

PAIR CORRELATION STATISTICS FOR DYNAMICAL SYSTEMS

SIMON BAKER AND MIKE TODD

ABSTRACT. We study the pair correlation statistics of orbits generated by maps on the interval. We show that under suitable mixing and multifractal assumptions, the pair correlation statistics of an orbit will almost surely exhibit the same asymptotic behaviour as a suitable sequence of i.i.d. random variables. We will also show that, under suitable hypotheses, the pair correlation statistics defined by two orbits will almost surely exhibit the same behaviour as two suitable sequences of i.i.d. random variables. Specific dynamical systems to which our results apply to include Gibbs-Markov maps and the Gauss map. We also give an example of a slowly mixing system for which the pair correlation statistics of an orbit almost surely behave distinctly to an i.i.d. sequence.

1. INTRODUCTION

Understanding the distribution of sequences of real numbers modulo one is a well-studied and important problem in mathematics [B, KP]. One of the first topics one encounters in this field is the notion of uniform distribution: a sequence of real numbers $(x_n)_n$ is *uniformly distributed modulo one* if for any $0 \leq a < b < 1$ we have

$$\lim_{n \rightarrow \infty} \frac{\#\{1 \leq i \leq n : x_i \bmod 1 \in [a, b]\}}{n} = b - a.$$

It is a consequence of the strong law of large numbers that if $(X_n)_n$ is a sequence of i.i.d. random variables distributed according to the uniform measure on $[0, 1]$, then $(X_n)_n$ will be uniformly distributed modulo one almost surely. This statement, and the concept of uniform distribution modulo one, can be generalised to other measures rather than just the uniform measure. It is a well studied problem to determine whether a sequence, often of some dynamical or number theoretic origins, is uniformly distributed with respect to some measure. The famous Birkhoff Ergodic Theorem [Wal, Theorem 1.14] shows that for ergodic, measure-preserving transformations

Date: 16th June 2026.

2020 *Mathematics Subject Classification.* 37B20, 37A05, 37E05 (Primary), 11K36 (Secondary).

Key words and phrases. Pair correlation statistics, Gibbs-Markov maps, Recurrence.

the orbit of a typical point will behave like a suitable sequence of i.i.d. random variables and be uniformly distributed with respect to the underlying measure.

A more fine scale understanding of the distribution of a sequence is provided by the pair correlation statistics. Namely, given a sequence of real numbers $(x_n)_n$, $\beta \in (0, 2)$ and $s > 0$, we would like to understand the asymptotic behaviour of

$$\# \left\{ 1 \leq i \neq j \leq n : \|x_i - x_j\| \leq \frac{s}{n^\beta} \right\}$$

as $n \rightarrow \infty$. Here and throughout we let $\|x\| = \min\{|x - n| : n \in \mathbb{Z}\}$ for $x \in \mathbb{R}$. It is known that if $\beta \in (0, 2)$ and $(X_n)_n$ is a sequence of i.i.d. random variables distributed according to the uniform measure on $[0, 1]$, then for any $s > 0$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2-\beta}} \# \left\{ 0 \leq i \neq j < n : \|X_i - X_j\| \leq \frac{s}{n^\beta} \right\} = 2s$$

almost surely. Just as in the case of uniform distribution, there is an analogue of this statement for more general measures than the uniform measure. Of particular interest is the case where $\beta = 1$. In this case a sequence of real numbers $(x_n)_n$ is said to have Poissonian pair correlations if

$$\lim_{n \rightarrow \infty} \frac{1}{n} \# \left\{ 0 \leq i \neq j < n : \|x_i - x_j\| \leq \frac{s}{n} \right\} = 2s$$

for all $s > 0$. It is known that if $(x_n)_n$ has Poissonian pair correlations then it is uniformly distributed modulo one [ALP, GL]. This result has been generalised to higher dimensions and sequence of points on manifolds in [M]. The opposite implication however does not hold, the sequence $(\alpha n)_n$ does not have Poissonian pair correlations for any $\alpha \in \mathbb{R}$, but is uniformly distributed modulo one whenever $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Thus the pair correlation statistics can be viewed as providing more detailed information about the distribution of a sequence than being uniformly distributed, and consequently they give a more sophisticated measure for how random a sequence is.

With the above in mind, suppose we were given a sequence of some dynamical or number theoretic origins: it is natural to ask whether it behaves like a sequence of i.i.d. random variables at the level of pair correlation statistics. This is a particularly natural question to ask in the dynamical context given that the discussion above demonstrates that pair correlation statistics can provide more detailed information than the conclusion of the Birkhoff Ergodic Theorem. Moreover, it fits into an ongoing and active area of research in dynamics that is concerned with determining to what extent orbits behave like a sequence of i.i.d random variables, for example Central Limit Theorems/Stable Laws and Large Deviations results stemming from works such as [Y, AD] and [MN, RY] respectively, and more recently functional limit theorems such as [MZ, FFT2]. We will also later mention results involving recurrence properties, which are more directly relevant to this paper, on hitting time statistics, shrinking target problems and extreme

value theory. This being said, very little is known about the pair correlation statistics of orbits generated by dynamical systems. Indeed to the best of the authors' knowledge, this is the first paper to consider this problem for general families of dynamical systems. The same cannot be said for number theory, where understanding the pair correlation statistics of a sequence of some number theoretic origin is an active area of research. The case when $(x_n)_n = (\alpha q_n)_n$ where (q_n) is a sequence of natural numbers and $\alpha \in \mathbb{R}$ has received significant attention [ALL, ALT, BCGW, W]. This is in part due to the connection between (αq_n) having Poissonian pair correlations for Lebesgue almost every α and the additive energy of the sequence $(q_n)_n$. There are also interesting connections between this problem and the Berry-Tabor conjecture from quantum mechanics [RS]. Another well studied problem is to determine those values of $\theta > 0$ for which $(n^\theta)_n$ has Poissonian pair correlations. We refer the reader to [LST] for more on this problem, and for a proof that (n^θ) has Poissonian pair correlations for all $\theta \in (0, 14/41)$. We also refer the reader to [EM] and [EBMV] where a deep connection between the distributional properties of $(n^{1/2})_n$ and homogeneous dynamics is exploited. In [AB] powers of real numbers were studied, and it was shown that for Lebesgue almost every $\alpha > 1$ the sequence $(\alpha^n)_n$ has Poissonian pair correlations.

In this paper we study the pair correlation statistics of orbits generated by maps $T : X \rightarrow X$ where $X \subset I = [0, 1]$. More formally, given $s > 0, \beta > 0, n \in \mathbb{N}$ and $x \in X$, we are interested in the almost sure asymptotic behaviour of

$$\# \left\{ 0 \leq i \neq j < n : |T^i x - T^j x| \leq \frac{s}{n^\beta} \right\}.$$

Proceeding via analogy with the i.i.d case, one would expect that if μ is a T -invariant probability measure that is suitably mixing, then for an interval of β , for μ -almost every x we should have

$$\lim_{n \rightarrow \infty} \frac{\# \left\{ 0 \leq i \neq j < n : |T^i x - T^j x| \leq \frac{s}{n^\beta} \right\}}{n^2 \int \mu(B(z, \frac{s}{n^\beta})) d\mu(z)} = 1 \quad (1.1)$$

for all $s > 0$. In Theorem B we verify this prediction for a family of interval maps and measures. Notice that the divisor here is n^2 multiplied by the probability that $|x - y| \leq s/n^\beta$ for x, y chosen independently according to μ , which if μ equals Lebesgue is $2sn^{2-\beta}$.

We will also consider the pair correlations statistics generated by two orbits. That is for $s > 0, \beta > 0, n \in \mathbb{N}$ and $x, y \in X$, we are interested in the almost sure asymptotic behaviour of

$$\# \left\{ 0 \leq i, j < n : |T^i x - T^j y| \leq \frac{s}{n^\beta} \right\}.$$

Again we would expect that if μ is a T -invariant probability measure that is suitably mixing, then for an interval of β , for $(\mu \times \mu)$ -almost every (x, y)

we should have

$$\lim_{n \rightarrow \infty} \frac{\#\{0 \leq i, j < n : d(T^i x, T^j y) \leq \frac{s}{n^\beta}\}}{n^2 \int \mu(B(z, \frac{s}{n^\beta})) d\mu(z)} = 1 \quad (1.2)$$

for all $s > 0$. Theorem A similarly verifies this prediction for a family of interval maps and measures.

Our main applications are to interval maps with good mixing properties, for example Gibbs-Markov interval maps and their absolutely continuous (with respect to Lebesgue measure) invariant probability measure (*acip*), but as we show, our results also apply more widely to invariant probability measures with non-trivial multifractal behaviour. We note here that a dynamical question similar to ours was considered in [NP], though for a special class of dynamical systems that were amenable to number theoretic techniques, in contrast to the dynamical systems we consider. We also consider cases where the conclusion of our theory does not hold, i.e. the pair correlation statistics of a μ -typical point do not coincide with those of a suitable i.i.d. sequence. Such examples are known to exist for ergodic transformations, e.g. by the discussion above irrational rotations of the circle have this property. This could however be attributed to the failure of the mixing property for these dynamical systems. It is natural to ask whether a quantitative rate of mixing would be sufficient to guarantee that the orbit of μ -almost every x behaves like an i.i.d. sequence from the perspective of pair correlation statistics. A natural place to look for examples, as in the case of the least distance problem (described below) in [RT], is interval maps with subexponential mixing. By considering a suitable class of Manneville-Pomeau maps, we show that even a polynomial rate of mixing is insufficient to guarantee this property (see Section 5.6).

We finish this opening discussion by highlighting a connection between pair correlation statistics and the notion of recurrence in dynamical systems. The foundational result in the study of recurrence is Poincaré's recurrence theorem [Wal, Theorem 1.4] which states that if (X, \mathcal{A}, μ) is a probability space and $T : X \rightarrow X$ is a measure preserving transformation, then for any $A \in \mathcal{A}$ for μ almost every $x \in A$ there exist infinitely many $n \in \mathbb{N}$ such that $T^n x \in A$. Under fairly modest assumptions, it can be shown that if X is equipped with a metric d , then Poincaré's recurrence theorem implies that $\liminf_{n \rightarrow \infty} d(T^n x, x) = 0$ for μ -almost every x . It is natural to ask whether a more quantitative statement holds. Following earlier work of Boshernitzan [B] this problem has recently received significant interest. We refer the reader to [ABB, BF, BK, BLR, B, FFT1, HT, KKP, HKKP, LLSV, Sau] and the references therein for more on this quantitative approach to recurrence. One approach to this problem is to study the smallest distance between all pairs of points chosen from an orbit, and between pairs of points chosen from two distinct orbits. More formally, this means understanding the quantities

$$M_n(x) := \min\{d(T^i x, T^j x) : 0 \leq i \neq j < n\}$$

and

$$M_n(x, y) := \min \{d(T^i x, T^j y) : 0 \leq i, j < n\}$$

for $x, y \in X$ and $n \in \mathbb{N}$. The study of $M_n(\cdot)$ and $M_n(\cdot, \cdot)$ is related to longest substring problems in probability [AW], which have applications in the comparison of DNA strands [RSW, Wat]. The study of pair correlation statistics can be viewed as a counting counterpart to this problem, where we count how many times a rate of recurrence is observed within an orbit. In [BLR], the authors considered how the quantity $M_n(x, y)$ scales with n for $(\mu \times \mu)$ -almost every (x, y) for μ a T -invariant probability measure. For sufficiently nice dynamical systems and measures it was shown that

$$\lim_{n \rightarrow \infty} \frac{-\log M_n(x, y)}{\log n} = \frac{2}{C_\mu}$$

for $(\mu \times \mu)$ -almost every (x, y) , where C_μ is the correlation dimension (defined below). This result is significant as it tells us that the optimal range of β for which we could expect the convergence in (1.1) and (1.2) to hold is $\beta \in (0, 2/C_\mu)$. Moreover, [BLR] introduces techniques which are useful for the problems addressed in the current paper. In the case of a single orbit, where there are extra challenges with dependence which don't exist in the two-orbit case, a similar result was proved in [Z]. A more precise analysis of the two-orbit case was done in [KKPT], where near-optimal results were proved for liminf and limsup subsets defined in terms of the behaviour of $M_n(\cdot, \cdot)$. We also note that the case of random iteration was considered in [GRS]. Recently for $(\alpha q_n)_n$, where $(q_n)_n$ is a sequence of integers and $\alpha \in \mathbb{R}$, the analogue of this minimal gaps problem was considered in [Ru, Re]. There the results relied on properties of the additive energy of the sequence.

1.1. Definitions and main theorems. We first define properties associated with our measures. Given a Borel probability measure μ supported on $[0, 1]$, we define the *upper correlation dimension* of μ to be

$$\overline{C}_\mu := \limsup_{r \searrow 0} \frac{\log \int \mu(B(x, r)) d\mu(x)}{\log r}.$$

Similarly we define \underline{C}_μ the *lower correlation dimension* of μ via the liminf. We generalise this to two Borel probability measures μ_1, μ_2 , by defining the *upper correlation dimension* of (μ_1, μ_2) to be

$$\overline{C}_{\mu_1, \mu_2} := \limsup_{r \searrow 0} \frac{\log \int \mu_1(B(x, r)) d\mu_2(x)}{\log r}$$

and similarly $\underline{C}_{\mu_1, \mu_2}$ using the liminf (note that by Fubini's theorem, switching μ_1 and μ_2 does not change these quantities).

Given $s > 0$ we say that a Borel probability measure μ is *s-Ahlfors regular* if there exists $C > 0$ such that

$$\frac{1}{C} r^s \leq \mu(B(x, r)) \leq C r^s$$

for all $x \in \text{supp}(\mu)$ and $0 < r < 1$. Particular examples of 1-Ahlfors regular measures are absolutely continuous invariant probability measures (acips) with density uniformly bounded away from zero and infinity. Acips are known to exist, and to satisfy the mixing assumptions appearing in our theorems, when T is a Gibbs-Markov map (see Definition 1.4). Gibbs-Markov maps include piecewise smooth full branched interval maps with bounded distortion. Some of the measures we will consider will not be s -Ahlfors regular for any $s > 0$. In this case a relevant scaling quantity is the Frostman dimension. We define the *Frostman dimension* of a Borel probability measure μ supported on $[0, 1]$ to be the supremum of those $s > 0$ for which there is $C > 0$ such that

$$\mu(B(x, r)) \leq Cr^s$$

for all $x \in [0, 1]$ and $0 < r < 1$. We denote the Frostman dimension of such a measure by F_μ .

Definition 1.1. We say that two Borel probability measures μ_1 and μ_2 have *continuous mean scaling* if for any $s, \beta, \varepsilon > 0$, there exists $\delta_0 > 0$ such that $0 \leq \delta \leq \delta_0$ implies

$$\left| 1 - \frac{\int \mu_1(B(x, (s \pm \delta)/n^\beta)) d\mu_2}{\int \mu_1(B(x, s/n^\beta)) d\mu_2} \right| < \varepsilon$$

for all $n \in \mathbb{N}$. If $\mu_1 = \mu_2$ we say that μ_1 has *continuous mean scaling*.

We next define the mixing properties of our dynamical systems.

Definition 1.2. Let (X, T, μ) denote a measure preserving system where $X \subset [0, 1]$, and let $(\mathcal{C}^1, \|\cdot\|_1)$ and $(\mathcal{C}^2, \|\cdot\|_2)$ be Banach spaces of observables on $[0, 1]$. If there exists $C, \theta > 0$ such that for all $\psi \in \mathcal{C}^1$ and $\phi \in \mathcal{C}^2$,

$$\left| \int \psi \cdot \phi \circ T^n d\mu - \int \psi d\mu \int \phi d\mu \right| \leq C \|\psi\|_1 \|\phi\|_2 e^{-\theta n},$$

then we say that (X, T, μ) has *exponential mixing for \mathcal{C}^1 against \mathcal{C}^2* .

We will usually assume that \mathcal{C}^1 are the functions of bounded variation BV with the usual norm $\|\psi\|_{BV} = |\psi|_{Var} + \|\psi\|_{L^1}$, and that $\mathcal{C}^2 = L^\infty$ with the sup norm.

Definition 1.3. Let (X, T, μ) denote a measure preserving system where $X \subset [0, 1]$. If there exist $C', \theta' > 0$ such that whenever $\mathbb{1}_1, \mathbb{1}_2, \mathbb{1}_3, \mathbb{1}_4$ are indicator functions on intervals and $0 \leq a < b \leq c$, then

$$\left| \int \mathbb{1}_1 \cdot \mathbb{1}_2 \circ T^a \cdot \mathbb{1}_3 \circ T^b \cdot \mathbb{1}_4 \circ T^c d\mu - \int \mathbb{1}_1 \cdot \mathbb{1}_2 \circ T^a d\mu \int \mathbb{1}_3 \cdot \mathbb{1}_4 \circ T^{c-b} d\mu \right| \leq C' e^{-\theta'(b-a)},$$

then we say that (X, T, μ) has *exponential 4-mixing for intervals*.

Note that in the definition of exponential 4-mixing for intervals that if we set $\mathbb{1}_1 = \mathbb{1}_4 = \text{Id}$ then we obtain a similar statement but for products of two indicator functions on intervals. This statement coincides with what one would obtain from assuming exponential mixing for BV against L^∞ . The exponential 4-mixing for intervals property is known to hold for Gibbs–Markov interval maps, see [Z, Lemma 4.16].

Our first main theorem is as follows; note that a version not assuming the existence of C_{μ_1, μ_2} is given later in Theorem 2.1.

Theorem A. Suppose that (X, T_1, μ_1) and (X, T_2, μ_2) satisfy the following properties:

- (X, T_i, μ_i) have exponential mixing for BV against L^∞ for $i = 1, 2$.
- C_{μ_1}, C_{μ_2} and C_{μ_1, μ_2} exist, belong to $(0, \infty)$, and satisfy $C_{\max} > 0$ for $C_{\max} := \max_{i=1,2} \{2C_{\mu_1, \mu_2} - C_{\mu_i}\}$.
- μ_1 and μ_2 have continuous mean scaling.

Then for $\beta \in (0, 2/C_{\max})$, for $(\mu_1 \times \mu_2)$ -a.e. (x, y) , for all $s > 0$,

$$\lim_{n \rightarrow \infty} \frac{\# \left\{ 0 \leq i, j < n : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta} \right\}}{n^2 \int \mu_1 \left(B \left(z, \frac{s}{n^\beta} \right) \right) d\mu_2(z)} = 1.$$

An example of an application of Theorem A is to the following type of interval map, which we give a particular version of.

Definition 1.4. Let $\tilde{X} \subset [0, 1]$ be a set that can be written as $\tilde{X} = \cup_{P \in \mathcal{P}} P$ where \mathcal{P} is an at most countable collection of intervals. Suppose $\tilde{T} : \tilde{X} \rightarrow I$ is a function satisfying the following properties:

- for each $P \in \mathcal{P}$, $\tilde{T}(P) \cap \tilde{X}$ is a union of elements of \mathcal{P} ;
- \tilde{T} is C^1 on each $P \in \mathcal{P}$ and there exist $\iota, C_\iota > 0$ such that for any $P \in \mathcal{P}$, if $x, y \in P$ then $|D\tilde{T}(x) - D\tilde{T}(y)| \leq C_\iota |x - y|^\iota$;
- there exists $\lambda > 1$ such that $|D\tilde{T}(x)| \geq \lambda$ for all $x \in \tilde{X}$;
- we have bounded distortion: there exists $C > 0$ such that if $x, y \in P \in \mathcal{P}$ then $\left| \frac{D\tilde{T}(x)}{D\tilde{T}(y)} - 1 \right| \leq C |x - y|$;
- there exists $b_0 > 0$ such that $\text{Diam}(\tilde{T}(P)) \geq b_0$ for all $P \in \mathcal{P}$.

Define the *attractor* to be

$$X = X_{\tilde{T}} := \left\{ x \in \tilde{X} : \tilde{T}^i x \in \cup_{P \in \mathcal{P}} P \text{ for all } i \geq 0 \right\}.$$

We denote the restriction of \tilde{T} to X by $T : X \rightarrow X$ and call T a *Gibbs–Markov map*.

Note that we will sometimes abuse notation and discuss T acting outside of the attractor X . If $X_{\tilde{T}} = I$, then these maps have a unique acip, with

density ρ in BV and bounded away from 0 and ∞ , and which satisfies the mixing conditions of our main theorems: in this case we will write $T = \tilde{T}$. Basic examples of these maps are $x \mapsto kx \pmod{1}$ for $k \in \mathbb{N}$, see also Farey maps and Lüroth maps (eg [KMS]). At the same time as considering these maps we will consider the *Gauss map* $x \mapsto 1/x \pmod{1}$ with acip given by the Gauss measure. We emphasise that the Gauss map behaves like a Gibbs-Markov map, see for example [BDT, Section 2.6.3], despite there being a point x where $|DT(x)| = 1$ (i.e., for $x = 1$). The proof of the following is given in Section 5.

Corollary 1.5. *Suppose that (I, T_i) are Gibbs-Markov maps or the Gauss map, with acips and densities denoted by μ_i and ρ_i for $i = 1, 2$, respectively. Then for $\beta \in (0, 2)$,*

$$\frac{1}{n^{2-\beta} \int (\rho_1 \rho_2)(t) dt} \# \left\{ 0 \leq i, j < n : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta} \right\} \rightarrow 2s,$$

for $\mu_1 \times \mu_2$ a.e. (x, y) .

So in the special case when $\mu_1 = \mu_2 = \text{Lebesgue}$, for $\beta \in (0, 2)$

$$\frac{1}{n^{2-\beta}} \# \left\{ 0 \leq i, j < n : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta} \right\} \rightarrow 2s.$$

A Gibbs-Markov map where \mathcal{P} is a finite set, its elements have disjoint closures and $\tilde{T}(P) = I$ for all $P \in \mathcal{P}$ is called a *cookie cutter*, a version of which is in [Ra]. Note that here $X_{\tilde{T}} \neq I$ and the Hausdorff dimension h of $X_{\tilde{T}}$ is strictly less than 1. If $\phi : \tilde{X} \rightarrow \mathbb{R}$ is Hölder on each $P \in \mathcal{P}$ then there is a unique equilibrium state $\mu = \mu_\phi$. Note one example of this is $\phi = -h \log |D\tilde{T}|$, in which case μ_ϕ is the natural ‘geometric measure’ and is h -Ahlfors regular. It is the analogue to the acips mentioned above in this case. The system (X, T, μ) then satisfies the conditions of our main theorems as we show in Section 5, and hence we have the conclusion of Theorem A for this family of maps and measures. This conclusion will also hold more generally for equilibrium states (see Section 5 for the definition).

Our second main theorem considers the case of a single orbit. This introduces significantly more dependence and has a much more involved proof. In this case we require an extra condition (as well as exponential 4-mixing for intervals). Given a measure preserving transformation (X, T, μ) where $X \subset [0, 1]$, let

$$A_r(n) := \{ x : |T^n x - x| < r \}.$$

We say that (X, T, μ) satisfies the *early return property* if there exist $C, s > 0$ such that

$$\mu(A_r(n)) \leq Cr^s \tag{1.3}$$

for all $r > 0$ and $n \in \mathbb{N}$. We let D_μ denote the supremum of those $s > 0$ for which (1.3) holds for some $C > 0$ and call D_μ the *early return exponent*. We give examples of systems satisfying the early return property in Section 5.

Theorem B. Assume that (X, T, μ) satisfies the following properties:

- (X, T, μ) has exponential mixing for BV against L^∞ and exponential 4-mixing for intervals.
- (X, T, μ) has the early return property with exponent D_μ .
- The correlation dimension of μ exists and is strictly positive. We denote it by C_μ .
- μ has continuous mean scaling.

Then for $\beta \in (0, 2/C_\mu)$ satisfying the following inequalities:

- $\beta(C_\mu - D_\mu) < 1$,
- $\beta(C_\mu - F_\mu) < 1$,

for μ almost every x , for all $s > 0$ we have

$$\lim_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j < n : |T^i x - T^j x| \leq \frac{s}{n^\beta}\}}{n^2 \int \mu(B(z, \frac{s}{n^\beta})) d\mu(z)} = 1.$$

As in the two-orbit case, if we have an acip with a density in BV , as for example in the case of Gibbs-Markov or Gauss maps, we have a result which is simpler to state.

Corollary 1.6. *Suppose that (I, T) is a Gibbs-Markov map or the Gauss map, with acip μ and density ρ . Then for all $\beta \in (0, 2)$, for μ -a.e. x and all $s > 0$,*

$$\frac{1}{n^{2-\beta} \int \rho(t)^2 dt} \#\left\{0 \leq i \neq j < n : |T^i x - T^j x| \leq \frac{s}{n^\beta}\right\} \rightarrow 2s.$$

So again in the special case when $\mu = \text{Lebesgue}$, for $\beta \in (0, 2)$

$$\frac{1}{n^{2-\beta}} \#\left\{0 \leq i \neq j < n : |T^i x - T^j x| \leq \frac{s}{n^\beta}\right\} \rightarrow 2s,$$

for Lebesgue-a.e. x . We also have applications to cookie cutters and their associated equilibrium states, see Section 5.

Remark 1.7. *We note that there are other possible shrinking sequences that we could use to define our pair correlation statistics that we can cover using our techniques, but the polynomial decay is a helpful property here, and the literature is mostly concerned with sequences of the form $(s/n^\beta)_n$, so we restrict ourselves to these sequences here.*

Structure of the paper: The rest of the paper is organised as follows. In Section 2 we prove Theorem 2.1 which implies Theorem A. Our proof of Theorem B is split across Sections 3 and 4. In Section 3 we introduce a conditioning argument that will underpin our proof of Theorem B and obtain some useful expectation bounds. Finally in Section 4 we complete

our proof by obtaining suitable second moment estimates. In Section 5 we explain why Corollaries 1.5 and 1.6 follow from our theorems and include some additional examples. We also include in this section the details for the aforementioned example of an interval map and invariant measure with a polynomial rate of mixing for which the orbit of almost every point does not have the expected pair correlation statistics.

Notation: For sequences $(a_n)_n$ and $(b_n)_n$ in $[0, \infty)$ we write $a_n \lesssim b_n$ if there is $C > 0$ such that $a_n \leq Cb_n$ for all n . The notation $a_n \gtrsim b_n$ is defined similarly. If $a_n \lesssim b_n$ and $a_n \gtrsim b_n$, we write $a_n \asymp b_n$. We write $a_n \sim b_n$ if $\frac{a_n}{b_n} \rightarrow 1$. For $(c_n)_n$ with $c_n \in \mathbb{R}$ we also write $c_n = \mathcal{O}(b_n)$ if $|c_n| \lesssim b_n$. Finally, if $(a_n)_n$ in $[0, \infty)$ satisfies $\lim_{n \rightarrow \infty} a_n = 0$, then we will sometimes write $a_n = o(1)$.

Acknowledgements. SB was partially funded by his EPSRC New Investigators Award (EP/W003880/1). MT was partially funded by his EPSRC grant UKRI1120. MT also thanks the University of Loughborough for their hospitality.

2. PROOF FOR THE TWO ORBIT CASE

Theorem A follows immediately from the following more general statement. We note that after setting γ to be any positive constant less than $2 - \beta C_{\mu_1, \mu_2}$ so that (2.1) is satisfied, to satisfy (2.2) for some $\alpha \in (0, 2\gamma)$ for $i = 1$ and $i = 2$, it is sufficient to satisfy that $\beta(2C_{\mu_1, \mu_2} - C_{\mu_1})$ and $\beta(2C_{\mu_1, \mu_2} - C_{\mu_2})$ are less than 2. This is precisely the second assumption in Theorem A.

Theorem 2.1. *Suppose that (X, T_i, μ_i) have exponential mixing for BV against L^∞ for $i = 1, 2$ and μ_1 and μ_2 have continuous mean scaling. Let $\beta > 0$. For $\gamma \in (0, 2)$, suppose that for all $s > 0$,*

$$n^2 \int \mu_1 \left(B \left(x, s/n^\beta \right) \right) d\mu_2(x) \gtrsim n^\gamma, \quad (2.1)$$

where the implied constant may depend on s . Moreover, suppose that there is $\alpha \in (0, 2\gamma)$ such that

$$n^2 \int \mu_i \left(B \left(x, s/n^\beta \right) \right) d\mu_i(x) \lesssim n^\alpha \quad (2.2)$$

for $i = 1$ and $i = 2$, where again the implied constant may depend on s .

Then for $\mu_1 \times \mu_2$ a.e. (x, y) , for all $s > 0$ we have

$$\lim_{n \rightarrow \infty} \frac{\# \left\{ 0 \leq i, j < n : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta} \right\}}{n^2 \int \mu_1 \left(B \left(z, s/n^\beta \right) \right) d\mu_2(z)} = 1.$$

The following lemma clarifies how (2.1), (2.2) and C_{μ_1, μ_2} are related.

- Lemma 2.2.** (a) If $\overline{C}_{\mu_1, \mu_2} < \infty$ then (2.1) is satisfied for s, β, γ whenever $\beta < \frac{2-\gamma}{\overline{C}_{\mu_1, \mu_2}}$.
- (b) If $\underline{C}_{\mu_1, \mu_2} > 0$ and (2.1) is satisfied for s, β, γ , then $\beta < \frac{2-\gamma}{\underline{C}_{\mu_1, \mu_2}}$.
- (c) If $\underline{C}_\mu > 0$ then (2.2) is satisfied for s, β, α whenever $\beta > \frac{2-\alpha}{\underline{C}_\mu}$.

Proof. For (a) it is sufficient to set $s = 1$. Then for $\varepsilon > 0$ and n sufficiently large,

$$\frac{\log \int \mu_1(B(x, 1/n^\beta)) d\mu_2(x)}{-\beta \log n} < \overline{C}_{\mu_1, \mu_2} + \varepsilon.$$

So

$$n^2 \int \mu_1(B(x, 1/n^\beta)) d\mu_2(x) > n^{2-\beta(\overline{C}_{\mu_1, \mu_2} + \varepsilon)}.$$

Therefore if $\beta < \frac{2-\gamma}{\overline{C}_{\mu_1, \mu_2} + \varepsilon}$, we are finished, so we conclude by noting that $\varepsilon > 0$ was arbitrary.

For (b), for $\varepsilon > 0$ and sufficiently large n , $\frac{\log \int \mu_1(B(x, s/n^\beta)) d\mu_2(x)}{\log(s/n^\beta)} \geq \underline{C}_{\mu_1, \mu_2} - \varepsilon$ and (2.1) imply

$$n^\gamma \lesssim n^2 \int \mu_1(B(x, s/n^\beta)) d\mu_2(x) \leq n^2 \left(\frac{s}{n^\beta}\right)^{\underline{C}_{\mu_1, \mu_2} - \varepsilon},$$

so $n^{-\beta} \gtrsim n^{\frac{\gamma-2}{\underline{C}_{\mu_1, \mu_2} - \varepsilon}}$ for all $\varepsilon > 0$, so $\beta < \frac{2-\gamma}{\underline{C}_{\mu_1, \mu_2}}$ as required.

Part (c) follows similarly. \square

We will require the following straightforward lemmas, see Lemmas 3.2 and 3.3 of [KKPT].

Lemma 2.3. *Suppose that $(X, \mu_1), (X, \mu_2)$ are Borel probability spaces with $X \subset \mathbb{R}^n$. Then there exists $K > 0$ such that if $r > 0$ is sufficiently small then for $y \in X$,*

$$\mu_i(B(y, r)) \leq K \left(\int \mu_i(B(x, r)) d\mu_i(x) \right)^{\frac{1}{2}} \quad \text{for } i = 1, 2.$$

Hence also

$$\int \mu_i(B(x, r))^2 d\mu_j(x) \leq K \left(\int \mu_i(B(x, r)) d\mu_i(x) \right)^{\frac{1}{2}} \int \mu_i(B(x, r)) d\mu_j(x).$$

for $(i, j) = (1, 2), (2, 1)$.

Lemma 2.4. *Let (X, μ) denote a Borel probability space where $X \subset \mathbb{R}$. For any $r > 0$, $\psi_r: X \rightarrow \mathbb{R}$ given by $y \mapsto \mu(B(y, r))$ is a function of bounded variation with total variation bounded above by 2.*

Remark 2.5. *We observe that the first part of Lemma 2.3 implies that if C_{μ_1}, C_{μ_2} and C_{μ_1, μ_2} exist and lie in $(0, \infty)$, as in Theorem A, then $C_{\max} \geq 0$. Recall that $C_{\max} = \max_{i=1,2} \{2C_{\mu_1, \mu_2} - C_{\mu_i}\}$.*

It will be useful in our proofs to use the following function.

Definition 2.6. Let $\ell : \mathbb{N} \rightarrow \mathbb{R}^+$ be a function satisfying the conditions:

- for any $\delta > 0$, $\frac{\ell(n)}{n^\delta} \rightarrow 0$ as $n \rightarrow \infty$;
- for any $p > 0$ and $\delta > 0$, $n^p e^{-\delta \ell(n)} \rightarrow 0$ as $n \rightarrow \infty$.

So for concreteness, we can choose $\ell(n) = \log n \log \log n$.

Proof of Theorem 2.1. Suppose, without loss of generality, that the exponential mixing constants are the same for both systems.

Let us fix $\beta > 0$, $\gamma \in (0, 2)$ and $\alpha \in (0, 2\gamma)$ so that our assumptions are satisfied. For $s > 0$, let $r_n = s/n^\beta$. Then for $x, y \in X$, define

$$S_n(x, y) := \sum_{i, j \in [0, n)} \mathbb{1}_{B(T_2^j y, r_n)}(T_1^i x).$$

Our proof depends on showing that the following second moment bound is satisfied: Let $\hat{\gamma} := \gamma - \frac{\alpha}{2} > 0$ (see (2.2)), then for all $s > 0$ there exists $C > 0$ such that

$$\mathbb{E} \left(\frac{S_n}{\mathbb{E}(S_n)} - 1 \right)^2 < \frac{C}{n^{\frac{\hat{\gamma}}{2}}}. \quad (2.3)$$

for all $n \in \mathbb{N}$. We will now explain why our result follows from (2.3) before returning to its proof. Using Markov's inequality, we deduce

$$\begin{aligned} (\mu_1 \times \mu_2) \left(\left| \frac{S_n}{\mathbb{E}(S_n)} - 1 \right| \geq \frac{1}{n^{\frac{\hat{\gamma}}{8}}} \right) &= (\mu_1 \times \mu_2) \left(\left(\frac{S_n}{\mathbb{E}(S_n)} - 1 \right)^2 \geq \frac{1}{n^{\frac{\hat{\gamma}}{4}}} \right) \\ &\leq n^{\frac{\hat{\gamma}}{4}} \mathbb{E} \left(\frac{S_n}{\mathbb{E}(S_n)} - 1 \right)^2 \leq \frac{C}{n^{\frac{\hat{\gamma}}{4}}}. \end{aligned}$$

So taking a subsequence $(n^K)_n$ where $K > 4/\hat{\gamma}$, we see that

$$(\mu_1 \times \mu_2) \left(\left| \frac{S_{n^K}}{\mathbb{E}(S_{n^K})} - 1 \right| \geq \frac{1}{n^{K \frac{\hat{\gamma}}{8}}} \right) \leq \frac{C}{n^{\frac{K\hat{\gamma}}{4}}},$$

which is summable. Hence, by the Borel-Cantelli lemma for $\mu_1 \times \mu_2$ a.e. (x, y) ,

$$\frac{S_{n^K}(x, y)}{\mathbb{E}(S_{n^K})} \rightarrow 1,$$

i.e.,

$$\lim_{n \rightarrow \infty} \frac{\# \left\{ 0 \leq i, j < n^K : |T_1^i x - T_2^j y| \leq \frac{s}{n^{\beta K}} \right\}}{n^{2K} \int \mu_1(B(z, s/n^{K\beta})) d\mu_2(z)} = 1. \quad (2.4)$$

This establishes our desired convergence along a subsequence for a fixed choice of s . It remains to establish this convergence along the integers and for arbitrary $s > 0$.

Recalling that our sequence of radii are $(s/n^\beta)_n$ where $s > 0$ was arbitrary, it follows from the above and the fact that a countable union of sets of measure zero has measure zero, that for $\mu_1 \times \mu_2$ a.e. (x, y) , for any $s \in \mathbb{Q}_{>0}$ we have

$$\lim_{n \rightarrow \infty} \frac{\#\left\{0 \leq i, j \leq n^K : |T_1^i x - T_2^j y| \leq \frac{s}{n^{K\beta}}\right\}}{n^{2K} \int \mu(B(z, s/n^{K\beta})) d\mu(z)} = 1. \quad (2.5)$$

We emphasise that here we are using the fact that $\hat{\gamma}$ and K are independent of s .

We will now explain why for $\mu_1 \times \mu_2$ a.e (x, y) for any $s > 0$ we have

$$\lim_{n \rightarrow \infty} \frac{\#\left\{0 \leq i, j \leq n : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta}\right\}}{n^2 \int \mu(B(z, s/n^\beta)) d\mu(z)} = 1.$$

Let us fix (x, y) belonging to the full measure set for which (2.5) holds for every $s \in \mathbb{Q}_{>0}$. For $n \in \mathbb{N}$ let $m \in \mathbb{N}$ be such that $m^K \leq n < (m+1)^K$, then for this choice of x, y we have

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{\#\{0 \leq i, j \leq n : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta}\}}{n^2 \int \mu(B(z, s/n^\beta)) d\mu(z)} \\ & \leq \limsup_{n \rightarrow \infty} \frac{\#\{0 \leq i, j \leq (m+1)^K : |T_1^i x - T_2^j y| \leq \frac{s}{n^\beta}\}}{m^{2K} \int \mu(B(z, s/(m+1)^{K\beta})) d\mu(z)} \\ & = \limsup_{n \rightarrow \infty} \frac{\#\{0 \leq i, j \leq (m+1)^K : |T_1^i x - T_2^j y| \leq \frac{s}{(m+1)^{K\beta}} \frac{(m+1)^{K\beta}}{n^\beta}\}}{m^{2K} \int \mu(B(z, s/(m+1)^{K\beta})) d\mu(z)} \\ & \leq \inf_{s' \in \mathbb{Q}: s' > s} \limsup_{n \rightarrow \infty} \frac{\#\{0 \leq i, j \leq (m+1)^K : |T_1^i x - T_2^j y| \leq \frac{s'}{(m+1)^{K\beta}}\}}{m^{2K} \int \mu(B(z, s'/(m+1)^{K\beta})) d\mu(z)} \\ & \leq \inf_{s' \in \mathbb{Q}: s' > s} \limsup_{n \rightarrow \infty} \frac{\#\{0 \leq i, j \leq (m+1)^K : |T_1^i x - T_2^j y| \leq \frac{s'}{(m+1)^{K\beta}}\}}{(m+1)^{2K} \int \mu(B(z, s'/(m+1)^{K\beta})) d\mu(z)} \\ & \quad \times \frac{\int \mu(B(z, s'/(m+1)^{K\beta})) d\mu(z)}{\int \mu(B(z, s/(m+1)^{K\beta})) d\mu(z)} \times \frac{(m+1)^{2K}}{m^{2K}} \\ & = 1 \times \inf_{s' \in \mathbb{Q}: s' > s} \limsup_{n \rightarrow \infty} \frac{\int \mu(B(z, s'/(m+1)^{K\beta})) d\mu(z)}{\int \mu(B(z, s/(m+1)^{K\beta})) d\mu(z)} = 1. \end{aligned}$$

In the final line we have used our assumption that μ_1 and μ_2 have continuous mean scaling. It can similarly be shown that

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} \frac{\#\{0 \leq i, j \leq n : |T^i x - T^j y| \leq \frac{s}{n^\beta}\}}{n^2 \int \mu(B(z, s/n^\beta)) d\mu(z)} \\
& \geq \liminf_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j \leq m^K : |T^i x - T^j y| \leq \frac{s}{n^\beta}\}}{(m+1)^{2K} \int \mu(B(z, s/m^{K\beta})) d\mu(z)} \\
& \geq \liminf_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j \leq m^K : |T^i x - T^j y| \leq \frac{s}{m^{K\beta}} \frac{m^{K\beta}}{n^\beta}\}}{(m+1)^{2K} \int \mu(B(z, s/m^{K\beta})) d\mu(z)} \\
& \geq \sup_{s' \in \mathbb{Q}: s' < s} \liminf_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j \leq m^K : |T^i x - T^j y| \leq \frac{s'}{m^{K\beta}}\}}{(m+1)^{2K} \int \mu(B(z, s/m^{K\beta})) d\mu(z)} \\
& \geq \sup_{s' \in \mathbb{Q}: s' < s} \liminf_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j \leq m^K : |T^i x - T^j y| \leq \frac{s'}{m^{K\beta}}\}}{m^{2K} \int \mu(B(z, s/m^{K\beta})) d\mu(z)} \\
& \quad \times \frac{\int \mu(B(z, s'/m^{K\beta})) d\mu(z)}{\int \mu(B(z, s/m^{K\beta})) d\mu(z)} \times \frac{m^{2K}}{(m+1)^{2K}} \\
& = 1 \times \sup_{s' \in \mathbb{Q}: s' < s} \liminf_{n \rightarrow \infty} \frac{\int \mu(B(z, s'/m^{K\beta})) d\mu(z)}{\int \mu(B(z, s/m^{K\beta})) d\mu(z)} = 1.
\end{aligned}$$

Where in the final line we used that μ_1 and μ_2 have continuous mean scaling. Thus (x, y) satisfies

$$\lim_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j \leq n : |T^i x - T^j y| \leq \frac{s}{n^\beta}\}}{n^2 \int \mu(B(z, s/n^\beta)) d\mu(z)} = 1$$

for any $s > 0$. This completes our proof up to verifying (2.3).

Let us fix $s > 0$ and define $r_n = s/n^\beta$. We now set out to prove (2.3) for this choice of s . Using T_1 -invariance of μ_1 and T_2 -invariance of μ_2 we first compute

$$\mathbb{E}(S_n) = n^2 \int \mu_1(B(y, r_n)) d\mu_2(y) \gtrsim n^\gamma. \tag{2.6}$$

Next, note that the left hand side of (2.3) is equal to $\frac{\mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2}{\mathbb{E}(S_n)^2}$. Since we already have an expression for $\mathbb{E}(S_n)$ we proceed to estimate $\mathbb{E}(S_n^2)$. We have

$$\mathbb{E}(S_n^2) = \sum_{i_1, i_2, j_1, j_2 \in [0, n]} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y).$$

In what follows we will often assume that $i_1 \leq i_2$ and $j_1 \leq j_2$ since for the three other combinations of inequalities, the upcoming arguments will give the same bound. Similarly to the proof of [BLR, Theorem 3], we will split the summation above according to the subcases corresponding to the following inequalities: $|i_2 - i_1| \leq \ell(n)$, $|i_2 - i_1| > \ell(n)$, $|j_2 - j_1| \leq \ell(n)$, $|j_2 - j_1| > \ell(n)$, where we recall ℓ from Definition 2.6.

For $i_2 - i_1 > \ell(n)$ and $j_2 - j_1 > \ell(n)$, i.e. *the totally separated case* (the cases $i_1 - i_2 > \ell(n)$ and/or $j_1 - j_2 > \ell(n)$ follow similarly), the following holds

$$\begin{aligned}
& \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ j_2 - j_1 > \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\
&= \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ j_2 - j_1 > \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(x) \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2 - i_1} x) d\mu_1(x) d\mu_2(y) \\
&\leq \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ j_2 - j_1 > \ell(n)}} \left(\int \mu_1(B(y, r_n)) \mu_1(B(T_2^{j_2 - j_1} y, r_n)) d\mu_2(y) \right. \\
&\qquad\qquad\qquad \left. + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right) \\
&\leq \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ j_2 - j_1 > \ell(n)}} \left(\left(\int \mu_1(B(y, r_n)) d\mu_2(y) \right)^2 + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right) \\
&\leq \frac{n^4}{4} \left(\left(\int \mu_1(B(y, r_n)) d\mu_2(y) \right)^2 + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right) \\
&= \frac{\mathbb{E}(S_n)^2}{4} + \mathcal{O}\left(n^4 e^{-\theta \ell(n)}\right),
\end{aligned}$$

where the first two inequalities follow from our exponential mixing assumptions, in the second inequality we have also used Lemma 2.4. Applying the same argument to the other three cases then gives us a bound of

$$\begin{aligned}
& \sum_{\substack{i_1, i_2 \in [0, n] \\ |i_2 - i_1| > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ |j_2 - j_1| > \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\
&\leq \mathbb{E}(S_n)^2 + \mathcal{O}\left(n^4 e^{-\theta \ell(n)}\right). \tag{2.7}
\end{aligned}$$

We emphasise that the second term on the right hand side of (2.7) decays faster than any polynomial by our choice of function ℓ (recall Definition 2.6).

For $0 \leq i_2 - i_1 \leq \ell(n)$ and $0 \leq j_2 - j_1 \leq \ell(n)$, i.e. *the totally non-separated case*, we have

$$\begin{aligned}
& \sum_{\substack{i_1, i_2 \in [0, n] \\ 0 \leq i_2 - i_1 \leq \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq j_2 - j_1 \leq \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\
& \leq \sum_{\substack{i_1, i_2 \in [0, n] \\ 0 \leq i_2 - i_1 \leq \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq j_2 - j_1 \leq \ell(n)}} \int \mu_1(B(T_2^{j_1} y, r_n)) d\mu_2(y) \\
& \leq \ell(n)^2 n^2 \int \mu_1(B(y, r_n)) d\mu_2(y) = \ell(n)^2 \mathbb{E}(S_n).
\end{aligned}$$

Again the cases where $0 \leq i_1 - i_2 \leq \ell(n)$ and/or $0 \leq j_1 - j_2 \leq \ell(n)$ follow similarly, and together yield

$$\begin{aligned}
& \sum_{\substack{i_1, i_2 \in [0, n] \\ 0 \leq |i_2 - i_1| \leq \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq |j_2 - j_1| \leq \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\
& \lesssim \ell(n)^2 \mathbb{E}(S_n). \tag{2.8}
\end{aligned}$$

For $i_2 - i_1 > \ell(n)$ and $0 \leq j_2 - j_1 \leq \ell(n)$, i.e. *a half-separated case*, we have

$$\begin{aligned}
& \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq j_2 - j_1 \leq \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\
& = \sum_{\substack{i_1, i_2, j_1, j_2 \in [0, n] \\ i_2 - i_1 > \ell(n) \\ 0 \leq j_2 - j_1 \leq \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2 - i_1} x) d\mu_1(x) d\mu_2(y) \\
& \leq \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq j_2 - j_1 \leq \ell(n)}} \left(\int \mu_1(B(T_2^{j_1} y, r_n)) \mu_1(B(T_2^{j_2} y, r_n)) d\mu_2(y) + \right. \\
& \qquad \qquad \qquad \left. + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right) \\
& \leq \sum_{\substack{i_1, i_2 \in [0, n] \\ i_2 - i_1 > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq j_2 - j_1 \leq \ell(n)}} \left(\int \mu_1(B(y, r_n))^2 d\mu_2(y) + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right) \\
& \leq n^3 \ell(n) \left(\int \mu_1(B(y, r_n))^2 d\mu_2(y) + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right),
\end{aligned}$$

where in the penultimate line we have used the Cauchy-Schwarz inequality and T_2 -invariance of μ_2 . In the case where $0 \leq i_2 - i_1 \leq \ell(n)$ and $j_2 - j_1 >$

$\ell(n)$, by an analogous argument we obtain:

$$\begin{aligned} & \sum_{\substack{i_1, i_2 \in [0, n] \\ 0 \leq i_2 - i_1 \leq \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ j_2 - j_1 > \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)} \\ & \leq n^3 \ell(n) \left(\int \mu_2(B(y, r_n))^2 d\mu_1(y) + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right). \end{aligned}$$

Similar estimates can be obtained in the half separated case when $i_2 \leq i_1$ and $j_2 \leq j_1$. In summary, in the half-separated case we have the following bound:

$$\begin{aligned} & \sum_{\substack{i_1, i_2 \in [0, n] \\ |i_2 - i_1| > \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ 0 \leq |j_2 - j_1| \leq \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\ & + \sum_{\substack{i_1, i_2 \in [0, n] \\ 0 \leq |i_2 - i_1| \leq \ell(n)}} \sum_{\substack{j_1, j_2 \in [0, n] \\ |j_2 - j_1| > \ell(n)}} \int \int \mathbb{1}_{B(T_2^{j_1} y, r_n)}(T_1^{i_1} x) \cdot \mathbb{1}_{B(T_2^{j_2} y, r_n)}(T_1^{i_2} x) d\mu_1(x) d\mu_2(y) \\ & \lesssim n^3 \ell(n) \left(\int \mu_1(B(y, r_n))^2 d\mu_2(y) + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right) \tag{2.9} \\ & + n^3 \ell(n) \left(\int \mu_2(B(y, r_n))^2 d\mu_1(y) + \mathcal{O}\left(e^{-\theta \ell(n)}\right) \right). \end{aligned}$$

Combining (2.7), (2.8), and (2.9) we have

$$\begin{aligned} \frac{\mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2}{\mathbb{E}(S_n)^2} & \lesssim \frac{n^4 e^{-\theta \ell(n)}}{\mathbb{E}(S_n)^2} + \frac{\ell(n)^2 \mathbb{E}(S_n)}{\mathbb{E}(S_n)^2} + \frac{n^3 \ell(n) \int \mu_1(B(y, r_n))^2 d\mu_2(y)}{\mathbb{E}(S_n)^2} \\ & + \frac{n^3 \ell(n) \int \mu_2(B(y, r_n))^2 d\mu_1(y)}{\mathbb{E}(S_n)^2} + \frac{n^3 e^{-\theta \ell(n)} \ell(n)}{\mathbb{E}(S_n)^2}. \end{aligned}$$

We now consider each of the terms on the right hand side of the above and show that they are each bounded above by a constant multiple of $n^{-\hat{\gamma}/2}$. This will in turn imply that (2.3) holds. The first and fifth terms decay to zero faster than any polynomial by our choice of function ℓ . Thus these two terms can be bounded above by a constant multiple of $n^{-\hat{\gamma}/2}$.

For the second term, using (2.6) we have

$$\frac{\ell(n)^2 \mathbb{E}(S_n)}{\mathbb{E}(S_n)^2} \lesssim \frac{\ell(n)^2}{n^\gamma} \lesssim \frac{1}{n^{\frac{\hat{\gamma}}{2}}}.$$

So this term satisfies the desired upper bound.

For the third term we apply Lemma 2.3 for $(i, j) = (1, 2)$ and use (2.1) and (2.2) to obtain

$$\begin{aligned} & \frac{n^3 \ell(n) \int \mu_1(B(y, r_n))^2 d\mu_2(y)}{\mathbb{E}(S_n)^2} \\ & \lesssim \frac{n^3 \ell(n) \left(\int \mu_1(B(y, r_n)) d\mu_1(y) \right)^{\frac{1}{2}} \int \mu_1(B(y, r_n)) d\mu_2(y)}{n^4 \left(\int \mu_1(B(y, r_n)) d\mu_2(y) \right)^2} \\ & = \frac{\ell(n) \left(n^2 \int \mu_1(B(y, r_n)) d\mu_1(y) \right)^{\frac{1}{2}}}{n^2 \int \mu_1(B(y, r_n)) d\mu_2(y)} \lesssim \frac{n^{\frac{\alpha}{2}} \ell(n)}{n^\gamma} \lesssim \frac{1}{n^{\frac{\gamma}{2}}}. \end{aligned}$$

So the third term satisfies the desired decay rate. By an analogous argument it can be shown that the fourth term is also bounded above by a constant multiple of $n^{-\tilde{\gamma}/2}$. This completes our proof. \square

3. ONE ORBIT CASE: CONDITIONING AND EXPECTATION

In this section and the next we fix (X, T, μ) satisfying the assumptions of Theorem B. We also fix $\beta > 0$ satisfying the inequalities listed in this theorem. By the definition of Frostman dimension and the early return exponent, we can choose $\tilde{F}_\mu, \tilde{D}_\mu, C > 0$ so that the following properties hold:

- $\mu(B(x, r)) \leq Cr^{\tilde{F}_\mu}$ for all $x \in X$ and $r > 0$.
- $\mu(A_r(n)) \leq Cr^{\tilde{D}_\mu}$ for all $r > 0$ and $n \in \mathbb{N}$.
- $\beta(C_\mu - \tilde{D}_\mu) < 1$.
- $\beta(C_\mu - \tilde{F}_\mu) < 1$.

We now set out to prove that μ -almost every x satisfies

$$\lim_{n \rightarrow \infty} \frac{\#\{0 \leq i \neq j < n : |T^i x - T^j x| \leq \frac{s}{n^\beta}\}}{n^2 \int \mu(B(z, \frac{s}{n^\beta})) d\mu(z)} = 1$$

for all $s > 0$. Assuming we have fixed $s > 0$ we associate the random variable

$$S_n(x) = \sum_{0 \leq i \neq j < n} \mathbb{1}_{B(T^i x, \frac{s}{n^\beta})}(T^j x).$$

We will prove Theorem B via a second moment analysis of S_n as in the proof of Theorem 2.1. However, in this case, because we are now considering a single orbit instead of two orbits, we have lost a degree of independence that was crucial in our earlier calculations, so for example, even computing the expectation of S_n becomes a significant task. To overcome this issue it is necessary to introduce a conditioning argument. We will outline this argument in the next subsection. In this section we use it to estimate the

expectation, or first moment, of our S_n . In the next section we will estimate the second moment and give the proof of Theorem B.

3.1. A conditioning argument. Given $s > 0$, in what follows we define the sequence $(r_n)_n$ via the equation

$$r_n = \frac{s}{n^\beta}$$

for all $n \in \mathbb{N}$. We suppress the dependence on s in this notation. When we refer to a sequence (r_n) we are implicitly making a choice of s .

The conditioning argument we will use throughout our proof involves introducing a parameter $\epsilon > 0$, and for each $n \in \mathbb{N}$ we let $m \in \mathbb{N}$ be the unique integer satisfying

$$2^{-m} < r_n^{1+\epsilon} \leq 2^{-m+1}. \quad (3.1)$$

In our notation we will suppress the dependence of m on ϵ . Our conditioning is provided by the dyadic partition provided by intervals of size 2^{-m} . With this in mind we introduce the following notation: let

$$I_0 := \left[0, \frac{1}{2^m}\right] \text{ and } I_k := \left(\frac{k}{2^m}, \frac{k+1}{2^m}\right] \text{ for } 1 \leq k \leq 2^m - 1.$$

Given $\epsilon > 0$ and $0 \leq k \leq 2^m - 1$ we also let

$$B_k^+ := B\left(\frac{k}{2^m}, r_n + \frac{1}{2^m}\right) \text{ and } B_k^- := B\left(\frac{k}{2^m}, r_n - \frac{1}{2^m}\right).$$

Again we suppress ϵ from our notation.

We observe that the partition $\{I_k\}_{k=0}^{2^m-1}$ has the property that for all $x, y \in I$ we have

$$\mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^-}(y) \leq \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B(x, r_n)}(y) \leq \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(y). \quad (3.2)$$

Moreover, it is an obvious fact that for $x \in [0, 1]$ and $i, j \in \mathbb{N}_0$ we have

$$\mathbb{1}_{B(T^i x, r_n)}(T^j x) = \sum_{k=0}^{2^m-1} \mathbb{1}_{I_k}(T^i x) \cdot \mathbb{1}_{B(T^i x, r_n)}(T^j x). \quad (3.3)$$

Combining (3.2) and (3.3) yields

$$\sum_{k=0}^{2^m-1} \mathbb{1}_{I_k}(T^i x) \mathbb{1}_{B_k^-}(T^j x) \leq \mathbb{1}_{B(T^i x, r_n)}(T^j x) \leq \sum_{k=0}^{2^m-1} \mathbb{1}_{I_k}(T^i x) \mathbb{1}_{B_k^+}(T^j x). \quad (3.4)$$

Crucially the left hand side and right hand side of (3.4) are in a form where our mixing assumptions can be applied. This is the important property that our conditioning approach provides.

We also highlight the following straightforward counting bounds that will be used throughout our proofs. Suppose that $\epsilon > 0$ and that m is defined

by (3.1), then for any $x \in [0, 1]$ we have

$$\sum_{k=0}^{2^m-1} \mathbb{1}_{B_k^+}(x) \lesssim 2^m r_n \lesssim n^{\beta\epsilon}. \quad (3.5)$$

We will also regularly use the trivial fact that for any $x \in [0, 1]$ we have

$$\sum_{k=0}^{2^m-1} \mathbb{1}_{I_k}(x) = 1. \quad (3.6)$$

We will also need the following technical lemma which controls the error when $\int \mu(B(x, r_n)) d\mu(x)$ is approximated by a sum over the dyadic partition corresponding to a choice of ϵ and m . In the proof of this lemma and at many points later on in our discussion, we will need the following equality which follows from the definition of correlation dimension:

$$\int \mu(B(x, r_n)) d\mu(x) = n^{-\beta C_\mu + o(1)}. \quad (3.7)$$

Lemma 3.1. *Let $s > 0$. Given $\epsilon > 0$ and m defined by (3.1), we have the following*

$$\left| \sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^+) - \int \mu(B(x, r_n)) d\mu(x) \right| \lesssim \int \mu\left(B\left(x, \frac{8}{2^m}\right)\right) d\mu(x) \\ \lesssim n^{-\beta C_\mu(1+\epsilon/2)}$$

and

$$\left| \sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^-) - \int \mu(B(x, r_n)) d\mu(x) \right| \lesssim \int \mu\left(B\left(x, \frac{8}{2^m}\right)\right) d\mu(x) \\ \lesssim n^{-\beta C_\mu(1+\epsilon/2)}.$$

Moreover, we always have the bound

$$\sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^+) \lesssim n^{-\beta C_\mu + o(1)}.$$

Proof. In the proof below we just prove the first statement for the set of balls $\{B_k^+\}_{k=0}^{2^m-1}$. The second statement follows by an analogous argument. Our final bound is a consequence of our first statement and (3.7).

Let $g_m : [0, 1] \rightarrow \{\frac{k}{2^m}\}_{k=0}^{2^m-1}$ be the function given by the rule $g_m(x) = \frac{k}{2^m}$ if $x \in I_k$. Then

$$\begin{aligned}
& \left| \sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^+) - \int \mu(B(x, r_n)) d\mu(x) \right| \\
&= \left| \int \mu \left(B \left(g_m(x), r_n + \frac{1}{2^m} \right) \right) - \mu(B(x, r_n)) d\mu(x) \right| \\
&= \int \mu \left(B \left(g_m(x), r_n + \frac{1}{2^m} \right) \setminus B(x, r_n) \right) d\mu(x) \\
&\leq \int \mu \left(B \left(g_m(x) + r_n, \frac{1}{2^m} \right) \right) + \mu \left(B \left(g_m(x) - r_n, \frac{1}{2^m} \right) \right) d\mu.
\end{aligned}$$

Thus to prove our result it remains to establish suitable upper bounds for the two integrals appearing in the final line of the above. We will just show how to bound the first of these integrals. The other is handled similarly. Let $p \in \mathbb{N}$ be the unique integer satisfying

$$\frac{p}{2^m} \leq r_n < \frac{p+1}{2^m}.$$

Then we have the following

$$\begin{aligned}
& \int \mu \left(B \left(g_m(x) + r_n, \frac{1}{2^m} \right) \right) d\mu(x) \\
&\leq \sum_{k=0}^{2^m-1} \mu(I_k) \mu \left(B \left(\frac{k+p}{2^m}, \frac{3}{2^m} \right) \right) \\
&\leq \left(\sum_{k=0}^{2^m-1} \mu(I_k)^2 \right)^{1/2} \left(\sum_{k=0}^{2^m-1} \mu \left(B \left(\frac{k+p}{2^m}, \frac{3}{2^m} \right) \right)^2 \right)^{1/2} \quad (\text{Cauchy-Schwarz}) \\
&\leq \left(\sum_{k=0}^{2^m-1} \mu(I_k)^2 \right)^{1/2} \left(\sum_{k=0}^{2^m-1} \mu \left(B \left(\frac{k}{2^m}, \frac{3}{2^m} \right) \right)^2 \right)^{1/2}. \tag{3.8}
\end{aligned}$$

We now bound the two terms appearing in (3.8). The following bound is obvious

$$\left(\sum_{k=0}^{2^m-1} \mu(I_k)^2 \right)^{1/2} \leq \left(\int \mu \left(B \left(x, \frac{8}{2^m} \right) \right) d\mu(x) \right)^{1/2}. \tag{3.9}$$

For the second term we have

$$\begin{aligned}
\sum_{k=0}^{2^m-1} \mu \left(B \left(\frac{k}{2^m}, \frac{3}{2^m} \right) \right)^2 &= \sum_{k=0}^{2^m-1} \sum_{q=-3}^2 \mu(I_{k+q}) \mu \left(B \left(\frac{k}{2^m}, \frac{3}{2^m} \right) \right) \\
&\leq \sum_{k=0}^{2^m-1} \sum_{q=-3}^2 \mu(I_{k+q}) \mu \left(B \left(\frac{k+q}{2^m}, \frac{7}{2^m} \right) \right) \\
&\leq 6 \sum_{k=0}^{2^m-1} \mu(I_k) \mu \left(B \left(\frac{k}{2^m}, \frac{7}{2^m} \right) \right) \\
&\leq 6 \int \mu \left(B \left(x, \frac{8}{2^m} \right) \right) d\mu(x). \tag{3.10}
\end{aligned}$$

Substituting (3.9) and (3.10) into (3.8), and using the definition of correlation dimension and (3.1) now yields

$$\begin{aligned}
\int \mu \left(B \left(g_m(x) + r_n, \frac{1}{2^m} \right) \right) d\mu(x) &\leq 7 \int \mu \left(B \left(x, \frac{8}{2^m} \right) \right) d\mu(x) \\
&\lesssim n^{-\beta C_\mu(1+\epsilon/2)}.
\end{aligned}$$

This completes our proof. \square

3.2. Expectation estimate. We next give the main result from this section, an estimate on $\mathbb{E}(S_n)$. Recall that $\beta, \tilde{D}_\mu, \tilde{F}_\mu$ and C_μ satisfy

$$\beta(C_\mu - \tilde{D}_\mu) < 1, \beta(C_\mu - \tilde{F}_\mu) < 1, \beta C_\mu < 2.$$

Proposition 3.2. *There exists $\epsilon_1 > 0$ such that for any $s > 0$ we have*

$$\left| \mathbb{E}(S_n) - n^2 \int \mu(B(x, r_n)) d\mu(x) \right| \lesssim n^{2-\beta C_\mu - \epsilon_1}.$$

Proof. We start by stating the following simple bound and equality:

$$\sum_{\substack{i, j \in [0, n) \\ |j-i| \geq \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x) \leq \mathbb{E}(S_n) \tag{3.11}$$

and

$$\begin{aligned}
\mathbb{E}(S_n) &= \sum_{\substack{i, j \in [0, n) \\ |j-i| \geq \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x) \\
&\quad + \sum_{\substack{i, j \in [0, n) \\ |j-i| < \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x). \tag{3.12}
\end{aligned}$$

Thus to prove our result it is sufficient to obtain suitably good upper and lower bounds for the sum over the well separated indices, and an upper bound for the sum over close indices. We begin by proving an upper bound

for the summation over well separated indices. Let $\epsilon > 0$ and m be defined by (3.1). By our exponential 4-mixing on intervals assumption, the right hand inequality in (3.4), and Lemma 3.1 we have:

$$\begin{aligned}
& \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x) \\
& \leq \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \sum_{k=0}^{2^m-1} \int \mathbb{1}_{I_k}(T^i x) \cdot \mathbb{1}_{B_k^+}(T^j x) d\mu(x) \\
& = \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \sum_{k=0}^{2^m-1} \left(\mu(I_k) \mu(B_k^+) + \mathcal{O}(e^{-\theta \ell(n)}) \right) \\
& \leq \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \left(\int \mu(B(x, r_n)) d\mu(x) + \mathcal{O}\left(n^{-\beta C_\mu(1+\epsilon/2)}\right) \right) + \mathcal{O}(2^m e^{-\theta \ell(n)}) \\
& \leq \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \int \mu(B(x, r_n)) d\mu(x) + \mathcal{O}\left(n^{2-\beta C_\mu(1+\epsilon/2)}\right). \tag{3.13}
\end{aligned}$$

In the final line we have used that $2^m e^{-\theta \ell(n)}$ decays to zero faster than any polynomial (recall Definition 2.6), and so can be absorbed into our final error term. It can similarly be shown using the first inequality in (3.4) that

$$\begin{aligned}
& \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x) \tag{3.14} \\
& \geq \sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \int \mu(B(x, r_n)) d\mu(x) + \mathcal{O}\left(n^{2-\beta C_\mu(1+\epsilon/2)}\right).
\end{aligned}$$

Combining (3.7), (3.13), (3.14), and using that

$$\#\{i, j \in [0, n] : |j - i| \geq \ell(n)\} = n^2 + \mathcal{O}(n\ell(n)),$$

we have that there exists $\epsilon'_1 > 0$ sufficiently small that does not depend on our choice of s such that

$$\sum_{\substack{i,j \in [0,n] \\ |j-i| \geq \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x) = n^2 \int \mu(B(x, r_n)) d\mu(x) + \mathcal{O}(n^{2-\beta C_\mu - \epsilon'_1}). \tag{3.15}$$

We now need to bound the contribution coming from those indices satisfying $|j - i| < \ell(n)$. Using the T -invariance of μ and the inequality $\beta(C_\mu - \tilde{D}_\mu) < 1$,

it follows that there exists $\epsilon'_2 > 0$ sufficiently small such that does not depend on s such that the following holds:

$$\begin{aligned}
& \sum_{\substack{i,j \in [0,n] \\ |j-i| < \ell(n)}} \int \mathbb{1}_{B(T^i(x), r_n)}(T^j x) d\mu(x) \\
& \leq \sum_{\substack{i,j \in [0,n], i > j \\ |j-i| < \ell(n)}} \int \mathbb{1}_{A_{r_n}(i-j)}(T^j x) d\mu(x) \\
& \quad + \sum_{\substack{i,j \in [0,n], j > i \\ |j-i| < \ell(n)}} \int \mathbb{1}_{A_{r_n}(j-i)}(T^i x) d\mu(x) \\
& \leq \sum_{\substack{i,j \in [0,n], i > j \\ |j-i| < \ell(n)}} \int \mathbb{1}_{A_{r_n}(i-j)}(x) d\mu(x) \\
& \quad + \sum_{\substack{i,j \in [0,n], j > i \\ |j-i| < \ell(n)}} \int \mathbb{1}_{A_{r_n}(j-i)}(x) d\mu(x) \\
& = \sum_{\substack{i,j \in [0,n] \\ |j-i| < \ell(n)}} \mu(A_{r_n}(|j-i|)) \lesssim r_n^{\tilde{D}_\mu} \cdot n\ell(n) \lesssim n^{1-\tilde{D}_\mu\beta} \cdot \ell(n) \\
& \lesssim n^{2-\beta C_\mu - \epsilon'_2}. \tag{3.16}
\end{aligned}$$

In the penultimate line we have used the definition of r_n . In the final line we have used that $\beta(C_\mu - \tilde{D}_\mu) < 1$ is equivalent to $2 - \beta C_\mu > 1 - \tilde{D}_\mu\beta$, and that $\ell(n) = n^{o(1)}$. Combining (3.15) and (3.16) it is clear that our result holds if we take $\epsilon_1 = \min\{\epsilon'_1, \epsilon'_2\}$. \square

4. THE ONE ORBIT CASE: SECOND MOMENT ARGUMENT

We now turn our attention to $\mathbb{E}(S_n^2)$. We start by observing that

$$\mathbb{E}(S_n^2) = 4 \sum_{i_1 < j_1 \in [0,n]} \sum_{i_2 < j_2 \in [0,n]} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1} x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2} x) d\mu(x).$$

When analysing this expression for $\mathbb{E}(S_n^2)$ it is useful to consider the ordering of the quadruple (i_1, i_2, j_1, j_2) . We start by highlighting that because our indices always satisfy $i_1 < j_1$ and $i_2 < j_2$, there are six possible orderings

for these indices:

$$\begin{aligned}
i_1 \leq j_1 \leq i_2 \leq j_2 & \quad (\text{Case A}) \\
i_1 \leq i_2 \leq j_1 \leq j_2 & \quad (\text{Case B}) \\
i_1 \leq i_2 \leq j_2 \leq j_1 & \quad (\text{Case C}) \\
i_2 \leq j_2 \leq i_1 \leq j_1 & \quad (\text{Case D}) \\
i_2 \leq i_1 \leq j_2 \leq j_1 & \quad (\text{Case E}) \\
i_2 \leq i_1 \leq j_1 \leq j_2 & \quad (\text{Case F})
\end{aligned}$$

We emphasise that after relabelling we see that Case A is equivalent to Case D, Case B is equivalent to Case E, and Case C is equivalent to Case F. In what follows whenever we speak of a quadruple of indices (i_1, i_2, j_1, j_2) we will always implicitly assume that $i_1 < j_1$ and $i_2 < j_2$. In addition to focusing on these cases it will be necessary to consider the gaps between successive indices.

4.1. Well separated indices. We say that a quadruple $(i_1, i_2, j_1, j_2) \in [0, n]^4$ is *well separated* if all terms are separated by a factor at least $\ell(n)$, i.e.

$$\min \{|i_1 - i_2|, |i_1 - j_1|, |i_1 - j_2|, |i_2 - j_1|, |i_2 - j_2|, |j_1 - j_2|\} \geq \ell(n).$$

The following proposition considers the case when the indices are well separated, and as we will see, it captures the dominant term in our expansion of $\mathbb{E}(S_n^2)$.

Proposition 4.1. *There exists $\epsilon_2 > 0$ such that for any $s > 0$, if (i_1, i_2, j_1, j_2) are well separated then*

$$\left| \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) - \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 \right| \lesssim n^{-2\beta C_\mu - \epsilon_2}.$$

Proof. In this proof we will assume that we are in Case A. The other cases are handled similarly. We first establish a suitable upper bound. Let $\epsilon > 0$ and m be defined by (3.1). The following inequality is a consequence of the right hand side of (3.4):

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(T^{i_1}x) \cdot \mathbb{1}_{I_l}(T^{i_2}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2}x) d\mu(x). \quad (4.1)
\end{aligned}$$

Let us now fix $0 \leq k \leq 2^{m-1}$ and $0 \leq l \leq 2^{m-1}$. Repeatedly using the T -invariance of μ and our exponential mixing for BV against L^∞ assumption,

we have

$$\begin{aligned}
& \int \mathbb{1}_{I_k}(T^{i_1}x) \cdot \mathbb{1}_{I_l}(T^{i_2}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2}x) d\mu(x) \\
&= \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x) \\
&= \mu(I_k) \int \mathbb{1}_{I_l}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_2}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_2}x) d\mu(x) + \mathcal{O}\left(e^{-\theta\ell(n)}\right) \\
&= \mu(I_k) \mu(I_l) \int \mathbb{1}_{B_k^+}(x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-j_1}x) d\mu(x) + \mathcal{O}\left(e^{-\theta\ell(n)}\right) \\
&= \mu(I_k)\mu(I_l)\mu(B_k^+)\mu(B_l^+) + \mathcal{O}\left(e^{-\theta\ell(n)}\right).
\end{aligned}$$

Using the equality above, together with Lemma 3.1, (3.7) and (4.1) yields

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
&\leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \left(\mu(I_k)\mu(I_l)\mu(B_k^+)\mu(B_l^+) + \mathcal{O}\left(e^{-\theta\ell(n)}\right) \right) \\
&= \left(\int \mu(B(x, r_n)) d\mu(x) + \mathcal{O}(n^{-\beta C_\mu(1+\epsilon/2)}) \right)^2 + \mathcal{O}\left(4^m e^{-\theta\ell(n)}\right) \\
&= \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 + \mathcal{O}(n^{-\beta C_\mu(2+\epsilon/4)}) + \mathcal{O}\left(4^m e^{-\theta\ell(n)}\right).
\end{aligned}$$

Using that our second error term decays to zero faster than any polynomial, it follows from the above that there exists $\epsilon'_1 > 0$ that does not depend on our choice of s such that

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
&\leq \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 + \mathcal{O}(n^{-2\beta C_\mu - \epsilon'_1}). \tag{4.2}
\end{aligned}$$

It can similarly be shown using the left hand side of (3.4) that there exists $\epsilon'_2 > 0$ that does not depend on s such that

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
&\geq \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 + \mathcal{O}(n^{-2\beta C_\mu - \epsilon'_2}). \tag{4.3}
\end{aligned}$$

Together (4.2) and (4.3) imply our result. \square

4.2. Not well separated indices. We now consider what happens when our indices are not well separated. For our purposes, we just need to bound from above the contribution to $\mathbb{E}(S_n^2)$ coming from those terms. We proceed

via a case analysis based on how many large gaps there are between successive indices. Before doing this it is necessary to introduce some terminology. Given a quadruple of indices (i_1, i_2, j_1, j_2) we inductively define

$$\gamma_1 = \min\{i_1, i_2, j_1, j_2\}$$

and

$$\gamma_i = \min\{\{i_1, i_2, j_1, j_2\} \setminus \cup_{l=1}^{i-1} \{\gamma_l\}\}$$

for $2 \leq i \leq 4$. When there isn't a unique minimum then we choose arbitrarily. Notice that the set $\{\gamma_i\}$ satisfies $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\} = \{i_1, i_2, j_1, j_2\}$ and

$$\gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \gamma_4.$$

It is this second property that will be more useful for our purposes. The cases we will study are defined as follows.

- We say that (i_1, i_2, j_1, j_2) has *no gaps* if

$$\gamma_{i+1} - \gamma_i < \ell(n)$$

for all $1 \leq i \leq 3$.

- We say that (i_1, i_2, j_1, j_2) has *one gap* if there exists a unique $1 \leq i \leq 3$ such that

$$\gamma_{i+1} - \gamma_i \geq \ell(n).$$

- We say that (i_1, i_2, j_1, j_2) has *two gaps* if there exists precisely two indices $i, j \in \{1, 2, 3\}$ satisfying

$$\gamma_{i+1} - \gamma_i \geq \ell(n) \text{ and } \gamma_{j+1} - \gamma_j \geq \ell(n).$$

We emphasise that a quadruple (i_1, i_2, j_1, j_2) is either well separated, has no gaps, has one gap, or has two gaps. We will obtain bounds for each of these cases in turn. We recall again, for use in the next propositions, that β, \tilde{D}_μ and \tilde{F}_μ satisfy

$$\beta(C_\mu - \tilde{D}_\mu) < 1, \beta(C_\mu - \tilde{F}_\mu) < 1, \beta C_\mu < 2.$$

4.2.1. *The no gaps case.* In the case of no gaps we have the following bound:

Proposition 4.2. *There exists $\epsilon_3 > 0$ such that for any $s > 0$ we have*

$$\sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ has} \\ \text{no gaps}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{4-2\beta C_\mu - \epsilon_3}.$$

Proof. We start our proof by introducing two estimates. The first is the following straightforward counting bound that follows from our choice of ℓ :

$$\#\{(i_1, i_2, j_1, j_2) : (i_1, i_2, j_1, j_2) \text{ has no gaps}\} \lesssim n^{1+o(1)}.$$

The second estimate is the following integral bound:

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{-\beta \tilde{D}_\mu}.$$

Recalling that $i_1 < j_1$, we now see that this bound is a consequence of the following inequalities:

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
& \leq \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) d\mu(x) \\
& = \int \mathbb{1}_{A_{r_n}(j_1 - i_1)}(T^{i_1}(x)) d\mu(x) \\
& = \int \mathbb{1}_{A_{r_n}(j_1 - i_1)}(x) d\mu(x) \\
& = \mu(A_{r_n}(j_1 - i_1)) \lesssim n^{-\beta \tilde{D}_\mu}.
\end{aligned}$$

Combining these two estimates yields:

$$\begin{aligned}
& \sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ has no gaps}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
& \lesssim n^{1 - \beta \tilde{D}_\mu + o(1)}.
\end{aligned}$$

The existence of ϵ_3 now follows from the inequality $4 - 2\beta C_\mu > 1 - \beta \tilde{D}_\mu$. This inequality holds because of our assumptions $1 > \beta(C_\mu - \tilde{D}_\mu)$ and $2 > \beta C_\mu$. \square

4.2.2. *The one gap case.* The following proposition considers the case when the quadruple of indices has one gap.

Proposition 4.3. *There exists $\epsilon_4 > 0$ such that for any $s > 0$ we have*

$$\sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ has} \\ \text{one gap}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{4 - 2\beta C_\mu - \epsilon_4}.$$

Proof. We begin by stating the counting bound that follows from our choice of function ℓ :

$$\#\{(i_1, i_2, j_1, j_2) : (i_1, i_2, j_1, j_2) \text{ has one gap}\} \lesssim n^{2+o(1)}.$$

It remains to show that if (i_1, i_2, j_1, j_2) has one gap then

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{2 - 2\beta C_\mu - \epsilon'} \quad (4.4)$$

for some $\epsilon' > 0$ that does not depend on s . We will do this via a case analysis. We will start by studying Cases B and C.

Cases B and C. In these two cases we have either $i_1 \leq i_2 \leq j_1 \leq j_2$ or $i_1 \leq i_2 \leq j_2 \leq j_1$. Now notice that it follows from these inequalities that if we have one gap, then we must have either $j_1 - i_1 \geq \ell(n)$ or $j_2 - i_2 \geq \ell(n)$. Let us assume $j_1 - i_1 \geq \ell(n)$. The other subcase is handled similarly. Let $\epsilon > 0$

and m be defined by (3.1). Then by Lemma 3.1, (3.4), (3.6), exponential 4-mixing on intervals, and the T -invariance of μ , we have

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) \\
& = \sum_{k=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) \\
& \lesssim \sum_{k=0}^{2^m-1} \mu(I_k)\mu(B_k^+) + \mathcal{O}(2^m e^{-\theta\ell(n)}) = \mathcal{O}(n^{-\beta C_\mu + o(1)}).
\end{aligned}$$

It follows from the above and our assumption $\beta C_\mu < 2$, which implies $-\beta C_\mu < 2 - 2\beta C_\mu$, that there exists ϵ'_1 such that for any $s > 0$ we have

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{2-2\beta C_\mu - \epsilon'_1}.$$

Case A. Recall that in this case we have $i_1 \leq j_1 \leq i_2 \leq j_2$. If $j_1 - i_1 \geq \ell(n)$ or $j_2 - i_2 \geq \ell(n)$ then the arguments given for Cases B and C above can be replicated to obtain the required bound. It remains to consider the subcase where $i_2 - j_1 \geq \ell(n)$. In this case, the following is a consequence of (3.4), T -invariance of μ , exponential 4-mixing on intervals and (3.5). Let $\epsilon > 0$

and m be defined by (3.1), then we have

$$\begin{aligned}
& \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) \int \mathbb{1}_{I_l}(x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_2}x) d\mu(x) \\
& \quad + \mathcal{O}(4^m e^{-\theta \ell(n)}) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{A_{4r_n}(j_1-i_1)}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) \\
& \quad \times \int \mathbb{1}_{A_{4r_n}(j_2-i_2)}(x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_2}x) d\mu(x) + \mathcal{O}(4^m e^{-\theta \ell(n)}) \\
& \leq n^{2\beta\epsilon} \int \mathbb{1}_{A_{4r_n}(j_1-i_1)}(x) d\mu(x) \int \mathbb{1}_{A_{4r_n}(j_2-i_2)}(x) d\mu(x) + \mathcal{O}(4^m e^{-\theta \ell(n)}) \\
& \lesssim n^{2\beta\epsilon-2\beta\tilde{D}_\mu}.
\end{aligned}$$

In the final line we have used that there exists $C > 0$ such $\mu(A_r(n)) \leq Cr^{\tilde{D}_\mu}$ for all $r > 0$ and $n \in \mathbb{N}$, and the fact our error term decays to zero faster than any polynomial. It is a consequence of our assumption $1 > \beta(C_\mu - \tilde{D}_\mu)$ that if we choose ϵ sufficiently small then $2\beta\epsilon - 2\beta\tilde{D}_\mu < 2 - 2\beta C_\mu$. Therefore it follows from the above that in this case there exists $\epsilon'_2 > 0$ such that for any $s > 0$ we have

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{2-2\beta C_\mu - \epsilon'_2}.$$

(4.4) can be verified for Cases D, E and F via a similar argument and appealing to the symmetry between these cases and Cases A, B and C.

We've shown that (4.4) holds for some ϵ' for all $s > 0$ in all cases, taking the minimum of these ϵ' yields a uniform estimate and we can conclude that (4.4) holds. This completes our proof. \square

4.2.3. The two gaps case. We now consider the case when we have two gaps between our indices. We split our analysis into two cases across the following two propositions.

We begin by considering the case where there are two gaps between our indices and one of these gaps is between the two smallest indices, i.e., $\gamma_2 - \gamma_1 \geq \ell(n)$.

Proposition 4.4. *Let (i_1, i_2, j_1, j_2) have two gaps and $\gamma_2 - \gamma_1 \geq \ell(n)$, then there exists $\epsilon_5 > 0$ such that for any $s > 0$ we have*

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{1-2\beta C\mu - \epsilon_5}.$$

Proof. For concreteness let us suppose we are in one of Case A, Case B or Case C so $\min\{i_1, i_2, j_1, j_2\} = i_1$. The other cases are handled similarly. Let $\epsilon > 0$ and m be defined by (3.1). By T -invariance and (3.4) we have

$$\begin{aligned} & \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x). \end{aligned}$$

By exponential mixing for BV against L^∞ , T -invariance and the inequality $i_2 \leq j_2$, we can find $a, b, c \geq 0$ such that either $a = 0$ or $b = 0$ (depending on which of i_2 and j_1 is larger) and

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \int \mathbb{1}_{I_l}(T^a x) \cdot \mathbb{1}_{B_k^+}(T^b x) \cdot \mathbb{1}_{B_l^+}(T^c x) d\mu(x) + \mathcal{O}(4^m e^{-\theta \ell(n)}). \end{aligned}$$

We now want to bound the integral over the triple product appearing in this summation. An application of exponential 4-mixing on intervals allows us to bound this integral from above by a product of two integrals plus some small error. This product over two integrals is of the form where one integral is over one of our indicator functions and the other is over the product of our two remaining indicator functions. The two indicator functions appearing in this latter integral correspond to the successive indices that are close, i.e. the γ_i, γ_{i+1} satisfying $|\gamma_{i+1} - \gamma_i| < \ell(n)$. Consequently after an application of our exponential 4-mixing of intervals assumption and T -invariance there are three possible forms the resulting bound can take:

- **Case 1:** There exist $d, e \geq 0$ such that either $d = 0$ or $e = 0$ and

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \int \mathbb{1}_{I_l}(T^a(x)) \cdot \mathbb{1}_{B_k^+}(T^b x) \cdot \mathbb{1}_{B_l^+}(T^c x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \mu(I_l) \int \mathbb{1}_{B_k^+}(T^d(x)) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) + \mathcal{O}(4^m e^{-\theta \ell(n)}). \end{aligned}$$

- **Case 2:** There exist $d, e \geq 0$ such that either $d = 0$ or $e = 0$ and

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \int \mathbb{1}_{I_l}(T^a(x)) \cdot \mathbb{1}_{B_k^+}(T^b x) \cdot \mathbb{1}_{B_l^+}(T^c x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \mu(B_k^+) \int \mathbb{1}_{I_l}(T^d(x)) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) + \mathcal{O}(4^m e^{-\theta \ell(n)}). \end{aligned}$$

- **Case 3:** There exist $d, e \geq 0$ such that either $d = 0$ or $e = 0$ and

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \int \mathbb{1}_{I_l}(T^a(x)) \cdot \mathbb{1}_{B_k^+}(T^b x) \cdot \mathbb{1}_{B_l^+}(T^c x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \mu(B_l^+) \int \mathbb{1}_{B_k^+}(T^d(x)) \cdot \mathbb{1}_{I_l}(T^e(x)) d\mu(x) + \mathcal{O}(4^m e^{-\theta \ell(n)}). \end{aligned}$$

We now consider each one of these cases in turn.

Case 1. It follows from the definition of m and our Frostman parameter \tilde{F}_μ that $\mu(I_l) \lesssim n^{-\beta \tilde{F}_\mu}$ for all $0 \leq l \leq 2^m - 1$. Using this bound together with the T -invariance of μ , (3.5) and Lemma 3.1 we see that the following holds:

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \mu(I_l) \int \mathbb{1}_{B_k^+}(T^d(x)) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) \\ & \lesssim n^{\beta \epsilon - \beta \tilde{F}_\mu} \sum_{k=0}^{2^m-1} \mu(I_k) \int \mathbb{1}_{B_k^+}(x) d\mu(x) \\ & \lesssim n^{\beta \epsilon - \beta F_\mu} \sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^+) \\ & \lesssim n^{\beta \epsilon - \beta \tilde{F}_\mu - \beta C_\mu + o(1)}. \end{aligned}$$

Using our assumption that $\beta(C_\mu - \tilde{F}_\mu) < 1$, which is equivalent to $-\beta \tilde{F}_\mu - C_\mu < 1 - 2\beta C_\mu$, and that $4^m e^{-\theta \ell(n)}$ decays to zero faster than any polynomial, we see that if we choose our original parameter ϵ to be sufficiently small then there exists $\epsilon'_1 > 0$ such that in this case, for any $s > 0$

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1} x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2} x) d\mu(x) \lesssim n^{1-2\beta C_\mu - \epsilon'_1}.$$

Case 2. In this case it follows from the inequality $i_2 \leq j_2$ that $d = 0$. Using Lemma 3.1, our early returns assumption, and (3.5) the following holds:

$$\begin{aligned}
& \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_k) \mu(B_k^+) \int \mathbb{1}_{I_l}(T^d(x)) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) \\
&= \sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^+) \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_l}(x) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) \\
&\lesssim n^{-\beta C_\mu + o(1)} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_l}(x) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) \\
&\leq n^{-\beta C_\mu + o(1)} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{A_{4r_n}(e)}(x) \cdot \mathbb{1}_{B_l^+}(T^e(x)) d\mu(x) \\
&\lesssim n^{\beta\epsilon - \beta C_\mu + o(1)} \int \mathbb{1}_{A_{4r_n}(e)}(x) d\mu(x) \\
&\lesssim n^{\beta\epsilon - \beta \tilde{D}_\mu - \beta C_\mu + o(1)}.
\end{aligned}$$

Now using that our error terms decays to zero faster than any polynomial, and our assumption $\beta(C_\mu - \tilde{D}_\mu) < 1$, which is equivalent to $-\beta \tilde{D}_\mu - \beta C_\mu < 1 - 2\beta C_\mu$, it follows that if we choose our parameter ϵ to be sufficiently small then in this case there exists $\epsilon'_2 > 0$ such that for any $s > 0$ we have

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{1-2\beta C_\mu - \epsilon'_2}.$$

Case 3. In this case we will use the Frostman bound $\mu(B_l^+) \lesssim n^{-\beta \tilde{F}_\mu}$. Using this bound, (3.6) Lemma 3.1, the T -invariance of μ , and duplicating the analysis given in Case 1, we obtain:

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{-\beta \tilde{F}_\mu - \beta C_\mu + o(1)}.$$

Now using our assumption $\beta(C_\mu - \tilde{F}_\mu) < 1$, which is equivalent to $-\beta \tilde{F}_\mu - \beta C_\mu < 1 - 2\beta C_\mu$, we see that in this case there exists $\epsilon'_3 > 0$ such that for any $s > 0$ we have

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{1-2\beta C_\mu - \epsilon'_3}.$$

Our result now follows from the three cases considered above by taking $\epsilon_5 = \min\{\epsilon'_1, \epsilon'_2, \epsilon'_3\}$. \square

The following proposition addresses the remaining case when there are two gaps but the first two indices satisfy $\gamma_2 - \gamma_1 < \ell(n)$.

Proposition 4.5. *Let (i_1, i_2, j_1, j_2) have two gaps and $\gamma_2 - \gamma_1 < \ell(n)$, then there exists $\epsilon_6 > 0$ such that for and $s > 0$ we have*

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{1-2\beta C_\mu - \epsilon_6}.$$

Proof. For concreteness let us again suppose we are in one of Case A, Case B or Case C so $\min\{i_1, i_2, j_1, j_2\} = i_1$. The other cases are handled similarly. Let $\epsilon > 0$ and m be defined by (3.1). We start with the familiar bound that follows from the T -invariance of μ and (3.4):

$$\begin{aligned} & \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x). \end{aligned}$$

We emphasise that it follows from our assumptions that $\gamma_3 - \gamma_2 \geq \ell(n)$ and $\gamma_4 - \gamma_3 \geq \ell(n)$. We will now apply exponential 4-mixing on intervals to bound the integrals appearing in the summation above by the product of two integrals plus some error. Using the fact that $i_2 < j_2$, we see that after an application of 4-mixing we are left with two possibilities:

- **Case 1:** If $j_1 \geq i_2$ then either $i_1 \leq i_2 \leq j_1 \leq j_2$ (Case B) or $i_1 \leq i_2 \leq j_2 \leq j_1$ (Case C). In which case there exists $a, b \geq 0$ such that $a = 0$ (first subcase) or $b = 0$ (second subcase), $|a - b| \geq \ell(n)$, and

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) d\mu(x) \int \mathbb{1}_{B_k^+}(T^a x) \cdot \mathbb{1}_{B_l^+}(T^b x) d\mu(x) \\ & \quad + \mathcal{O}(4^m e^{-\theta \ell(n)}). \end{aligned}$$

- **Case 2:** If $j_1 < i_2$ then we must be in Case A where $i_1 \leq j_1 \leq i_2 \leq j_2$ since we always have $i_2 \leq j_2$. We also have the bound $j_2 - i_2 \geq \ell(n)$ by our underlying assumptions. In this case, an application of exponential 4-mixing on intervals yields

$$\begin{aligned} & \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_1}x) d\mu(x) \\ & \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) \int \mathbb{1}_{I_l}(x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_2}x) d\mu(x) \\ & \quad + \mathcal{O}(4^m e^{-\theta \ell(n)}). \end{aligned}$$

We consider these two cases in turn. Because the error terms appearing in these two cases decay to zero faster than any polynomial they can be ignored in what follows. As such, we will just focus on bounding the integral terms.

Case 1. In this case we use our mixing assumption in the second integral and then the Cauchy-Schwarz inequality to obtain:

$$\begin{aligned}
& \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) d\mu(x) \int \mathbb{1}_{B_k^+}(T^a x) \cdot \mathbb{1}_{B_l^+}(T^b x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(B_k^+) \mu(B_l^+) \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) d\mu(x) + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \leq \left(\sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(B_k^+)^2 \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) d\mu(x) \right)^{1/2} \\
& \quad \times \left(\sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(B_l^+)^2 \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) d\mu(x) \right)^{1/2} + \mathcal{O}(4^m e^{-\theta\ell(n)}).
\end{aligned}$$

Now using the Frostman bounds $\mu(B_k^+) \lesssim n^{-\beta\tilde{F}_\mu}$, $\mu(B_l^+) \lesssim n^{-\beta\tilde{F}_\mu}$ together with (3.6), and Lemma 3.1, it follows from the above that

$$\begin{aligned}
& \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{I_l}(T^{i_2-i_1}x) d\mu(x) \int \mathbb{1}_{B_k^+}(T^a x) \cdot \mathbb{1}_{B_l^+}(T^b x) d\mu(x) \\
& \leq \left(n^{-\beta\tilde{F}_\mu} \sum_{k=0}^{2^m-1} \mu(I_k) \mu(B_k^+) \right)^{1/2} \left(n^{-\beta\tilde{F}_\mu} \sum_{l=0}^{2^m-1} \mu(I_l) \mu(B_l^+) \right)^{1/2} \\
& \quad + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{-\beta\tilde{F}_\mu - \beta C_\mu + o(1)} + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{-\beta\tilde{F}_\mu - \beta C_\mu + o(1)}.
\end{aligned}$$

In the final line we have used that our error term decays to zero faster than any polynomial. Now using our assumption $\beta(C_\mu - \tilde{F}_\mu) < 1$, which is equivalent to $-\beta\tilde{F}_\mu - \beta C_\mu < 1 - 2\beta C_\mu$, it follows that in this case there exists $\epsilon'_1 > 0$ such that for all $s > 0$ we have

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{1-2\beta C_\mu - \epsilon'_1}. \quad (4.5)$$

Case 2. Using our mixing assumptions in the second integral, Lemma 3.1, our early return assumption, and (3.5) we obtain

$$\begin{aligned}
& \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) \int \mathbb{1}_{I_l}(x) \cdot \mathbb{1}_{B_l^+}(T^{j_2-i_2}x) d\mu(x) \\
& \leq \sum_{k=0}^{2^m-1} \sum_{l=0}^{2^m-1} \mu(I_l)\mu(B_l^+) \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{-\beta C_\mu + o(1)} \sum_{k=0}^{2^m-1} \int \mathbb{1}_{I_k}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{-\beta C_\mu + o(1)} \sum_{k=0}^{2^m-1} \int \mathbb{1}_{A_{4r_n}(j_1-i_1)}(x) \cdot \mathbb{1}_{B_k^+}(T^{j_1-i_1}x) d\mu(x) + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{\beta\epsilon - \beta C_\mu + o(1)} \int \mathbb{1}_{A_{4r_n}(j_1-i_1)}(x) d\mu(x) + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{\beta\epsilon - \beta\tilde{D}_\mu - \beta C_\mu + o(1)} + \mathcal{O}(4^m e^{-\theta\ell(n)}) \\
& \lesssim n^{\beta\epsilon - \beta\tilde{D}_\mu - \beta C_\mu + o(1)}.
\end{aligned}$$

In the final line we used that our error term decays to zero faster than any polynomial. It follows now from our assumption $\beta(C_\mu - \tilde{D}_\mu) < 1$, which is equivalent to $-\beta\tilde{D}_\mu - \beta C_\mu < 1 - 2\beta C_\mu$, that if ϵ is chosen to be sufficiently small, then there exists $\epsilon'_2 > 0$ such that for all $s > 0$ we have

$$\int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{1-2\beta C_\mu - \epsilon'_2}. \quad (4.6)$$

Our result now follows from (4.5) and (4.6) by choosing ϵ sufficiently small and letting $\epsilon_5 = \min\{\epsilon'_1, \epsilon'_2\}$. \square

Combining Propositions 4.4 and 4.5 we have the following result.

Proposition 4.6. *There exists $\epsilon_7 > 0$ such that for all $s > 0$ we have*

$$\sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ has} \\ \text{two gaps}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{4-2\beta C_\mu - \epsilon_7}.$$

Proof. This result follows immediately from Propositions 4.4, 4.5 and the following straightforward upper bound that follows from our choice of function ℓ (recall Definition 2.6):

$$\#\{(i_1, i_2, j_1, j_2) : (i_1, i_2, j_1, j_2) \text{ has two gaps}\} \lesssim n^{3+o(1)}.$$

\square

4.2.4. *The not well separated conclusion.* Now combining Propositions 4.2, 4.3 and 4.6 we can conclude the following statement.

Proposition 4.7. *There exists $\epsilon_8 > 0$ such that for all $s > 0$ we have*

$$\sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ is not} \\ \text{well separated}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \lesssim n^{4-2\beta C_\mu - \epsilon_8}.$$

4.3. Proof of Theorem B. Equipped with the expectation estimates established in the previous section we are now in a position to prove Theorem B. Our proof is similar to the proof of Theorem 2.1.

Proof of Theorem B. Our proof relies upon showing that there exists $\delta > 0$ such that for any $s > 0$ we have the following second moment estimate:

$$\mathbb{E} \left(\left(\frac{S_n}{E(S_n)} - 1 \right)^2 \right) = \frac{\mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2}{E(S_n)^2} \lesssim n^{-\delta}. \quad (4.7)$$

Once (4.7) is established we can apply the same arguments as used in the proof of Theorem 2.1 to complete our proof. In particular, we can apply Markov's inequality and the Borel-Cantelli lemma to show that for a fixed $s > 0$ we have the desired convergence for μ -almost every x along some sequence $(n^K)_n$ where K does not depend on s . We then use an approximation argument and our assumption that μ has continuous mean scaling to upgrade this statement to show that for μ -almost every x , for any $s > 0$ we have the desired convergence along the integers. The only minor difference between the proofs is that in the proof of Theorem 2.1 we use the simple formula for $\mathbb{E}(S_n)$ provided by (2.6). Whereas in the one orbit case we have to use the following more complicated formula for $\mathbb{E}(S_n)$ which is a consequence of Proposition 3.2:

$$\mathbb{E}(S_n) = n^2 \int \mu(B(x, r_n)) d\mu(x) + \mathcal{O}(n^{2-\beta C_\mu - \epsilon_1}).$$

Because the error term is of a lower order than the integral term this causes no significant issues.

We now turn our attention to establishing (4.7). By Proposition 3.2 and (3.7), to show that (4.7) holds for some $\delta > 0$ for all $s > 0$, it is sufficient to show that

$$\frac{\mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2}{n^4 \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2} \lesssim n^{-\delta} \quad (4.8)$$

for some $\delta > 0$ for all $s > 0$. Recall that

$$\mathbb{E}(S_n^2) = 4 \sum_{i_1 < j_1 \in [0, n]} \sum_{i_2 < j_2 \in [0, n]} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x).$$

Combining Propositions 4.1 and 4.7 with (3.7), and using the fact that

$$\begin{aligned} \#\{(i_1, i_2, j_1, j_2) : (i_1, i_2, j_1, j_2) \text{ is well separated } i_1 < j_1 \text{ and } i_2 < j_2\} \\ = \frac{n^4}{4} + \mathcal{O}(n^{3+o(1)}), \end{aligned}$$

for any $s > 0$ we have

$$\begin{aligned} & \left| \mathbb{E}(S_n^2) - n^4 \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 \right| \\ & \leq \left| 4 \sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ is} \\ \text{well separated}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \right. \\ & \quad \left. - n^4 \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 \right| \\ & + \left| 4 \sum_{\substack{(i_1, i_2, j_1, j_2) \\ (i_1, i_2, j_1, j_2) \text{ is not} \\ \text{well separated}}} \int \mathbb{1}_{B(T^{i_1}(x), r_n)}(T^{j_1}x) \mathbb{1}_{B(T^{i_2}(x), r_n)}(T^{j_2}x) d\mu(x) \right| \\ & \lesssim n^{4-2\beta C_\mu - \epsilon_2} + n^{4-2\beta C_\mu - \epsilon_8} + n^{3-2\beta C_\mu + o(1)} \\ & \lesssim n^{4-2\beta C_\mu - \min\{\epsilon_3, \epsilon_8\}}. \end{aligned} \tag{4.9}$$

In the last line we have assumed that $\min\{\epsilon_3, \epsilon_8\} < 1$, which we may freely do without loss of generality. Moreover by Proposition 3.2, (3.7) and the well known equality $x^2 - y^2 = (x - y)^2 + 2(x - y)y$ for $x, y \in \mathbb{R}$, for any $s > 0$ we have

$$\begin{aligned} \left| \mathbb{E}(S_n)^2 - n^4 \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2 \right| & \lesssim n^{4-2\beta C_\mu - 2\epsilon_1} + n^{4-2\beta C_\mu - \epsilon_1 + o(1)} \\ & \lesssim n^{4-2\beta C_\mu - \epsilon_1 + o(1)}. \end{aligned} \tag{4.10}$$

Combining the triangle inequality, (4.9), (4.10) and (3.7), for any $s > 0$ we have that

$$\frac{\mathbb{E}(S_n^2) - E(S_n)^2}{n^4 \left(\int \mu(B(x, r_n)) d\mu(x) \right)^2} \lesssim \frac{n^{4-2\beta C_\mu - \min\{\epsilon_1, \epsilon_3, \epsilon_8\} + o(1)}}{n^{4-2\beta C_\mu + o(1)}} \lesssim n^{-\min\{\epsilon_1, \epsilon_3, \epsilon_8\}/2}.$$

Taking $\delta = \min\{\epsilon_1, \epsilon_3, \epsilon_8\}/2$ we see that (4.8) holds and so (4.7) also holds. This completes our proof. \square

5. APPLICATIONS

Here we will first apply our results to acips for Gibbs-Markov maps and the Gauss map, which is the content of Corollaries 1.5 and 1.6. We then give examples of non-acip cases where our results apply, and finally an (acip) example of a slow mixing system where the conclusion of Corollary 1.6 does not hold.

5.1. Acips for Gauss and Gibbs-Markov maps. Here we will prove Corollaries 1.5 and 1.6.

Lemma 5.1. *If μ_1, μ_2 are acips on $[0, 1]$ with densities ρ_1, ρ_2 in BV , then*

$$\int \mu_1(B(x, r)) d\mu_2(x) \sim 2r \int (\rho_1 \rho_2)(x) dx.$$

Proof. Let $m \in \mathbb{N}$ be arbitrary. Then for $r < 1/2m$ we have

$$\begin{aligned} \int \mu_1(B(x, r)) d\mu_2(x) &= \sum_{l=0}^{m-1} \int_{[\frac{l}{m}+r, \frac{l+1}{m}-r]} \mu_1(B(x, r)) d\mu_2(x) \\ &\quad + \sum_{l=0}^{m-1} \int_{[\frac{l}{m}, \frac{l}{m}+r] \cup [\frac{l+1}{m}-r, \frac{l+1}{m}]} \mu_1(B(x, r)) d\mu_2(x). \end{aligned}$$

For the second term we have

$$\sum_{l=0}^{m-1} \int_{[\frac{l}{m}, \frac{l}{m}+r] \cup [\frac{l+1}{m}-r, \frac{l+1}{m}]} \mu_1(B(x, r)) d\mu_2(x) \leq 2r^2 m \|\rho_1\|_\infty \|\rho_2\|_\infty.$$

In our analysis of the first term we will use $\text{Var}_a^b(\rho_1)$ to denote the total variation of ρ_1 on a closed interval $[a, b]$. In this case we have

$$\begin{aligned}
& \sum_{l=0}^{m-1} \int_{[\frac{l}{m}+r, \frac{l+1}{m}-r]} \mu_1(B(x, r)) d\mu_2(x) \\
&= \sum_{l=0}^{m-1} \int_{[\frac{l}{m}+r, \frac{l+1}{m}-r]} 2r\rho_1(x) d\mu_2(x) \\
&\quad + \sum_{l=0}^{m-1} \int_{[\frac{l}{m}+r, \frac{l+1}{m}-r]} \int_{x-r}^{x+r} \rho_1(y) - \rho_1(x) dy d\mu_2(x) \\
&\leq 2r \int \rho_1(x)\rho_2(x) dx + \sum_{l=0}^{m-1} \int_{[\frac{l}{m}+r, \frac{l+1}{m}-r]} 2r \cdot \text{Var}_{\frac{l}{m}}^{\frac{l+1}{m}}(\rho_1) d\mu_2(x) \\
&\leq 2r \int \rho_1(x)\rho_2(x) dx + \frac{2r}{m} \|\rho_2\|_\infty \sum_{l=0}^{m-1} \text{Var}_{\frac{l}{m}}^{\frac{l+1}{m}}(\rho_1) \\
&= 2r \left(\int \rho_1(x)\rho_2(x) dx + \frac{1}{m} \|\rho_2\|_\infty \text{Var}_0^1(\rho_1) \right).
\end{aligned}$$

In the final line we have used that the total variation is additive on closed intervals. Combining the estimates above we have

$$\begin{aligned}
& \int \mu_1(B(x, r)) d\mu_2(x) \\
&\leq 2r \left(\int \rho_1(x)\rho_2(x) dx + \frac{1}{m} \|\rho_2\|_\infty \text{Var}_0^1(\rho_1) + rm \|\rho_1\|_\infty \|\rho_2\|_\infty \right).
\end{aligned}$$

We see that all of the terms in the large bracket apart from $\int \rho_1\rho_2 dx$ can either be made arbitrarily small by choosing m sufficiently large, or tend to zero as $r \rightarrow 0$. This implies

$$\limsup_{r \rightarrow 0} \frac{\int \mu_1(B(x, r)) d\mu_2(x)}{2r \int (\rho_1\rho_2)(x) dx} \leq 1.$$

The corresponding lower bound for the liminf follows from a similar argument. \square

We are now in a position to prove the corollaries.

Proof of Corollary 1.5. Since our maps are Gibbs-Markov or the Gauss map, the acips μ_1 and μ_2 have densities ρ_1 and ρ_2 in BV. Moreover, ρ_1 and ρ_2 are both bounded away from 0 and bounded above. In particular this means that $C_{\mu_1, \mu_2} = 1$ and μ_1 and μ_2 have continuous mean scaling. Furthermore, (X, T_i, μ_i) have exponential mixing for BV against L^∞ for $i = 1, 2$, see for example [Ry] for the Gibbs-Markov case and [BDT, Section 2.6] for that case, done in more details, and for the Gauss map case. Therefore this result follows from Theorem A and Lemma 5.1. \square

Proof of Corollary 1.6. This proof is almost the same as that for Corollary 1.5. The correlation dimension exists and μ has continuous mean scaling by the arguments above. By [Z, Lemma 4.16] we satisfy the required exponential 4-mixing on intervals assumption. The Frostman dimension is 1 since the density is bounded above. Moreover, by [HNT, Lemma 3.4] the early return property holds with exponent $D_\mu = 1$. Our result now follows from Theorem B and Lemma 5.1. \square

5.2. The continuous mean scaling property. Since our applications require the continuous mean scaling property, which is not always easy to check in non-acip settings, we give an alternative criterion for the single measure case in the following lemma.

Lemma 5.2. *Suppose that μ has the property that there exists $r_0 > 0$, $D > 1$ and $C > 1$ such that for any $x \in \text{supp}(\mu)$ and $r \in (0, r_0)$ we have*

$$\mu(B(x, Dr)) \geq C\mu(B(x, r)).$$

Then μ has the continuous mean scaling property.

Note that we can iterate the relation above: so long as $D^{n-1}r < r_0$,

$$\mu(B(x, D^n r)) \geq C^n \mu(B(x, r)). \quad (5.1)$$

As we will see, this condition is satisfied by many dynamically defined measures, but the fact that it holds for acips with density bounded above and below and Ahlfors regular measures is immediate.

Proof. Fix μ satisfying our assumptions and $s, \epsilon, \beta > 0$. Then for any $\delta > 0$ we have

$$\begin{aligned} 0 &\leq \frac{\int \mu(B(x, (s + \delta)/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} - 1 \\ &= \frac{\int \mu(B(x, (s + \delta)/n^\beta) \setminus B(x, s/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} \\ &\leq \frac{\int \mu(B(x + s/n^\beta, \delta/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} \\ &\quad + \frac{\int \mu(B(x - s/n^\beta, \delta/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)}. \end{aligned}$$

We will now show that we can choose $\delta_0 > 0$ such that when $0 < \delta \leq \delta_0$ then

$$\frac{\int \mu(B(x + s/n^\beta, \delta/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} < \epsilon/2$$

for all $n \in \mathbb{N}$. The argument we present can easily be adapted to find $\delta_0 > 0$ for which

$$\frac{\int \mu(B(x - s/n^\beta, \delta/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} < \epsilon/2$$

for all $0 < \delta \leq \delta_0$ and $n \in \mathbb{N}$, and to show that for any $s, \beta, \epsilon > 0$ there exists $\delta_0 > 0$ such that

$$\left| \frac{\int \mu(B(x, (s - \delta)/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} - 1 \right| < \epsilon$$

for all $0 < \delta \leq \delta_0$ for all $n \in \mathbb{N}$. Thus μ will have continuous mean scaling.

Given $\delta > 0$ and $n \in \mathbb{N}$ let $L \in \mathbb{N}$ be such that

$$2^{-L-1} < \frac{\delta}{n^\beta} \leq 2^{-L}$$

and $w \in \mathbb{N}_0$ be such that

$$\frac{w}{2^L} \leq \frac{s}{n^\beta} \leq \frac{w+1}{2^L}.$$

Letting $I_j = \left[\frac{j}{2^L}, \frac{j+1}{2^L} \right)$ for $0 \leq j \leq 2^L - 1$, we have

$$\begin{aligned} \int \mu(B(x + s/n^\beta, \delta/n^\beta)) d\mu(x) &\leq \sum_{j=0}^{2^L-1} \mu(I_j) \sum_{p=-3}^3 \mu(I_{j+w+p}) \\ &\leq \sum_{p=-3}^3 \sum_{j=0}^{2^L-1} \mu(I_j) \mu(I_{j+w+p}) \\ &\leq \sum_{p=-3}^3 \left(\sum_{j=0}^{2^L-1} \mu(I_j)^2 \right)^{1/2} \left(\sum_{j=0}^{2^L-1} \mu(I_{j+w+p})^2 \right)^{1/2} \\ &\leq 7 \sum_{j=0}^{2^L-1} \mu(I_j)^2. \end{aligned}$$

Where we have used the Cauchy-Schwarz inequality in the penultimate line. For any $P > 0$ and interval I we let $P \cdot I$ be the interval with the same centre as I expanded by a factor of P . Since s is fixed, for any $P > 1$ we can choose $\delta_0 > 0$ to be sufficiently small so that for any $0 < \delta \leq \delta_0$ and $n \in \mathbb{N}$ we have

$$\int \mu(B(x, s/n^\beta)) d\mu(x) \geq \sum_{j=0}^{2^L-1} \mu(I_j) \mu(P \cdot I_j).$$

Combining the estimates above, we have shown that for any $P > 1$ we can choose $\delta_0 > 0$ so that for any $0 < \delta \leq \delta_0$ and $n \in \mathbb{N}$ we have

$$\frac{\int \mu(B(x + s/n^\beta, \delta/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} \leq \frac{7 \sum_{j=0}^{2^L-1} \mu(I_j)^2}{\sum_{j=0}^{2^L-1} \mu(I_j) \mu(P \cdot I_j)}. \quad (5.2)$$

It follows now from (5.1) that for any $\epsilon > 0$, we can choose $\delta_0 > 0$ and $P > 1$ so that (5.2) holds for all $0 < \delta \leq \delta_0$ and $n \in \mathbb{N}$, and for any interval I of

at length at most $2\delta_0$ satisfying $\mu(I) > 0$, we have $\mu(I) < \frac{\epsilon}{2} \cdot \mu(P \cdot I)$. Using this final inequality in (5.2) implies

$$\frac{\int \mu(B(x + s/n^\beta, \delta/n^\beta)) d\mu(x)}{\int \mu(B(x, s/n^\beta)) d\mu(x)} < \epsilon/2$$

for all $0 < \delta \leq \delta_0$ for all $n \in \mathbb{N}$. Which was what we wanted to show. \square

5.3. Cookie cutters: the general case. Here we outline how we can apply our results to cookie cutters and their associated equilibrium states.

Recall Definition 1.4 and the discussion on cookie cutters after Corollary 1.5. We begin by remarking that there is some minimal gap $G > 0$ such that $\text{dist}(P, P') \geq G$ for distinct $P, P' \in \mathcal{P}$. Now let \mathcal{P}_n denote the set of n -cylinders, i.e., maximal intervals P such that for each $i = 0, \dots, n-1$, $\tilde{T}^i P \subset P_i$ for some $P_i \in \mathcal{P}$. The bounded distortion property, along with the regularity and expansion properties, implies that $\frac{D\tilde{T}^n x}{D\tilde{T}^n y}$ for $x, y \in P \in \mathcal{P}_n$ is uniformly bounded from above and below. Hence there is $G' > 0$ such that if $P, P' \in \mathcal{P}_n$ are distinct and $P \subset P'' \in \mathcal{P}_{n-1}$, then $\text{dist}(P, P') \geq G'|P''|$. Note that by the bounded distortion property there also exists $b_T > 0$ such that $|P| \geq b_T|P''|$.

We next briefly explain the thermodynamic formalism associated to cookie cutters, see for example [PU, Chapter 5] for proofs of the claims made below.

We note that for $\phi : \tilde{X} \rightarrow \mathbb{R}$ Hölder continuous there is a unique equilibrium state μ_ϕ , that is μ_ϕ is the unique T -invariant measure satisfying

$$h(\mu_\phi) + \int \phi d\mu_\phi = P(\phi) := \sup \left\{ h(\nu) + \int \phi d\nu : \nu \in \mathcal{M} \right\}$$

where \mathcal{M} is the space of T -invariant probability measures. Moreover, there is a Hölder continuous ρ (bounded away from zero and infinity) such that for $\psi = \phi + \log \rho - \log \rho \circ T - P(\phi) \leq -c_\phi < 0$, for some $c_\phi > 0$, μ_ϕ is also an equilibrium state for ψ , but it is furthermore ψ -conformal: this means that if $T : A \rightarrow T(A)$ is bijective on a measurable set A , then

$$\mu_\phi(T(A)) = \int_A e^{-\psi} d\mu_\phi.$$

Iterating this we obtain that if $T^n : A \rightarrow T^n(A)$ is bijective, then

$$\mu_\phi(T^n(A)) = \int_A e^{-S_n \psi} d\mu_\phi$$

where $S_n \psi(x) := \sum_{k=0}^{n-1} \psi(T^k x)$.

Noting that the above means that if ϕ is α -Hölder then so is ψ , we will immediately obtain some additional regularity for $S_n \psi$ which is provided by the following lemma.

Lemma 5.3. *If ψ is α -Hölder then there exists $K \geq 1$ such that for any $n \in \mathbb{N}$, for $x, y \in P \in \mathcal{P}_n$,*

$$|S_n\psi(x) - S_n\psi(y)| \leq K|\tilde{T}^n x - \tilde{T}^n y|^\alpha.$$

Proof. Suppose $|\psi(x) - \psi(y)| \leq \tilde{K}|x - y|^\alpha$. Then recalling that $|D\tilde{T}| \geq \lambda > 1$,

$$\begin{aligned} |S_n\psi(x) - S_n\psi(y)| &\leq \tilde{K} \sum_{k=0}^{n-1} |\tilde{T}^k(x) - \tilde{T}^k(y)|^\alpha \leq \tilde{K} |\tilde{T}^n x - \tilde{T}^n y|^\alpha \sum_{k=0}^{n-1} \frac{1}{\lambda^{\alpha k}} \\ &\leq K |\tilde{T}^n x - \tilde{T}^n y| \end{aligned}$$

for $K = \tilde{K}/(1 - \lambda^{-\alpha})$. \square

We will now prove that the assumptions appearing in our main theorems hold for equilibrium states associated to Hölder continuous potentials. First note that by [Ry] and [Z, Lemma 4.16], the mixing properties required in Theorems A, B and 2.1 hold. As in [PW], C_μ exists for equilibrium states associated to cookie cutters, see also [PU, Chapter 9]. Note that $D_q(\mu)$, the q -Renyi dimension was first rigorously shown to exist for equilibrium states associated to cookie cutters in [Ra], and the proof that $D_2(\mu) = C_\mu$ was given in the other references.

Lemma 5.4. *Let μ be an equilibrium state with Frostman exponent F_μ . For any $\tilde{F}_\mu < F_\mu$, there exists $C > 0$ such that $\mu(A_r(n)) \leq Cr^{\tilde{F}_\mu}$ for all $r > 0$ and $n \in \mathbb{N}$. Thus the early return exponent satisfies $D_\mu \geq F_\mu$.*

The proof is adapted from [HNT, Lemma 3.4].

Proof. Fix $\tilde{F}_\mu < F_\mu$ and let $\tilde{C} > 0$ be such that

$$\mu(B(x, r)) \leq \tilde{C}r^{\tilde{F}_\mu} \tag{5.3}$$

for all $x \in [0, 1]$ and $r > 0$. Let $r > 0$ and $n \in \mathbb{N}$. We consider $A_r(n) \cap P$ for some $P \in \mathcal{P}_n$. For simplicity assume $\tilde{T}^n|_P$ is orientation preserving, otherwise the proof is similar. Since \tilde{T}^n is expanding on P the map $x \rightarrow \tilde{T}^n x - x$ is strictly increasing on P . Thus, if a solution exists there exists a unique $x_P^+ \in P$ such that $\tilde{T}^n x_P^+ = x_P^+ + r$, and similarly, if a solution exists there exists a unique $x_P^- \in P$ such that $\tilde{T}^n x_P^- = x_P^- - r$. If no solution to $\tilde{T}^n x = x + r$ exists then we define $x_P^+ \in P$ to be such that $\tilde{T}^n x_P^+ - x_P^+ = \max_{x \in P} \tilde{T}^n x - x$. If no solution to $\tilde{T}^n x = x - r$ exists then we define $x_P^- \in P$ to be such that $\tilde{T}^n x_P^- - x_P^- = \min_{x \in P} \tilde{T}^n x - x$. Since there always exists $x \in P$ satisfying $\tilde{T}^n x - x = 0$, it follows that we always have $\tilde{T}^n x_P^+ - x_P^+ \leq r$ and $\tilde{T}^n x_P^- - x_P^- \geq -r$. Thus in all cases we have $\tilde{T}^n(x_P^-, x_P^+) \subseteq (x_P^- - r, x_P^+ + r)$. Then due to the monotonicity of $x \rightarrow \tilde{T}^n x - x$ on P , if $x \in P$ satisfies $x < x_P^-$ then $x - \tilde{T}^n(x) > r$, and if $x > x_P^+$ then $\tilde{T}^n(x) - x > r$.

Hence $A_r(n) \cap P = (x_P^-, x_P^+)$. So we wish to estimate $\mu(x_P^-, x_P^+)$. Firstly we estimate the length of $\tilde{T}^n(x_P^-, x_P^+)$. Using that $x \mapsto \tilde{T}^n(x) - x$ is expanding by at least λ^n and the inclusion $\tilde{T}^n(x_P^-, x_P^+) \subseteq (x_P^- - r, x_P^+ + r)$, it can be shown that $|(x_P^-, x_P^+)| \leq 2r/(\lambda^n - 1)$. Using this inequality and the inclusion $\tilde{T}^n(x_P^-, x_P^+) \subset (x_P^- - r, x_P^+ + r)$ again, we have the estimate

$$\begin{aligned} |\tilde{T}^n(x_P^-, x_P^+)| &\leq |(x_P^- - r, x_P^+ + r)| \\ &= |(x_P^- - r, x_P^-)| + |(x^-, x^+)| + |(x^+, x^+ + r)| \\ &\leq 2r + \frac{2r}{\lambda^n - 1} \leq \frac{2\lambda r}{\lambda - 1}. \end{aligned}$$

Let $x_P \in P$ be the unique point such that $\tilde{T}^n(x_P) = x_P$. It is clear that $x_P \in (x_P^-, x_P^+)$. Using this fact, the bound for $|\tilde{T}^n(x_P^-, x_P^+)|$ above, and Lemma 5.3, the following holds

$$|S_n \psi(x_P) - S_n \psi(x)| \leq K \left(\frac{2\lambda r}{\lambda - 1} \right)^\alpha$$

whenever $x \in (x_P^-, x_P^+)$. So by ψ -conformality,

$$\begin{aligned} \mu(x_P^- - r, x_P^+ + r) &\geq \mu\left(\tilde{T}^n(x_P^-, x_P^+)\right) = \int_{(x_P^-, x_P^+)} e^{-S_n \psi} d\mu \\ &\geq e^{-K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} e^{-S_n \psi(x_P)} \mu(x_P^-, x_P^+), \end{aligned}$$

Manipulating the above and using (5.3) implies

$$\begin{aligned} \mu(x_P^-, x_P^+) &\leq e^{K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} e^{S_n \psi(x_P)} \left(\mu(x_P^-, x_P^+) + \mu(x_P^- - r, x_P^-) + \mu(x_P^+, x_P^+ + r) \right) \\ &\leq e^{K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} e^{S_n \psi(x_P)} \left(\mu(x_P^-, x_P^+) + 2\tilde{C}r^{\tilde{F}_\mu} \right). \end{aligned} \quad (5.4)$$

Since we know that $\psi \leq -c_\phi < 0$, there exists $\eta > 0$ such that for r small enough

$$e^{K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} e^{S_n \psi(x_P)} \leq e^{K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} e^{-c_\phi} < (1 - \eta). \quad (5.5)$$

Combining (5.4), (5.5) and the trivial bound $e^{K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} \leq e^K$ which holds for all r sufficiently small, we have the bound

$$\mu(x_P^-, x_P^+) \leq \frac{2\tilde{C}e^K e^{S_n \psi(x_P)} r^{\tilde{F}_\mu}}{1 - e^{K\left(\frac{2\lambda r}{\lambda - 1}\right)^\alpha} e^{S_n \psi(x_P)}} \leq \frac{2\tilde{C}e^K}{\eta} e^{S_n \psi(x_P)} r^{\tilde{F}_\mu} \quad (5.6)$$

whenever r is sufficiently small. Since ψ -conformality also implies that there exists $y_P \in P$ such that $e^{S_n \psi(y_P)} = \mu(P)$, and this y_P satisfies $|S_n \psi(x_P) - S_n \psi(y_P)| \leq K$ by Lemma 5.3, (5.6) implies

$$\mu(x_P^-, x_P^+) \leq \frac{2\tilde{C}e^{2K}\mu(P)}{\eta} r^{\tilde{F}_\mu}.$$

Using this expression and summing over $P \in \mathcal{P}_n$ now yields

$$\mu(A_r(n)) = \sum_{P \in \mathcal{P}_n} \mu(x_P^-, x_P^+) \leq \left(\frac{2\tilde{C}e^{2K}}{\eta} \right) r^{\tilde{F}_\mu}$$

for all r sufficiently small. Thus our result holds for r sufficiently small. This implies our result for an arbitrary choice of $r > 0$ after choosing a potentially larger constant. \square

Lemma 5.5. *μ satisfies the continuous mean scaling property.*

Proof. We begin by proving a claim on how cylinders scale.

Claim. There exists $\Phi > 1$ such that if $P \in \mathcal{P}_n$ and $P' \in \mathcal{P}_{n-1}$ with $P \subset P'$, then $\mu(P') \geq \Phi\mu(P)$.

Proof. There must be some $Q \in \mathcal{P}_n$ where $Q \subset P'$ and $Q \neq P$. By conformality,

$$\frac{1}{\mu(\tilde{T}^{n-1}Q)} = \frac{\mu(\tilde{T}^{n-1}P')}{\mu(\tilde{T}^{n-1}Q)} = \frac{\int_{P'} e^{-S_{n-1}\psi} d\mu}{\int_Q e^{-S_{n-1}\psi} d\mu} = \frac{e^{-S_{n-1}\psi(x)}\mu(P')}{e^{-S_{n-1}\psi(y)}\mu(Q)}$$

for some $x \in P'$ and $y \in Q$. By Lemma 5.3 we know that $|S_{n-1}\psi(x) - S_{n-1}\psi(y)| \leq K$. This implies

$$\mu(Q) \geq e^{-K} \mu(\tilde{T}^{n-1}Q) \mu(P') \geq e^{-K} M \mu(P')$$

for $M := \inf_{\hat{Q} \in \mathcal{P}_1} \mu(\hat{Q}) \in (0, 1)$, where we are using the fact that $\tilde{T}^{n-1}Q \in \mathcal{P}_1$. So as

$$\mu(P') \geq \mu(Q) + \mu(P) \geq e^{-K} M \mu(P') + \mu(P),$$

setting $\Phi := \frac{1}{1 - e^{-K}M}$ completes the claim. \square

Let $x \in \text{supp}(\mu)$ and $r > 0$. Suppose that n is maximal such that $B(x, r) \cap X$ is contained in some $P \in \mathcal{P}_n$. Here we also assume that r is small enough that $n \geq 2$. So $\mu(B(x, r)) \leq \mu(P)$. By construction, $r \geq G'|P|$, since otherwise $B(x, r) \cap X$ is contained in some $(n+1)$ -cylinder. Then for $P' \in \mathcal{P}_{n-1}$ such that $P \subset P'$, defining $D := 1/(b_T G')$ we have $Dr \geq |P'|$ so $B(x, Dr)$ contains P' . Hence by the claim above

$$\mu(B(x, Dr)) \geq \mu(P') \geq \Phi\mu(P) \geq \Phi\mu(B(x, r)).$$

Combining the above with Lemma 5.2 we see that μ satisfies the continuous mean scaling property. \square

The following gives a very rough Frostman bound, which seems only likely to be optimal in special cases, but it is sufficient for our purposes here. We first define

$$\lambda_{\max} := \sup_{x \in \tilde{X}} |D\tilde{T}(x)| \geq \lambda > 1$$

and recall Φ in the proof of the lemma above.

Lemma 5.6. *Let Φ be as in the claim in Lemma 5.5. The Frostman dimension of μ satisfies $F_\mu \geq \frac{\log \Phi}{\log \lambda_{\max}}$.*

Proof. It is sufficient to prove that $\mu(P) \leq |P|^{\frac{\log \Phi}{\log \lambda_{\max}}}$ where P is a cylinder, since a suitable bound then passes to any interval via an approximation argument. If $P \in \mathcal{P}_n$ then $\mu(P) \leq \Phi^{-n}$ by repeatedly applying the claim in Lemma 5.5, and $|P| \geq \lambda_{\max}^{-n}$ by an application of the mean value theorem and the chain rule. Hence $\mu(P) \leq |P|^{\frac{\log \Phi}{\log \lambda_{\max}}}$. \square

Hence we can apply Theorem B to these systems. Note that by Lemma 5.4, the condition $\beta(C_\mu - D_\mu) < 1$ is implied by $\beta(C_\mu - F_\mu) < 1$.

We have also shown that Theorem A holds when our systems are both the same. In the other case, we have not proved the existence of C_{μ_1, μ_2} , see [AW, Section 5] to see the types of issues two different measures can cause. Nevertheless, if we have some estimates in these cases, we may apply Theorem 2.1.

5.4. Cookie cutters: geometric potentials. A particularly natural class of equilibrium state for a cookie cutter corresponds to the case when the potential is given by $\phi(x) = \phi_\delta = -\delta \log |D\tilde{T}(x)|$ for $\delta \in \mathbb{R}$ and $x \in \tilde{X}$. Note that this map is always Hölder. Bowen's formula, see [Bow] implies that there is a unique $h > 0$ such that $P(\phi_h) = 0$ and moreover the Hausdorff dimension of X equals h . Setting $\mu = \mu_{\phi_h}$ we see that as in Section 5.3 μ is ψ -conformal where $\psi(x) = -h \log |D\tilde{T}(x)| + \log \rho(x) - \log \rho \circ \tilde{T}(x)$ where ρ is the Hölder density function for μ . So for example, if $T^n : A \rightarrow T^n(A)$ is bijective, then

$$\mu(T^n(A)) = \int_A |D\tilde{T}^n|^h \cdot \frac{\rho \circ \tilde{T}^n}{\rho} d\mu.$$

It can therefore be shown that in this case μ is h -Ahlfors regular. Hence C_μ exists and equals h . Moreover, by h -Ahlfors regularity and Lemma 5.4 we must have $F_\mu = h$ and $D_\mu \geq h$. Consequently the only constraint in applying Theorems A and B to this system is that $\beta \in (0, 2/h)$. This is the optimal parameter space for β .

5.5. Cookie cutters: an explicit example. Here we will give a simple example of a linear cookie cutter with a Bernoulli measure exhibiting multifractal behaviour, where we can compute our quantities explicitly.

Let $I_1 = [0, 1/4]$ and $I_2 = [3/4, 1]$ and $\tilde{T}(x) = 4x$ for $x \in I_1$ and $\tilde{T}(x) = 4x - 3$ for $x \in I_2$. Then \tilde{T} is a linear cookie cutter. Let μ be the corresponding $(\alpha, 1 - \alpha)$ -Bernoulli measure for $\alpha \in (0, 1/2]$. We first note that by standard arguments, for example an analogue of [GRS, Lemma 13] or an adaptation

of the conditioning argument in Section 3.1, we have

$$\underline{C}_\mu = \liminf_{n \rightarrow \infty} \frac{\log \sum_{Q \in \mathcal{Q}_n} \mu(Q)^2}{-n \log 4},$$

where \mathcal{Q}_n is the set of intervals of the form $[\frac{k}{4^n}, \frac{k+1}{4^n}]$ for some $k \in \mathbb{N}$. We have a similar expression for \overline{C}_μ involving the limsup. Using these expressions we can show that in this case C_μ exists and that it equals $-\log(\alpha^2 + (1-\alpha)^2)/\log 4$. For the Frostman exponent we can compute $F_\mu = -\log(1-\alpha)/\log 4$.

Suppose now that $\alpha = 1/3$. Then $C_\mu = \frac{\log(\frac{9}{5})}{\log 4}$ and $F_\mu = \frac{\log \frac{3}{2}}{\log 4}$. Then by Lemma 5.4, $C_\mu - D_\mu \leq C_\mu - F_\mu = \frac{\log \frac{18}{15}}{\log 4} \approx 0.132$. So the conditions $\beta(C_\mu - D_\mu) < 1$ and $\beta(C_\mu - F_\mu) < 1$ become $\beta < 7.604\dots$, and $\beta C_\mu < 2$ becomes $\beta < \frac{2 \log 4}{\log \frac{9}{5}} \approx 3.538$. Hence the conclusion of Theorem B holds in the optimal range of parameters $\beta \in (0, \frac{2}{C_\mu})$. We note that Theorem A also applies for $\tilde{T}_1 = \tilde{T}_2 = \tilde{T}$ and $\mu_1 = \mu_2 = \mu$, and Theorem 2.1 may also apply if μ_1 and μ_2 have different α corresponding to them.

5.6. A counterexample. Consider the version of the Manneville-Pomeau map given by

$$T(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & \text{if } x \in [0, 1/2), \\ 2x - 1 & \text{if } x \in [1/2, 1]. \end{cases}$$

with parameter $\alpha \in (0, 1)$. Then let μ be the associated acip. It is shown in [RT, Proposition 3.2] that $C_\mu = 2(1-\alpha)$ when $\alpha \in (1/2, 1)$: indeed there exists $K > 0$ such that for all $r \in (0, 1)$ we have

$$\frac{r^{2(1-\alpha)}}{K} \leq \int \mu(B(z, r)) \, d\mu(z) \leq Kr^{2(1-\alpha)}. \quad (5.7)$$

The continuous mean scaling property can be easily checked here, and it can be shown that the Frostman dimension satisfies $F_\mu = 1 - \alpha$, which would not cause problems in terms of the statement of Theorem B for $\beta \in (0, 2]$; however, as in [S, Corollary 1], the mixing rate here is polynomial, so the conditions for our main results do not hold. We will consider the particular case of $\beta = 1$ and show that the conclusion of Theorem B does not hold.

As in, for example, [S, Corollary 1], let $y_0 = 1$, and for $n \geq 1$ let $y_n \in (0, y_{n-1})$ be such that $T(y_n) = y_{n-1}$. As proved there, $y_n \sim c/n^{1/\alpha}$. Let $k_n := \lfloor (n \log n)^\alpha \rfloor$ so that $y_{k_n} \sim \frac{c}{n \log n}$. Let $\gamma \in (0, \alpha)$ and observe that

$$k_n - n^\alpha (\log n)^\gamma = k_n \left(1 - \frac{n^\alpha (\log n)^\gamma}{\lfloor (n \log n)^\alpha \rfloor} \right) > \frac{k_n}{2}$$

for n sufficiently large. This inequality and the fact $y_n \sim c/n^{1/\alpha}$ mean that for any $s > 0$, for n sufficiently large, if $T^n x \in (0, y_{k_n})$ then $T^{n+i} x \in$

$(0, y_{k_n-i}) \in (0, s/2n)$ for all $0 \leq i \leq n^\alpha(\log n)^\gamma$. Therefore for all $s > 0$, for n sufficiently large, if $T^n x \in (0, y_{k_n})$ then

$$\# \left\{ 0 \leq i \neq j \leq n + n^\alpha(\log n)^\gamma : |T^i x - T^j x| < \frac{s}{2n} \right\} \gtrsim n^{2\alpha}(\log n)^{2\gamma}$$

Using this counting bound and (5.7), we see that for any $s > 0$, for n sufficiently large, if $T^n x \in (0, y_{k_n})$ then

$$\begin{aligned} & \frac{\# \left\{ 0 \leq i \neq j < 2n : |T^i x - T^j x| \leq \frac{s}{2n} \right\}}{(2n)^2 \int \mu \left(B \left(z, \frac{s}{2n} \right) \right) d\mu(z)} \\ & \asymp \frac{1}{n^{2\alpha}} \# \left\{ 0 \leq i \neq j < 2n : |T^i x - T^j x| \leq \frac{s}{2n} \right\} \gtrsim (\log n)^{2\gamma}. \end{aligned} \tag{5.8}$$

Since the Lebesgue measure of $(0, y_{k_n})$ is at least a constant multiple of $\frac{1}{n \log n}$ for all $n \in \mathbb{N}$, hence not summable, by [G, Theorem 1.1] for μ almost every x , there are infinitely many n such that $T^n x \in (0, y_{k_n})$. Hence for any $s > 0$ there exists infinitely many n so that (5.8) is satisfied. Hence for $\beta = 1$, for μ almost every x the conclusion of Theorem B does not hold for any $s > 0$.

REFERENCES

- [AD] J. Aaronson, M. Denker, *Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps*, Stoch. Dyn. **1** (2001) 193–237. 2
- [AB] C. Aistleitner, S. Baker, *On the pair correlations of powers of real numbers*, Israel J. Math. **242** (2021) no. 1, 243–268. 3
- [ALP] C. Aistleitner, T. Lachmann, F. Pausinger, *Pair correlations and equidistribution*, J. Number Theory **182** (2018) 206–220. 2
- [ALL] C. Aistleitner, G. Larcher, M. Lewko, *Additive energy and the Hausdorff dimension of the exceptional set in metric pair correlation problems*, Israel J. Math. **222** (2017) 463–485, With an appendix by Jean Bourgain. 3
- [ALT] C. Aistleitner, G. Larcher, N. Technau, *There is no Khintchine threshold for metric pair correlations*, Mathematika **65** (2019) 929–949. 3
- [ABB] D. Allen, S. Baker, B. Bárány, *Recurrence rates for shifts of finite type*, Adv. Math. **460** (2025), Paper No. 110039, 36 pp. 4
- [AW] R. Arratia, M.S. Waterman, *Critical phenomena in sequence matching*, Ann. Probab. **13** (1985) 1236–1249. 5, 47
- [BF] S. Baker, M. Farmer, *Quantitative recurrence properties for self-conformal sets*, Proc. Amer. Math. Soc. **149** (2021) 1127–1138. 4
- [BK] S. Baker, H. Koivusalo, *Quantitative recurrence and the shrinking target problem for overlapping iterated function systems*, Adv. Math. **442** (2024), Paper No. 109538, 65 pp. 4
- [BLR] V. Barros, L. Liao, J. Rousseau, *On the shortest distance between orbits and the longest common substring problem*, Adv. Math. **344** (2019) 311–339. 4, 5, 14
- [BCGW] T.F. Bloom, S. Chow, A. Gafni, A. Walker, *Additive energy and the metric Poissonian property*, Mathematika **64** (2018) 679–700. 3
- [B] M.D. Boshernitzan, *Quantitative recurrence results*, Invent. Math. **113** (1993) 617–631. 4

- [Bow] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Springer Lect. Notes in Math. 470 (1975). 47
- [BDT] H. Bruin, M. Demers, M. Todd, *Hitting and escaping statistics: mixing, targets and holes*, Adv. Math. **328** (2018) 1263–1298. 8, 40
- [B] Y. Bugeaud, *Distribution modulo one and Diophantine approximation*, Cambridge Tracts in Mathematics, 193, Cambridge Univ. Press, Cambridge, 2012. 1
- [EBMV] D. El-Baz, J. Marklof, I. Vinogradov, *The two-point correlation function of the fractional parts of \sqrt{n} is Poisson*, Proc. Amer. Math. Soc. **143** (2015), no. 7, 2815–2828. 3
- [EM] N. Elkies, C. McMullen, *Gaps in $\sqrt{n} \bmod 1$ and ergodic theory*, Duke Math. J. **123** (2004) 95–139. 3
- [FFT1] A.C.M. Freitas, J.M. Freitas, M. Todd, *Hitting time statistics and extreme value theory*, Probab. Theory Related Fields **147** (2010) 675–710. 4
- [FFT2] A.C.M. Freitas, J.M. Freitas, M. Todd, *Enriched functional limit theorems for dynamical systems*, Online at Ann. Sc. Norm. Super. Pisa Cl. Sci. (5). 2
- [G] S. Gouëzel, *A Borel-Cantelli lemma for intermittent interval maps*, Nonlinearity **20** (2007) 1491–1497. 49
- [GRS] S. Gouëzel, J. Rousseau, M. Stadlbauer, *Minimal distance between random orbits*, Probab. Theory Related Fields **189** (2024) 811–847. 5, 47
- [GL] S. Grepstad, G. Larcher, *On pair correlation and discrepancy*, Arch. Math. (Basel) **109** (2017), no. 2, 143–149. 2
- [HKKP] M. Holland, M. Kirsebom, P. Kunde, T. Persson, *Dichotomy results for eventually always hitting time statistics and almost sure growth of extremes*, Trans. Amer. Math. Soc. **377** (2024) 3927–3982. 4
- [HNT] M. Holland, M. Nicol, A. Török, *Extreme value theory for non-uniformly expanding dynamical systems*, Trans. Amer. Math. Soc. **364** (2012) 661–688. 41, 44
- [HT] M. Holland, M. Todd, *On distributional limit laws for recurrence*, Nonlinearity **38** (2025) Paper No. 075028. 4
- [KMS] M. Kesseböhmer, S. Munday, B.O. Stratmann, *Strong renewal theorems and Lyapunov spectra for α -Farey and α -Lüroth systems*, Ergodic Theory Dynam. Systems **32** (2012) 989–1017. 8
- [KKP] M. Kirsebom, P. Kunde, T. Persson, *On shrinking targets and self-returning points*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **24** (2023) 1499–1535. 4
- [KKPT] M. Kirsebom, P. Kunde, T. Persson, M. Todd, *Almost sure orbits closeness*, Preprint (arXiv:2510.13277). 5, 11
- [KP] L. Kuipers, H. Niederreiter, *Uniform distribution of sequences*, Pure and Applied Mathematics, Wiley-Intersci., New York-London-Sydney, 1974. 1
- [LLSV] J. Levesley, B. Li, D. Simmons, S. Velani, *Shrinking targets versus recurrence: the quantitative theory*, Mathematika **71** (2025) Paper No. e70039, 16 pp. 4
- [LST] C. Lutsko, A. Sourmelidis, N. Technau, *Pair correlation of the fractional parts of αn^θ* , J. Eur. Math. Soc. (JEMS) **27** (2025) 4069–4082. 3
- [M] J. Marklof, *Pair correlation and equidistribution on manifolds*, Monatsh. Math. **191** (2020) 279–294. 2
- [MN] I. Melbourne, M. Nicol, *Large deviations for nonuniformly hyperbolic systems*, Trans. Amer. Math. Soc. **360** (2008) 6661–6676. 2
- [MZ] I. Melbourne, R. Zweimüller, *Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015) 545–556. 2
- [NP] R. Nair, M. Pollicott, *Pair correlations of sequences in higher dimensions*, Israel J. Math. **157** (2007) 219–238. 4
- [PW] Y. Pesin, H. Weiss, *A multifractal analysis of equilibrium measures for conformal expanding maps and Moran-like geometric constructions*, J. Statist. Phys. **86** (1997) 233–275. 44

- [PU] F. Przytycki, M. Urbański, *Conformal fractals: ergodic theory methods*, London Mathematical Society Lecture Note Series, vol. 371, Cambridge University Press, Cambridge, 2010. 43, 44
- [Ra] D.A. Rand, *The singularity spectrum $f(\alpha)$ for cookie-cutters*, Ergodic Theory Dynam. Systems **9** (1989) 527–541. 8, 44
- [Re] S. Regavim, *Minimal gaps and additive energy in real-valued sequences*, Q. J. Math. **74** (2023) 825–866. 5
- [RSW] G. Reinert, S. Schbath, M. Waterman, *Applied Combinatorics on Words*, volume 105 of Encyclopedia of Mathematics and its Applications, chapter Statistics on Words with Applications to Biological Sequences, Cambridge University Press, 2005. 5
- [RY] L. Rey-Bellet, L.-S. Young, *Large deviations in non-uniformly hyperbolic dynamical systems*, Ergodic Theory Dynam. Systems **28** (2008) 587–612. 2
- [RT] J. Rousseau, M. Todd, *Orbits closeness for slowly mixing dynamical systems*, Ergodic Theory Dynam. Systems **44** (2024) 1192–1208. 4, 48
- [Ru] Z. Rudnick, *A metric theory of minimal gaps*, Mathematika **64** (2018) 628–636. 5
- [RS] Z. Rudnick, P. Sarnak, *The pair correlation function of fractional parts of polynomials*, Comm. Math. Phys. **194** (1998) 61–70. 3
- [Ry] M. Rychlik, *Bounded variation and invariant measures*, Studia Math. **76** (1983) 69–80. 40, 44
- [S] O. Sarig, *Subexponential decay of correlations*, Invent. Math. **150** (2002) 629–653. 48
- [Sau] B. Saussol, *An introduction to quantitative Poincaré recurrence in dynamical systems*, Rev. Math. Phys. **21** (2009) 949–979. 4
- [W] A. Walker, *The primes are not metric Poissonian*, Mathematika **64** (2018) 230–236. 3
- [Wal] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, 79, Springer, New York-Berlin, 1982 1, 4
- [Wat] M. Waterman, *Introduction to Computational Biology: Maps, Sequences and Genomes*, Chapman and Hall, London, 1995 5
- [Y] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147** (1998) 585–650. 2
- [Z] B. Zhao, *Closest distance between iterates of typical points*, Discrete Contin. Dyn. Syst. **44** (2024) 2252–2279. 5, 7, 41, 44

MATHEMATICAL SCIENCES, LOUGHBOROUGH UNIVERSITY, LOUGHBOROUGH, LEICESTERSHIRE, LE11 3TU, UK

Email address: simonbaker412@gmail.com

URL: <https://simonbakermaths.wordpress.com/>

MATHEMATICAL INSTITUTE, UNIVERSITY OF ST ANDREWS, NORTH HAUGH, ST ANDREWS, KY16 9SS, SCOTLAND

Email address: m.todd@st-andrews.ac.uk

URL: <https://mtoddm.github.io>