorenz attractor*, Ann. Henri Poincaré **17** (2016), no. 11, 2975–3004.
- [BD17] Jean Bourgain and Semyon Dyatlov, *Fourier dimension and spectral gaps for hyperbolic surfaces*, Geom. Funct. Anal. **27** (2017), no. 4, 744–771.
- [BDL18] Viviane Baladi, Mark F. Demers, and Carlangelo Liverani, *Exponential decay of correlations for finite horizon Sinai billiard flows*, Invent. Math. **211** (2018), no. 1, 39–177.
- [BL12] Viviane Baladi and Carlangelo Liverani, *Exponential decay of correlations for piecewise cone hyperbolic contact flows*, Comm. Math. Phys. **314** (2012), no. 3, 689–773.
- [Bow93] B.H. Bowditch, *Geometrical finiteness for hyperbolic groups*, Journal of Functional Analysis **113** (1993), no. 2, 245–317.
- [But16a] Oliver Butterley, *A note on operator semigroups associated to chaotic flows*, Ergodic Theory Dynam. Systems **36** (2016), no. 5, 1396–1408.
- [But16b] ———, *A note on operator semigroups associated to chaotic flows—corrigendum*, Ergodic Theory Dynam. Systems **36** (2016), no. 5, 1409–1410.
- [BY] Amlan Banaji and Han Yu, *Fourier transforms and nonlinear images of self-similar sets and measures*.
- [Cor90] Kevin Corlette, *Hausdorff dimensions of limit sets. I*, Invent. Math. **102** (1990), no. 3, 521–541.
- [CS22] Michael Chow and Pratyush Sarkar, *Exponential mixing of frame flows for convex cocompact locally symmetric spaces*, arXiv e-prints (2022), arXiv:2211.14737.
- [DFSU21] Tushar Das, Lior Fishman, David Simmons, and Mariusz Urbański, *Extremality and dynamically defined measures, part II: measures from conformal dynamical systems*, Ergodic Theory Dynam. Systems **41** (2021), no. 8, 2311–2348.
- [DG16] Semyon Dyatlov and Colin Guillarmou, *Pollicott-Ruelle resonances for open systems*, Ann. Henri Poincaré **17** (2016), no. 11, 3089–3146.
- [Dol98] Dmitry Dolgopyat, *On decay of correlations in Anosov flows*, Ann. of Math. (2) **147** (1998), no. 2, 357–390.
- [ELO22] Sam Edwards, Minju Lee, and Hee Oh, *Torus counting and self-joinings of Kleinian groups*, arXiv e-prints (2022), arXiv:2210.10229.
- [EO21] Samuel C. Edwards and Hee Oh, *Spectral gap and exponential mixing on geometrically finite hyperbolic manifolds*, Duke Math Journal, to appear (2021), arXiv:2001.03377.
- [FL09] De-Jun Feng and Ka-Sing Lau, *Multifractal formalism for self-similar measures with weak separation condition*, J. Math. Pures Appl. (9) **92** (2009), no. 4, 407–428.
- [FNW02] De-Jun Feng, Nhu T. Nguyen, and Tonghui Wang, *Convolutions of equicontractive self-similar measures on the line*, Illinois J. Math. **46** (2002), no. 4, 1339–1351.
- [FS90] L. Flaminio and R. J. Spatzier, *Geometrically finite groups, Patterson-Sullivan measures and Ratner’s rigidity theorem*, Invent. Math. **99** (1990), no. 3, 601–626.
- [GL06] Sébastien Gouëzel and Carlangelo Liverani, *Banach spaces adapted to Anosov systems*, Ergodic Theory Dynam. Systems **26** (2006), no. 1, 189–217.
- [GL08] ———, *Compact locally maximal hyperbolic sets for smooth maps: fine statistical properties*, J. Differential Geom. **79** (2008), no. 3, 433–477.
- [GLP13] P. Giulietti, C. Liverani, and M. Pollicott, *Anosov flows and dynamical zeta functions*, Ann. of Math. (2) **178** (2013), no. 2, 687–773.
- [GPL22] Paolo Giulietti, Mark Pollicott, and Carlangelo Liverani, *Anosov flows and dynamical zeta functions (errata)*, 2022.
- [Hoc14] Michael Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Ann. of Math. (2) **180** (2014), no. 2, 773–822.

- [Hoc15] ———, *On self-similar sets with overlaps and inverse theorems for entropy in \mathbb{R}^d* , arXiv e-prints (2015), arXiv:1503.09043.
- [JN12] Dmitry Jakobson and Frédéric Naud, *On the critical line of convex co-compact hyperbolic surfaces*, Geom. Funct. Anal. **22** (2012), no. 2, 352–368.
- [Kat94] Anatole Katok, *Infinitesimal Lyapunov functions, invariant cone families and stochastic properties of smooth dynamical systems*, Ergodic Theory Dynam. Systems **14** (1994), no. 4, 757–785, With Keith Burns.
- [Kau84] Robert Kaufman, *On Bernoulli convolutions*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math., vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 217–222.
- [Kha23] Osama Khalil, *Mixing, resonances, and spectral gaps on geometrically finite locally symmetric spaces*, 2023.
- [Kna02] Anthony W. Kna, *Lie groups beyond an introduction*, second ed., Progress in Mathematics, vol. 140, Birkhäuser Boston, Inc., Boston, MA, 2002.
- [Liv04] Carlangelo Liverani, *On contact Anosov flows*, Ann. of Math. (2) **159** (2004), no. 3, 1275–1312.
- [LNP21] Jialun Li, Frédéric Naud, and Wenyu Pan, *Kleinian Schottky groups, Patterson-Sullivan measures, and Fourier decay*, Duke Math. J. **170** (2021), no. 4, 775–825, With an appendix by Li.
- [LP82] Peter D. Lax and Ralph S. Phillips, *The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces*, J. Functional Analysis **46** (1982), no. 3, 280–350.
- [LP23] Jialun Li and Wenyu Pan, *Exponential mixing of geodesic flows for geometrically finite hyperbolic manifolds with cusps*, Invent. Math. **231** (2023), no. 3, 931–1021.
- [LPS23] Jialun Li, Wenyu Pan, and Pratyush Sarkar, *Exponential mixing of frame flows for geometrically finite hyperbolic manifolds*, arXiv e-prints (2023), arXiv:2302.03798.
- [MN20] Michael Magee and Frédéric Naud, *Explicit spectral gaps for random covers of Riemann surfaces*, Publ. Math. Inst. Hautes Études Sci. **132** (2020), 137–179.
- [MN21] Michael Magee and Frédéric Naud, *Extension of Alon’s and Friedman’s conjectures to Schottky surfaces*, arXiv e-prints (2021), arXiv:2106.02555.
- [MO15] Amir Mohammadi and Hee Oh, *Matrix coefficients, counting and primes for orbits of geometrically finite groups*, J. Eur. Math. Soc. (JEMS) **17** (2015), no. 4, 837–897.
- [Mos55] G. D. Mostow, *Self-adjoint groups*, Ann. of Math. (2) **62** (1955), 44–55.
- [Mos73] ———, *Strong rigidity of locally symmetric spaces*, Annals of Mathematics Studies, No. 78, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1973.
- [MS18] Carolina A. Mosquera and Pablo S. Shmerkin, *Self-similar measures: asymptotic bounds for the dimension and Fourier decay of smooth images*, Ann. Acad. Sci. Fenn. Math. **43** (2018), no. 2, 823–834.
- [Nau05] Frédéric Naud, *Expanding maps on Cantor sets and analytic continuation of zeta functions*, Ann. Sci. École Norm. Sup. (4) **38** (2005), no. 1, 116–153.
- [Pat76] S. J. Patterson, *The limit set of a Fuchsian group*, Acta Math. **136** (1976), no. 3-4, 241–273.
- [PPS15] Frédéric Paulin, Mark Pollicott, and Barbara Schapira, *Equilibrium states in negative curvature*, Astérisque (2015), no. 373, viii+281.
- [Rob03] Thomas Roblin, *Ergodicité et équidistribution en courbure négative*, Mém. Soc. Math. Fr. (N.S.) (2003), no. 95, vi+96.
- [RS20] Eino Rossi and Pablo Shmerkin, *On measures that improve L^q dimension under convolution*, Rev. Mat. Iberoam. **36** (2020), no. 7, 2217–2236.
- [Sch04] Barbara Schapira, *Lemme de l’ombre et non divergence des horosphères d’une variété géométriquement finie*, Ann. Inst. Fourier (Grenoble) **54** (2004), no. 4, 939–987.
- [Shm19] Pablo Shmerkin, *On Furstenberg’s intersection conjecture, self-similar measures, and the L^q norms of convolutions*, Ann. of Math. (2) **189** (2019), no. 2, 319–391.
- [Shm21] ———, *Slices and distances: on two problems of Furstenberg and Falconer*, arXiv (2021), arXiv:2109.12157.
- [Sto11] Luchezar Stoyanov, *Spectra of Ruelle transfer operators for axiom A flows*, Nonlinearity **24** (2011), no. 4, 1089–1120.
- [Sul79] Dennis Sullivan, *The density at infinity of a discrete group of hyperbolic motions*, Inst. Hautes Études Sci. Publ. Math. (1979), no. 50, 171–202.
- [SW20] Pratyush Sarkar and Dale Winter, *Exponential mixing of frame flows for convex cocompact hyperbolic manifolds*, arXiv e-prints (2020), arXiv:2004.14551.
- [Tit71] J. Tits, *Représentations linéaires irréductibles d’un groupe réductif sur un corps quelconque*, J. Reine Angew. Math. **247** (1971), 196–220.
- [Tsu15] Masato Tsujii, *On the Fourier transforms of self-similar measures*, Dyn. Syst. **30** (2015), no. 4, 468–484.
- [TV06] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.