

# USING BERNOULLI MAPS TO ACCELERATE MIXING OF A RANDOM WALK ON THE TORUS

GAUTAM IYER, ETHAN LU, AND JAMES NOLEN

*Dedicated to Robert L. Pego, whose life and work is an inspiration.*

ABSTRACT. We study the mixing time of a random walk on the torus, alternated with a Lebesgue measure preserving Bernoulli map. Without the Bernoulli map, the mixing time of the random walk alone is  $O(1/\varepsilon^2)$ , where  $\varepsilon$  is the step size. Our main results show that for a class of Bernoulli maps, when the random walk is alternated with the Bernoulli map  $\varphi$  the mixing time becomes  $O(|\ln \varepsilon|)$ . We also study the *dissipation time* of this process, and obtain  $O(|\ln \varepsilon|)$  upper and lower bounds with explicit constants.

## 1. Introduction

The aim of this paper is to study how certain dynamical systems can accelerate convergence of a random walk to its stationary distribution. Explicitly, let  $\mathbb{T}^d$  denote the  $d$ -dimensional torus, and consider a (discrete time) dynamical system that preserves Lebesgue measure and whose evolution is determined by a map  $\varphi: \mathbb{T}^d \rightarrow \mathbb{T}^d$ . Let  $\varepsilon > 0$  be small, and  $\varepsilon\zeta_n$  be a sequence of i.i.d.  $\mathbb{T}^d$  valued random variables with mean 0 and “size  $\varepsilon$ ” (in a sense that will be made precise shortly). We consider the Markov process on the torus defined by

$$(1.1) \quad X_{n+1} = \varphi(X_n) + \varepsilon\zeta_{n+1}.$$

Throughout this paper, we will assume that the law of the random variables  $\varepsilon\zeta_n$  has a density function  $K_\varepsilon$ . Under our assumptions, the unique, stationary distribution of the process  $X$  is the Lebesgue measure on  $\mathbb{T}^d$ , and we will denote this measure by  $\pi$ .

We will measure the accelerated mixing of  $X$  by estimating the *mixing time* and the *dissipation time* as  $\varepsilon \rightarrow 0$ . Recall the *mixing time* of  $X$ , denoted by  $t_{\text{mix}}(\delta)$ , is defined by

$$t_{\text{mix}}(\delta) = \min \left\{ n \in \mathbb{N} \mid \|\text{dist}(X_n) - \pi\|_{\text{TV}} \leq \delta \text{ for all initial distributions of } X_0 \right\}.$$

Here  $\delta \in (0, 1)$  is any fixed constant, and when  $\delta = 1/2$ , we will drop the argument and simply write  $t_{\text{mix}}$  instead of  $t_{\text{mix}}(1/2)$ .

If  $\varphi$  is the identity map, then  $X$  is simply a random walk on the torus. In this case, if the distribution of  $\varepsilon\zeta_n$  is non-degenerate and regular it is not hard to show that  $t_{\text{mix}} \approx O(1/\varepsilon^2)$  as  $\varepsilon \rightarrow 0$  (see for instance Lemma 1 in [FW03]). For more interesting choices of  $\varphi$ , however, the mixing time may be dramatically smaller.

The simplest example of this is when  $\varphi$  is the one dimensional doubling map:

$$\varphi(x) = 2x \pmod{1}.$$

In this case it is not hard to see  $t_{\text{mix}} \approx O(|\ln \varepsilon|)$ . Indeed, the linear structure of  $\varphi$  allows us to write

$$X_n = 2^n X_0 + \sum_{k=1}^n 2^{n-k} (\varepsilon\zeta_k) \pmod{1}.$$

If  $2^{n-1}\varepsilon \approx 1$ , then the first term in the above sum is roughly uniformly distributed, and can be used to show  $t_{\text{mix}} \approx |\ln_2 \varepsilon|$  as  $\varepsilon \rightarrow 0$ . A similar argument can be used to show that if  $\varphi$  is an *ergodic toral automorphism* (see for instance §4.2 in [KH95]), then  $t_{\text{mix}} = O(|\ln \varepsilon|)$  as  $\varepsilon \rightarrow 0$ . In fact, for both these examples, the distribution of  $X_n$  converges to the stationary distribution at a rate that is faster than exponential (see for instance [FW03, FI19], or Appendix A, below).

The analysis of both the above examples relies crucially on the linear structure of  $\varphi$ . Even when  $\varphi$  is piecewise affine linear, the analysis breaks down and estimating the mixing time is a much harder problem. One general result that can be deduced from [FI19] (see Corollary 1.4, below), is that when  $\varphi$  is  $C^1$  and generates an *exponentially mixing* dynamical system, and the distribution of  $\varepsilon\zeta_k$  is a periodized rescaled non-degenerate Gaussian, then

$$(1.2) \quad t_{\text{mix}} \leq O(|\ln \varepsilon|^3).$$

This bound, however, is not expected to be sharp. To briefly explain why, suppose  $X_0$  is concentrated at a point  $x \in \mathbb{T}^d$ . After one time step, the noise should ensure that some fraction of the mass of  $X_1$  becomes uniformly spread over the ball  $B(\varphi(x), \varepsilon)$ . Since  $\varphi$  generates an exponentially mixing dynamical system, we know that for  $n \geq O(|\ln \varepsilon|)$ , a constant fraction of the set  $\varphi^n(B(\varphi(x), \varepsilon))$  intersects any given  $\varepsilon$  ball. So, after one more time step, the noise should spread a fraction of the mass uniformly over this  $\varepsilon$  ball. This suggests that the optimal bounds should be  $t_{\text{mix}} = O(|\ln \varepsilon|)$ , and not  $O(|\ln \varepsilon|^3)$  as stated above. To turn the above argument into a rigorous proof of  $O(|\ln \varepsilon|)$  bounds, one needs to control the distance between  $X_n$  and  $\varphi^n(X_0)$  for  $n = O(|\ln \varepsilon|)$ . This does not always seem possible, and thus showing  $t_{\text{mix}} \leq O(|\ln \varepsilon|)$  in this generality is still open.

The main purpose of this paper is to improve (1.2) and obtain  $O(|\ln \varepsilon|)$  mixing time bounds for a class of exponentially mixing maps with non-degenerate noise. Roughly speaking, we will need to assume that  $\varphi$  is a piecewise affine linear Bernoulli map. In dimensions higher than 1 we will also require that all cylinder sets are axis-aligned cubes (see Section 2 for a precise description). A typical example (when  $d = 1$ ) is

2020 *Mathematics Subject Classification*. Primary: 60J05. Secondary: 37A25.

*Key words and phrases*. enhanced dissipation, mixing time.

This work has been partially supported by the National Science Foundation under grant DMS-2108080, and the Center for Nonlinear Analysis.

the map

$$\varphi(x) = \begin{cases} 3x & x \in [0, \frac{1}{3}), \\ \frac{3(1-x)}{2} & x \in [\frac{1}{3}, 1). \end{cases}$$

In dimension two, another example is shown graphically on the left of Figure 1. The map shown on the right of Figure 1 does not satisfy our assumptions, as some cylinder sets are rectangles, and not axis-aligned squares.

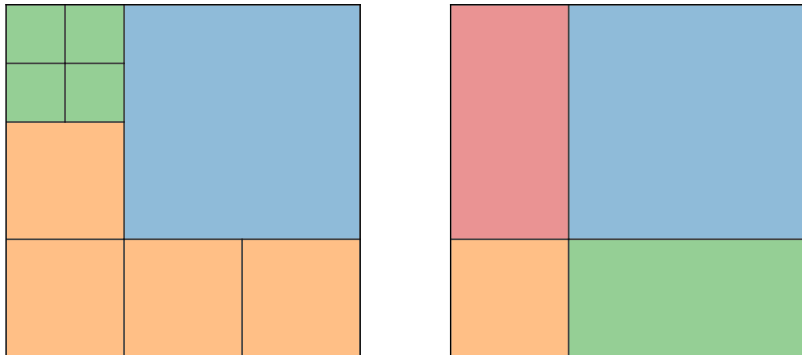


FIGURE 1. Two examples of the map  $\varphi$  in two dimensions. On each smaller region (colored by size), the function  $\varphi$  is an affine linear bijection onto the big square.

The first main result in this paper is an  $O(|\ln \varepsilon|)$  upper bound on the mixing time, under certain assumptions on the distribution of  $\varepsilon \zeta_n$ . Roughly speaking, we assume that the distribution of  $\varepsilon \zeta_n$  is either a periodized rescaled non-degenerate Gaussian (Assumption 2.2), or a rescaled separated kernel (Assumption 2.3).

**Theorem 1.1.** *Suppose either Assumption 2.2 or 2.3 (below) hold. Then there exists a constant  $C$  such that for all sufficiently small  $\varepsilon$  we have*

$$(1.3) \quad t_{\text{mix}} \leq C |\ln \varepsilon|.$$

Of course, the above implies for any  $\delta \in (0, 1)$  we have

$$t_{\text{mix}}(\delta) \leq C |\ln \varepsilon| (1 + |\ln \delta|).$$

While this bound is of the right order in  $\varepsilon$ , the constant  $C$  is not explicit. We believe the sharp bound is of the form

$$(1.4) \quad \frac{d \ln \varepsilon}{\ln p_{\max}} - C(\delta) \leq t_{\text{mix}}(\delta) \leq \frac{d \ln \varepsilon}{\ln p_{\max}} + C(\delta) \ln(1 + |\ln \varepsilon|),$$

where  $p_{\max}$  is the measure of the largest domain on which  $\varphi$  is affine linear, and  $C(\delta) < \infty$  is a finite constant that is independent of  $\varepsilon$ . Although we are presently unable to prove the upper bound in (1.4), we are able to prove the lower bound, under less restrictive assumptions than that of Theorem 1.1.

**Theorem 1.2.** *Suppose Assumption 2.1 holds, and for all sufficiently small  $\varepsilon > 0$  we have*

$$(1.5) \quad \int_{\mathbb{T}^d} d(y, 0) K_\varepsilon(y) \pi(dy) \leq \varepsilon.$$

(Here  $d(y, 0)$  denotes the distance between  $y$  and 0 on the torus  $\mathbb{T}^d$ .) Then for every  $\delta \in (0, 1)$  the mixing time of (1.1) is bounded below by

$$(1.6) \quad \frac{d \ln \varepsilon}{\ln p_{\max}} - C(1 + |\ln(1 - \delta)|) \leq t_{\text{mix}}(\delta),$$

for some constant  $C$  and all sufficiently small  $\varepsilon > 0$ .

Thus far we have been discussing the mixing time  $t_{\text{mix}}$ , which quantifies convergence to the stationary distribution in terms of total variation. On the other hand, in many physical contexts (such as passive scalar advection [Obu49, Cor51, Pie94, SS00, HV05, Thi12, MD18]),  $L^2$  convergence is more natural. The *dissipation time* (see [FW03, FNW04, FI19]) measures the rate of convergence of the distribution of  $X_n$  to  $\pi$  in  $L^2$ , for initial distributions which are also  $L^2$ . Explicitly, for any  $\delta \in (0, 1)$ , define the dissipation time  $t_{\text{dis}}(\delta)$  by

$$t_{\text{dis}}(\delta) = \min \left\{ n \in \mathbb{N} \mid \|\theta_n\|_{L^2} \leq \delta \|\theta_0\|_{L^2} \text{ for all } \theta_0 \in \dot{L}^2 \right\}.$$

Here  $\theta_n(x) = \mathbf{E}^x \theta_0(X_n) = \mathbf{E}(\theta_0(X_n) \mid X_0 = x)$ , and  $\dot{L}^2 \subseteq L^2(\mathbb{T}^d, \pi)$  is the sub-space of all mean-zero functions. For notational convenience, when  $\delta = 1/2$ , we will drop the argument and write  $t_{\text{dis}}$  instead of  $t_{\text{dis}}(1/2)$ .

Before stating our result concerning the dissipation time of (1.1), we mention that the dissipation time and mixing time are related in a general setting (see also [IZ22]).

**Proposition 1.3.** *For any Markov process, the dissipation time is bounded in terms of the mixing time by the inequality*

$$(1.7) \quad t_{\text{dis}}(\delta) \leq t_{\text{mix}}\left(\frac{\delta^2}{4}\right) \leq (2 - \log_\delta 4) t_{\text{mix}}(\delta).$$

For the Markov process (1.1), if the densities  $K_\varepsilon$  satisfy

$$\sup_{\varepsilon > 0} \varepsilon^{d/2} \|K_\varepsilon\|_{L^2} = \mathbf{K} < \infty,$$

then for any  $\delta, \delta' \in (0, 1)$  we have

$$(1.8) \quad t_{\text{mix}}(\delta') \leq 2 + \log_\delta \left( \frac{\delta' \varepsilon^{d/2}}{\mathbf{K}} \right) t_{\text{dis}}(\delta).$$

The  $L^2$  convergence of systems in the form (1.1) have been studied by many authors [TC03, FW03, FNW04, FNW06, FI19, OTD21]. The work of Fannjiang et al. [FW03, FNW04, FNW06] studies the dissipation time under certain spectral assumptions on the associated Koopman operator. These assumptions, however, are hard to verify.

Previous work of Feng and the first author [FI19] used a Fourier splitting method to show that if  $\varphi$  is  $C^1$  and generates an exponentially mixing dynamical system, then the dissipation time of the Markov process (1.1) is bounded by

$$(1.9) \quad t_{\text{dis}} \leq C |\ln \varepsilon|^2,$$

for some finite constant  $C$  and all sufficiently small  $\varepsilon > 0$ . An immediate consequence of this is the mixing time bound (1.2) mentioned above.

**Corollary 1.4.** *Suppose  $K_\varepsilon$  is a periodized rescaled standard Gaussian (i.e.  $K_\varepsilon$  is defined by (2.4), below, with  $\tilde{K}$  being a standard Gaussian in  $\mathbb{R}^d$ ). If  $\varphi: \mathbb{T}^d \rightarrow \mathbb{T}^d$  is  $C^1$  and generates an exponentially mixing, dynamical system, then there exists a constant  $C$  such that (1.2) holds for all sufficiently small  $\varepsilon$ .*

*Proof.* Corollary 2.5 in [FI19] proves (1.9), and now Proposition 1.3 implies (1.2).  $\square$

We now return to studying the process (1.1) when  $\varphi$  is the piecewise affine linear Bernoulli map described earlier. Such maps generate exponentially mixing dynamical systems (see for instance [KH95, SOW06]), and hence [FI19] guarantees the  $O(|\ln \varepsilon|^2)$  upper bound (1.9). Of course, if  $K_\varepsilon$  satisfies the assumptions in Theorem 1.1, then Proposition 1.3 will imply the stronger upper bound

$$(1.10) \quad t_{\text{dis}} \leq t_{\text{mix}} \left( \frac{1}{16} \right) \leq 4t_{\text{mix}} \leq 4C|\ln \varepsilon|,$$

for some constant  $C > 0$ . We will obtain now improve (1.10) by obtaining the exact constant on the right, and making less restrictive assumptions on  $K_\varepsilon$ .

**Theorem 1.5.** *Suppose Assumption 2.1 holds and  $\text{supp}(K_\varepsilon) \subseteq B(0, \varepsilon)$ . Moreover, assume that the Fourier coefficients of  $K_\varepsilon$  satisfy*

$$(1.11) \quad \sup_{|k| > \frac{\varepsilon}{c}} |\hat{K}_\varepsilon(k)| < 1,$$

for any  $c > 0$  and all sufficiently small  $\varepsilon > 0$ . Then, for every  $\delta \in (0, 1)$ , there exists an explicit constant  $C$ , such that

$$(1.12) \quad \frac{d \ln \varepsilon}{\ln p_{\min}} - C(1 + |\ln(1 - \delta)|) \leq t_{\text{dis}}(\delta) \leq \frac{d \ln \varepsilon}{\ln p_{\max}} + C \left( 1 + \frac{|\ln \delta|}{\delta^2} \right)$$

for all sufficiently small  $\varepsilon > 0$ . Here  $p_{\min}$  is the measure of the smallest domain on which  $\varphi$  is affine linear.

*Remark 1.6.* Since  $t_{\text{dis}}(\delta)$  is a decreasing function of  $\delta$ , the lower bound in (1.12) is useful when  $\delta$  is close to 0, and the upper bound is useful when  $\delta$  is close to 1.

In general,  $p_{\min} < p_{\max}$ , and so there may be a gap between the upper and lower bounds in (1.12). However, in the symmetric case (when each of the regions on which  $\varphi$  is affine linear have equal size), the upper and lower bounds in (1.12) match. Obtaining matching upper and lower bounds for the mixing time is indicative of a *cutoff phenomenon* [Dia96, LPW09, BHP17]. Although we believe the cutoff phenomenon happens even when  $p_{\min} < p_{\max}$ , we are presently unable to prove it.

The main idea behind our proofs is to find a sufficiently large family of initial distributions which are both mixed well by  $\varphi$ , and not perturbed too much by the noise. For Theorem 1.1 the family we use are bump functions on ‘‘cylinder sets’’ (which will be defined shortly). These bump functions will vanish linearly near the boundary, giving a good estimate on the rate at which  $X$  leaves cylinder sets. The key to using this is an eigenfunction like inequality (3.5), explained in Section 3, below.

For Theorems 1.2 and 1.5 the family of initial distributions we use are piecewise constant on cylinder sets. This family of distributions are better mixed by the deterministic dynamics of  $\varphi$ , which is what leads to the explicit coefficients of  $|\ln \varepsilon|$  coefficients in (1.12). The drawback of this method is that to apply it one needs to first approximate the initial distribution by one which is piecewise constant to a high degree of accuracy. Using Fourier projections, we are able to obtain convergence of the density of  $X$  in  $L^2$ , but not in  $L^1$ . This allows us to obtain the dissipation time bounds (1.12), but not the mixing time bound (1.3). In order to obtain a bound on the mixing time by this method, one would need to obtain suitable  $W^{1,1}$  decay of the density, which we are presently unable to do.

There are several elementarily stated variants of our main results that *can not* be proved by our methods, and require new ideas.

- (1) What is the mixing time if the cylinder sets are cuboids, and not cubes (e.g. the map  $\varphi$  shown on the right of Figure 1)? More generally, what is the mixing time if the map  $\varphi$  is Bernoulli and the cylinder sets ‘‘shrink nicely’’? This is particularly interesting in light of a recent striking result [DKRH21] which shows that any  $C^{1+\alpha}$  exponentially mixing map is Bernoulli. For general Bernoulli systems the cylinder sets could be very irregular. Our methods can not be used to address this situation and new ideas are required.
- (2) If the map  $\varphi$  is a bi-infinite Bernoulli shift (such as the bakers map [SOW06, KH95], or the folded bakers map appearing in [EZ19]), then is the mixing time still  $O(|\ln \varepsilon|)$ ? Numerical evidence suggests this is true, but we are presently unable to generalize our methods to this situation.
- (3) Does the Markov process  $X$  exhibit a cutoff phenomenon? When the map  $\varphi$  is a uniformly expanding map, this is easy to show. When  $p_{\max} = p_{\min}$ , the bound (1.12) shows a cutoff phenomenon in the  $L^2$  sense, but not necessarily in the usual total variation sense. In general, when  $p_{\max} \neq p_{\min}$ , and  $\varphi$  is only piecewise affine linear, we suspect  $X$  exhibits a cutoff phenomenon, but are presently unable to prove it.
- (4) In the spatially discrete setting, a recent result of Chatterjee and Diaconis [CD20] shows that choosing  $\varphi$  to be almost any bijection on the state space results in an exponential speedup. Is there a continuous space analog of this result?

Finally, we mention related questions in the continuous time setting. In the context of fluid dynamics, there has been a lot of recent activity studying the notion and applications of *enhanced dissipation* (see for instance [CKRZ08, Zla10, BCZ17, FI19, Wei19, CZDE20, FFIT20, BBPS21, CCZW21, IXZ21, FM22, IZ22, Sei22]). This notion measures faster convergence of the associated PDE to its equilibrium distribution. The motivation for the present paper was to provide a different perspective and study this phenomenon using convergence of Markov processes. While we are able to obtain optimal bounds in one discrete time setting, there are many related continuous time settings where the optimal bounds are open.

**Plan of this paper.** In Section 2 we describe  $\varphi$ , and state the assumptions required for our main results precisely. In Section 3 we prove Theorem 1.1 by constructing a family of bump functions supported on cylinder sets. In Section 4 we prove

Theorem 1.2 by estimating the rate at which piecewise constant functions get mixed. In Section 5 we prove Proposition 1.3 relating the mixing time and dissipation time, and in Section 6 we prove Theorem 1.5 obtaining upper and lower bounds on  $t_{\text{dis}}$ . Finally, in Appendix A we show that when  $\varphi$  is a linear expanding map (or an ergodic toral automorphism) we show that the convergence of  $X_n$  happens at a double exponential rate.

## 2. Notation and Preliminaries

**2.1. Piecewise affine linear Bernoulli maps.** We begin by precisely describing the map  $\varphi : \mathbb{T}^d \rightarrow \mathbb{T}^d$  used in (1.1). Let  $M \geq 2$  and  $\check{E}_1, \dots, \check{E}_M \subseteq \mathbb{R}^d$  be a partition of  $[0, 1)^d$ . When  $d = 1$  we assume each  $\check{E}_i$  is an interval, and when  $d > 1$  we assume each  $\check{E}_i$  is a  $d$ -dimensional cube with sides parallel to the coordinate axes (as in the left figure in Figure 1).

Let  $p_i = \check{\pi}(\check{E}_i) \in (0, 1)$ , where  $\check{\pi}$  denotes the Lebesgue measure on  $\mathbb{R}^d$ . We will assume  $\check{\varphi}_i : \check{E}_i \rightarrow [0, 1)^d$  are affine linear bijections of the form

$$(2.1) \quad \check{\varphi}_i(\check{x}) = \frac{\check{D}_i \check{x}}{p_i^{1/d}} + \check{e}_i,$$

for some vectors  $\check{e}_i \in \mathbb{R}^d$ , and orthogonal matrices  $\check{D}_i$ .

We now project onto the torus  $\mathbb{T}^d$ , and define the map  $\varphi : \mathbb{T}^d \rightarrow \mathbb{T}^d$ . Let  $\Pi : \mathbb{R}^d \rightarrow \mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$  denote the canonical projection,  $E_i = \Pi(\check{E}_i)$ , and let

$$\varphi_i(x) = \Pi \circ \check{\varphi}_i(\check{x}),$$

where  $\check{x} \in [0, 1)^d$  is the unique point such that  $\Pi(\check{x}) = x \in \mathbb{T}^d$ . Define the expanding map  $\varphi : \mathbb{T}^d \rightarrow \mathbb{T}^d$  by

$$(2.2) \quad \varphi(x) = \varphi_i(x), \quad \text{if } x \in E_i.$$

With this notation, we define quantities  $p_{\min}$  and  $p_{\max}$  (appearing in Theorems 1.2 and 1.5) by

$$p_{\min} \stackrel{\text{def}}{=} \min_{i \in \mathcal{I}} \pi(E_i), \quad \text{and} \quad p_{\max} \stackrel{\text{def}}{=} \max_{i \in \mathcal{I}} \pi(E_i).$$

Note that

$$\sum_{i \in \mathcal{I}} p_i = \sum_{i \in \mathcal{I}} \pi(E_i) = \pi(\mathbb{T}^d) = 1.$$

Hence, for any Borel set  $A \subseteq \mathbb{T}^d$ ,

$$\pi(\varphi^{-1}(A)) = \sum_{i=1}^M \pi(\varphi_i^{-1}(A)) = \pi(A) \sum_{i=1}^M p_i = \pi(A),$$

which shows that  $\varphi$  preserves the Lebesgue measure  $\pi$ . This implies that the unique stationary distribution of the process (1.1) is also the Lebesgue measure  $\pi$ .

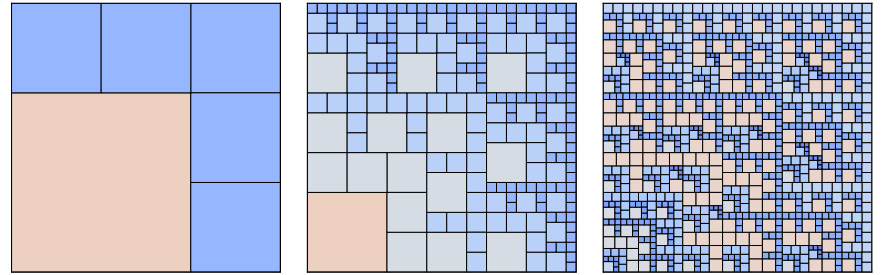


FIGURE 2. A few examples of cylinder sets. The leftmost figure shows order 0 cylinder sets, which are simply the domains  $E_1, \dots, E_6$ . The middle figure shows all cylinder sets of order 2. The rightmost figure shows a partition of the torus into cylinder sets of different orders, that will be used in the proof.

**2.2. Cylinder sets and the Bernoulli shift.** Our analysis of  $X_n$  relies on the fact that the map  $\varphi$  has the structure of a *Bernoulli shift* on the space of one-sided sequences (see for instance [KH95, SOW06]). The building block for functions that are controllably mixed will be based on *cylinder sets*, which we define in this section.

Let  $\mathcal{I} = \{1, \dots, M\}$ , and  $\mathcal{T}$  denote the set of all finite length  $\mathcal{I}$ -valued tuples. Explicitly,

$$\mathcal{T} = \{\mathbf{0}\} \cup \bigcup_{m=1}^{\infty} \mathcal{I}^m,$$

where  $\mathbf{0}$  denotes the empty tuple. Given a tuple  $s = (s_0, \dots, s_{m-1}) \in \mathcal{T}$  we use  $|s| = m$  to denote the length of the tuple  $s$ , with  $|\mathbf{0}| = 0$  by convention.

Let  $\sigma : \mathcal{T} \rightarrow \mathcal{T}$  be the *Bernoulli left shift*. That is,  $\sigma(s)$  removes the first coordinate of  $s$  and shifts the other coordinates left. More precisely, we define

$$\sigma(s_0, \dots, s_{m-1}) = (s_1, \dots, s_{m-1}), \quad \text{and} \quad \sigma(\mathbf{0}) = \mathbf{0}.$$

Let  $\sigma^k$  denote the  $k$ -fold composition of the map  $\sigma$ .

Now given a map  $\varphi$  as in (2.2), we can define the associated cylinder set associated to a tuple  $s \in \mathcal{T}$  by

$$(2.3) \quad \mathcal{C}_s \stackrel{\text{def}}{=} \{x \in \mathbb{T}^d \mid \varphi^n(x) \in E_{s_n} \text{ for all } n \leq |s|\}.$$

When  $s = \mathbf{0}$ , the associated cylinder set  $\mathcal{C}_s$  is the whole torus  $\mathbb{T}^d$ . When  $s = (s_0)$  is a tuple of length 1, the associated cylinder set  $\mathcal{C}_s$  is simply the domain  $E_{s_0}$  (see for instance the leftmost figure in Figure 2). When  $|s| > 1$ , each of these cylinder sets get subdivided further, forming finer and finer partitions of the torus (e.g. the middle figure in Figure 2).

Note that the action of  $\varphi$  on cylinder sets is simply the Bernoulli shift. That is, for any  $s \in \mathcal{T}$ ,

$$\varphi(\mathcal{C}_s) = \mathcal{C}_{\sigma s}.$$

In particular, this means that  $\varphi^{|s|}(\mathcal{C}_s) = \mathbb{T}^d$ , and so an initial distribution that is supported on  $\mathcal{C}_s$  becomes spread over  $\mathbb{T}^d$  after  $|s|$  iterations of  $\varphi$ . This can be

used to show that  $\varphi$  generates an *exponentially mixing* dynamical system (see for instance [KH95, SOW06]).

Since the process  $X$  is constructed by intertwining the action of  $\varphi$  with noise, this suggests that if  $X_0$  is concentrated on one cylinder set  $\mathcal{C}_s$ , then the distribution of  $X_{|s|}$  should be spread out over the whole torus. This, however, is not easy to prove as the action of the noise does not necessarily commute with the dynamics of  $\varphi$ . The main idea behind the proof of Theorem 1.1 is to construct a special distribution that is supported on  $\mathcal{C}_s$  and is provably mixed after  $|s|$  iterations. We do this in Section 3.1, below.

**2.3. Assumptions on  $\varphi$  and the noise.** We now precisely describe the assumptions that are required for our results. The first assumption is on the geometry of the cylinder sets.

**Assumption 2.1.** *Assume that the following hold:*

- (1) Each  $\varphi_i: E_i \rightarrow \mathbb{T}^d$  (defined in (2.2)) is a bijection.
- (2) For every  $\check{x} \in \partial[0, 1]^d$ , there exists  $i \in \mathcal{I}$  and  $\check{y} \in (0, 1)^d$  such that  $\check{\varphi}_i(\check{y}) = \check{x}$ .
- (3) If  $d \geq 2$ , every cylinder set  $\mathcal{C}_s$  (as defined in (2.3)) is an axis-aligned cube.

If the cylinder sets are not exactly cubes, Theorems 1.6 and 1.5 can still be used provided the *volume to perimeter* ratio of cylinder sets is controlled by the diameter. Specifically, the quantity that needs to stay bounded is the right hand side of (4.5) that appears in Lemma 4.1 from Section 4 below. This condition, however, is hard to verify for general Bernoulli systems, and so we restrict our attention to the piecewise affine linear case in this paper.

The assumptions on the noise distribution in Theorem 1.1 are a little more restrictive. The noise has to either be a periodized, non-degenerate Gaussian, or a sum of separated kernels with controlled first moments. We state this as our next two assumptions.

**Assumption 2.2.** *In addition to Assumption 2.1, suppose the following hold:*

- (1) The function  $\check{K}$  is the density of a Gaussian in  $\mathbb{R}^d$  with mean 0 and non-degenerate covariance matrix, and the densities  $K_\varepsilon$  are obtained by rescaling and periodizing  $\check{K}$ . Explicitly,  $K_\varepsilon$  is defined by

$$K_\varepsilon(x) = \sum_{n \in \mathbb{Z}^d} \check{K}_\varepsilon(\check{x} + n), \quad \check{K}_\varepsilon(\check{x}) = \frac{1}{\varepsilon^d} \check{K}\left(\frac{\check{x}}{\varepsilon}\right). \quad (2.4)$$

- (2) The function  $\check{K}$  is invariant under the action of each orthogonal matrix  $\check{D}_i$  in (2.1). That is, we assume

$$\check{K}(\check{D}_i \check{x}) = \check{K}(x) \quad \text{for all } \check{x} \in \mathbb{R}^d. \quad (2.5)$$

*Remark.* The invariance assumption (2.5) is satisfied if each of the orthogonal matrices  $\check{D}_i$  commute with the covariance matrix of  $\check{K}$ .

**Assumption 2.3.** *In addition to Assumption 2.1, suppose the following hold:*

- (1) The distributions  $K_\varepsilon$  are obtained by rescaling and periodizing a distribution  $\check{K}$  (as in (2.4)), and  $\check{K}$  satisfies the conditions below.

- (2) For every  $\eta > 0$  we have

$$\inf_{\check{x}, \check{y} \in [0, \eta]^d} \check{K}(\check{x} - \check{y}) > 0.$$

- (3) There exists a family of densities  $\{\check{K}_n\}_{n \geq 1}$  such that  $\check{K}_n \rightarrow \check{K}$  in  $L^1(\mathbb{R}^d)$ , and each  $\check{K}_n$  is in the form

$$\check{K}_n(\check{x}) = \sum_{i=1}^n b_{n,i} \prod_{j=1}^d \check{k}_{n,i}(\check{x}_j), \quad (2.6)$$

for some even functions  $\check{k}_{n,i}: \mathbb{R} \rightarrow \mathbb{R}$ , and constants  $b_{n,i} \geq 0$ . Moreover, there exists a constant  $A > 0$  such that for every  $n, i$  we have

$$\int_{\mathbb{R}} \check{k}_{n,i}(\check{x}_j) d\check{x}_j = 1, \quad \int_{\mathbb{R}} |\check{x}_i| \check{k}_{n,i}(\check{x}_j) d\check{x}_j \leq A, \quad \int_0^{\frac{1}{2}} \check{x}_i \check{k}_{n,i}(\check{x}_j) d\check{x}_j \geq \frac{1}{A}. \quad (2.7)$$

- (4) The density  $\check{K}$  is invariant under the action of each  $\check{D}_i$  as in (2.5).

We do not presently have a simple description of the class of probability densities that satisfy condition 3 in Assumption 2.3. If (2.7) did not require a *lower bound* on the first moments, then any compactly supported  $L^1$  probability distribution can be expressed as the  $L^1$  limit of distributions  $\check{K}_n$  in the form (2.6). The lower bound, however, breaks the standard approximation arguments.

### 3. Upper bounds on the Mixing Time.

**3.1. Proof of Theorem 1.1.** As mentioned above, the main idea behind the proof of Theorem 1.1 is to construct a family of ‘‘bump functions’’ supported on cylinder sets whose behavior is controlled under the evolution of  $X$ .

To state this, it is convenient to introduce the operators  $T_*$  and  $U_*$ . The operator  $T_*$  is the push forward of a distribution by the transition kernel of  $X$ . Explicitly, we define

$$(3.1) \quad T_* \mu(y) \stackrel{\text{def}}{=} \int_{\mathbb{T}^d} \mu(dx) \rho(x, y),$$

where  $\rho(x, y)$  is the transition density of the process  $X$ , and  $\mu$  is a finite measure. If  $\mu \ll \pi$  and  $\frac{d\mu}{d\pi} = f \in L^1(\pi)$ , we define the action of  $T_*$  on the density  $f$  by

$$(3.1') \quad T_* f(y) = \int_{\mathbb{T}^d} \pi(dx) f(x) \rho(x, y).$$

From (1.1) we note that we may also write

$$(3.2) \quad T_* f = K_\varepsilon * U_* f, \quad \text{where} \quad U_* f \stackrel{\text{def}}{=} \sum_{i=1}^M f \circ \varphi_i^{-1} |\det(D\varphi_i^{-1})|.$$

Recall by Assumption 2.1, all cylinder sets are intervals for  $d = 1$  and axis-aligned cubes for  $d > 1$ . We will use  $\ell_s$  denote the length of the interval  $\mathcal{C}_s$  when  $d = 1$ ,

and the side length of the cube  $\mathcal{C}_s$  when  $d > 1$ . For convenience define  $\lambda_s = 1/\ell_s$ . Explicitly,

$$(3.3) \quad \ell_s = \pi(\mathcal{C}_s)^{1/d}, \quad \text{and} \quad \lambda_s = \frac{1}{\ell_s} = \frac{1}{\pi(\mathcal{C}_s)^{1/d}}.$$

**Lemma 3.1.** *Suppose either Assumption 2.2 or Assumption 2.3 hold. Then there exist a family of continuous functions  $\{\mathbf{F}_s \mid s \in \mathcal{T}\}$  with the following properties. Each function  $\mathbf{F}_s$  is supported on  $\mathcal{C}_s$ , is strictly positive on the interior of  $\mathcal{C}_s$ , is normalized so that  $\|\mathbf{F}_s\|_{L^1} = 1$ , and satisfies the identity*

$$(3.4) \quad U_* \mathbf{F}_s = \mathbf{F}_{\sigma_s}, \quad \text{for all } s \in \mathcal{T} - \{\mathbf{0}\}.$$

Moreover, there exists  $a < \infty$ ,  $\gamma > 0$  such that for all  $s \in \mathcal{T}$  we have

$$(3.5) \quad K_\varepsilon * \mathbf{F}_s \geq (1 - a(\lambda_s \varepsilon)^\gamma) \mathbf{F}_s.$$

When  $K_\varepsilon$  is a periodized rescaled Gaussian, one can simply choose  $\mathbf{F}_s$  to be the principal eigenfunctions of an elliptic operator (see the proof in Section 3.4, below). In the other case the construction is more involved and is presented in Section 3.5, below.

The key to proving Theorem 1.1 is to show that if the density of the distribution of  $X_0$  is  $\mathbf{F}_s$ , then after time  $|s| + 1$  the density of the distribution of  $X_{|s|+1}$  is bounded below, away from 0. This is our next lemma.

**Lemma 3.2.** *Let  $a, \gamma$  be as in Lemma 3.1 and set  $\eta = (2a)^{1/\gamma}$ . There exist constants  $\beta, \beta' \in (0, 1]$  such that for all  $\varepsilon > 0$  and all  $s \in \mathcal{T}$  such that  $\ell_s \geq \eta\varepsilon$ , we have*

$$(3.6) \quad T_*^n \mathbf{F}_s \geq \begin{cases} \beta \mathbf{F}_{\sigma^n s}, & \forall n \leq |s|, \\ \beta', & \forall n \geq |s| + 1. \end{cases}$$

By bounding the density of  $X_1$  below by a combination of the functions  $\mathbf{F}_s$  above, we claim that any initial distribution becomes bounded away from 0 in time  $O(|\ln \varepsilon|)$ .

**Lemma 3.3.** *Let  $\eta > 0$  be as in Lemma 3.2, and define*

$$(3.7) \quad N \stackrel{\text{def}}{=} \left\lceil \frac{d \ln(\varepsilon \eta)}{\ln(p_{\max})} \right\rceil.$$

There is  $\chi > 0$  such that for all sufficiently small  $\varepsilon > 0$  and every probability measure  $\mu$ , we have

$$\frac{d(T_*^{N+2} \mu)}{d\pi} \geq \chi.$$

The expression  $d(T_*^{N+2} \mu)/d\pi$  denotes the Radon–Nikodym derivative of  $T_*^{N+2} \mu$  with respect to the Lebesgue measure  $\pi$ . Momentarily postponing the proofs of Lemmas 3.1–3.3, we prove Theorem 1.1.

*Proof of Theorem 1.1.* Let  $\mu_0$  be any probability measure, and inductively define

$$\mu_{n+1} = \frac{T_*^{N+2} \mu_n - \chi \pi}{1 - \chi}.$$

By Lemma 3.3 we see that  $\mu_n$  is a positive measure, and hence by our normalization must be a probability measure. Since

$$T_*^{n(N+2)} \mu_0 = (1 - (1 - \chi)^n) \pi + (1 - \chi)^n \mu_n,$$

we note

$$\|T_*^{n(N+2)} \mu_0 - \pi\|_{\text{TV}} \leq (1 - \chi)^n,$$

which immediately implies

$$t_{\text{mix}} \leq \frac{(N+2) \ln 2}{|\ln(1 - \chi)|}.$$

Using the definition of  $N$  (equation (3.7)) we obtain (1.3) as claimed.  $\square$

**3.2. A lower bound for  $T_*^n \mathbf{F}_s$  (Lemma 3.2).** The main idea behind the proof of Lemma 3.2 is to control the mass that escapes the envelope of  $\mathbf{F}_s$  through the noise. Once this is established, repeated application of  $U_*$  to  $\mathbf{F}_s$  will give a function that bounded away from 0.

*Proof of Lemma 3.2.* Using (3.4) and (3.5) we see

$$(3.8) \quad T_* \mathbf{F}_s = K_\varepsilon * U_* \mathbf{F}_s \geq (1 - a(\lambda_{\sigma_s} \varepsilon)^\gamma) \mathbf{F}_{\sigma_s}.$$

Since  $\ell_s \geq \varepsilon \eta = \varepsilon(2a)^{1/\gamma}$  by assumption, we must have  $(1 - a(\lambda_s \varepsilon)^\gamma) \geq 1/2$ . Therefore,

$$1 - a(\lambda_{\sigma_s} \varepsilon)^\gamma \geq e^{-C(\lambda_{\sigma_s} \varepsilon)^\gamma}$$

where  $C = 2 \ln 2$  is independent of  $\varepsilon$ . Using this and iterating (3.8) gives

$$(3.9) \quad T_*^n \mathbf{F}_s \geq \left( \prod_{k=1}^n (1 - a(\lambda_{\sigma^k s} \varepsilon)^\gamma) \right) \mathbf{F}_{\sigma^n s} \geq \exp\left(-C\varepsilon^\gamma \sum_{k=1}^n \lambda_{\sigma^k s}^\gamma\right) \mathbf{F}_{\sigma^n s}.$$

for any  $n \in \mathbb{N}$ .

The sum in the exponential is easily bounded for  $n \leq |s|$ . Indeed, if  $s = (s_0, \dots, s_{n'-1}) \in \mathcal{T}$ , then

$$\lambda_{\sigma s} = \frac{1}{\pi(\mathcal{C}_{\sigma s})^{1/d}} = \frac{p_{s_0}^{1/d}}{\pi(\mathcal{C}_s)^{1/d}} \leq p_{\max}^{1/d} \lambda_s.$$

Hence for every  $n \leq |s|$  we have

$$\varepsilon^\gamma \sum_{k=1}^n \lambda_{\sigma^k s}^\gamma \leq \frac{(\varepsilon \lambda_{\sigma s})^\gamma}{1 - p_{\max}^{\gamma/d}} \leq \frac{1}{\eta^\gamma (1 - p_{\max}^{\gamma/d})},$$

which is finite and independent of  $n$  and  $\varepsilon$ . Using this in (3.9) implies (3.6) holds, for all  $n \leq |s|$  and some nonnegative,  $\varepsilon$ -independent constant  $\beta > 0$  that can be computed explicitly to be

$$\beta = \exp\left(\frac{-2 \ln(2)}{2a(1 - p_{\max}^{\gamma/d})}\right).$$

To handle the case  $n > |s|$  we note first that we have already proved

$$(3.10) \quad T_*^{|s|} \mathbf{F}_s \geq \beta \mathbf{F}_0.$$

By assumption  $\check{\mathbf{F}}_0$  is continuous and strictly positive in  $(0, 1)^d$ , and hence condition 3 in Assumption 2.1 implies  $U_*\mathbf{F}_0 > 0$  on  $\mathbb{T}^d$ . This in turn implies  $T_*\mathbf{F}_0 > 0$  on  $\mathbb{T}^d$ . Since  $\mathbf{F}_0$  is continuous, this implies  $\min_{\mathbb{T}^d} \mathbf{F}_0 > 0$ , and hence applying  $T_*$  to (3.10) shows

$$T_*^{|s|+1} \mathbf{F}_s \geq \beta \left( \min_{\mathbb{T}^d} T_* \mathbf{F}_0 \right) = \beta'.$$

This implies (3.6) for  $n = |s| + 1$ , and some finite constant  $\beta'$  that is independent of  $s, \varepsilon$ . Since  $T_*$  can not decrease a nonnegative minimum, we also obtain (3.6) for all  $n \geq |s| + 1$ , concluding the proof.  $\square$

**3.3. A lower bound on  $T_*^n \mu$  (Lemma 3.3).** The main idea behind the proof of Lemma 3.3 is to partition the torus into cylinder sets  $\mathcal{C}_s$  with side length  $\ell_s = O(\varepsilon)$ . If we apply  $T_*$  once to the initial measure  $\mu$ , then  $X_1$  has a density that is roughly uniform on sets at the scale  $\varepsilon$ , and hence can be bounded from below by linear combination of functions  $\mathbf{F}_s$  with non-negative coefficients. Applying Lemma 3.2 to this will allow us to show that the distribution eventually becomes bounded away from 0.

*Proof of Lemma 3.3.* We first define  $\mathcal{S}_\varepsilon$  by

$$\mathcal{S}_\varepsilon = \{s \in \mathcal{T} \mid \ell_{\sigma s} > \eta\varepsilon, \quad \ell_s \leq \eta\varepsilon\},$$

where we recall that  $a$  is the constant appearing in Lemma 3.1 (see the rightmost figure in Figure 2 for an illustration). For all  $s \in \mathcal{T}$  with  $|s| \geq 1$ , the side length  $\ell_s$  satisfies

$$\left(p_{\min}^{1/d}\right)^{|s|} \leq \ell_s \leq \left(p_{\max}^{1/d}\right)^{|s|}, \quad p_{\min}^{1/d} \ell_{\sigma s} \leq \ell_s \leq p_{\max}^{1/d} \ell_{\sigma s}.$$

So, for every  $s \in \mathcal{S}_\varepsilon$  we must have

$$(3.11) \quad p_{\min}^{1/d} \eta\varepsilon \leq \ell_s \leq \eta\varepsilon,$$

and

$$(3.12) \quad |s| - 1 \leq N = \left\lfloor \frac{d \ln(\varepsilon\eta)}{\ln p_{\max}} \right\rfloor.$$

Since  $\check{K} \in L^1$ , the measure  $T_*\mu$  is absolutely continuous with respect to the Lebesgue measure  $\pi$ , and we let

$$f \stackrel{\text{def}}{=} \frac{d(T_*\mu)}{d\pi},$$

denote the Radon–Nikodym derivative of  $T_*\mu$ . Define

$$(3.13) \quad c_s \stackrel{\text{def}}{=} \frac{1}{\|\mathbf{F}_s\|_{L^\infty(\mathcal{C}_s)}} \inf_{x \in \mathcal{C}_s} f(x), \quad s \in \mathcal{S}_\varepsilon$$

and note each  $c_s$  is nonnegative and

$$(3.14) \quad f(x) \geq \sum_{s \in \mathcal{S}_\varepsilon} c_s \mathbf{F}_s(x).$$

We now claim that

$$(3.15) \quad \sum_{s \in \mathcal{S}_\varepsilon} c_s \geq \frac{1}{C_0},$$

for some constant  $C_0 > 0$  that is independent of  $\varepsilon$ .

To prove (3.15), let  $\nu = U_*\mu$ . Note that the lower bound in (3.11) implies that for any  $s \in \mathcal{S}_\varepsilon$ ,  $\mathcal{C}_s$  is a cube with side length at most  $\eta\varepsilon$ . Thus, for any  $s \in \mathcal{S}_\varepsilon$  and  $x \in \mathcal{C}_s$ , we have

$$(3.16) \quad f(x) = K_\varepsilon * \nu(x) \geq \int_{\mathcal{C}_s} K_\varepsilon(x-y) \nu(dy) \geq \frac{\nu(\mathcal{C}_s)}{\varepsilon^d} \kappa$$

where

$$\kappa = \inf_{\check{x}, \check{y} \in [0, \eta]^d} \check{K}(\check{x} - \check{y}) > 0.$$

By (3.4) and (3.11) we note

$$(3.17) \quad \|\mathbf{F}_s\|_{L^\infty} = \frac{1}{\pi(\mathcal{C}_s)} \|\mathbf{F}_0\|_{L^\infty} \leq \frac{C_1}{\varepsilon^d},$$

for some constant  $C_1$  that is independent of  $\varepsilon$ . Using (3.17) and (3.16) in (3.13) we obtain

$$c_s \geq \frac{\nu(\mathcal{C}_s)\kappa}{C_1}, \quad \forall s \in \mathcal{S}_\varepsilon.$$

Since the sets  $\{\mathcal{C}_s \mid s \in \mathcal{S}_\varepsilon\}$  form a partition and  $\nu$  is a probability measure, summing the above over  $s \in \mathcal{S}_\varepsilon$  yields (3.15) as desired, with  $C_0 = C_1/\kappa$ .

Now, to finish we note that  $T_*$  is monotone on nonnegative functions. That is, if  $g_1, g_2$  are any two functions such that  $0 \leq g_1 \leq g_2$ , then  $0 \leq T_*g_1 \leq T_*g_2$ . By (3.12),  $|s| \leq 1 + N$  for all  $s \in \mathcal{S}_\varepsilon$ . Therefore, using (3.11), (3.14) and Lemma 3.2 implies

$$T_*^{N+1} f \geq \sum_{s \in \mathcal{S}_\varepsilon} c_s T_*^{N+1} \mathbf{F}_s \geq \sum_{s \in \mathcal{S}_\varepsilon} c_s \beta' \geq \frac{\beta'}{C_0}.$$

where  $C_0 > 0$  is the constant from (3.15), independent of  $\varepsilon$ . Choosing  $\chi = \beta'/C_0$ , the proof is complete.  $\square$

**3.4. Constructing  $\mathbf{F}_s$  (Lemma 3.1).** We now construct the family of functions  $\mathbf{F}_s$ . In light of the property (3.4), it is enough to find one function  $\check{\mathbf{F}}_0$  that satisfies a bound like (3.5).

**Lemma 3.4.** *Suppose either Assumption 2.2 or 2.3 holds. There exists a continuous function  $\check{\mathbf{F}}_0: \mathbb{R}^d \rightarrow [0, \infty)$ , and constants  $a < \infty, \gamma > 0$  such that*

$$(3.18) \quad \{\check{\mathbf{F}}_0 > 0\} = (0, 1)^d, \quad \int_{\mathbb{R}^d} \check{\mathbf{F}}_0 d\check{\pi} = 1, \\ \text{and} \quad \check{K}_\varepsilon * \check{\mathbf{F}}_0 \geq (1 - a\varepsilon^\gamma) \check{\mathbf{F}}_0,$$

for all  $\varepsilon > 0$ .

Given the function  $\check{\mathbf{F}}_0$ , we construct the functions  $\mathbf{F}_s$  by rescaling, and projecting to the torus.

*Proof of Lemma 3.1.* For any  $s \in \mathcal{T}$  we define

$$\check{\mathbf{F}}_s = \frac{1}{\pi(\mathcal{C}_s)} \mathbf{1}_{\mathcal{C}_s} \check{\mathbf{F}}_0 \circ \check{\varphi}^{|s|},$$

where

$$\check{C}_s = \{\check{x} \in [0, 1]^d \mid \Pi(\check{x}) \in C_s\}, \quad \text{and} \quad \check{\varphi}(\check{x}) = \check{\varphi}_i(\check{x}) \quad \text{if } \check{x} \in \check{E}_i.$$

We project these functions to the torus by defining

$$\mathbf{F}_s(x) \stackrel{\text{def}}{=} \check{\mathbf{F}}_s(\check{x}),$$

where as before  $\check{x} \in [0, 1]^d$  is the unique point such that  $\Pi(\check{x}) = x$ . We claim the functions  $\mathbf{F}_s$  satisfy all the properties required in the statement of Lemma 3.1.

Clearly  $\mathbf{F}_s$  is supported on  $C_s$ , and is strictly positive in the interior. By definition of  $U_*$  we see (3.4) holds, which in turn implies  $\|\mathbf{F}_s\|_{L^1} = \|\check{\mathbf{F}}_0\|_{L^1} = 1$ . It only remains to verify (3.5). To see this, note first that for all  $x \in C_s$  we have

$$\check{\varphi}^{|\mathbf{s}|}(x) = \lambda_s \check{D}\check{x} + \check{e}_s$$

for some  $\check{e}_s \in \mathbb{R}^d$ , and orthogonal matrix  $D$  that is a product of the matrices  $\check{D}_i$  in (2.1). Thus, setting  $n = |\mathbf{s}|$ , we see

$$\begin{aligned} \check{K}_\varepsilon * \check{\mathbf{F}}_s(\check{x}) &= \frac{1}{\pi(C_s)\varepsilon^d} \int_{\mathbb{R}^d} \check{K}\left(\frac{\check{y}}{\varepsilon}\right) \check{\mathbf{F}}_0\left(\lambda_s \check{D}(\check{x} - \check{y}) + \check{e}_s\right) d\check{y} \\ &= \frac{1}{\varepsilon^d} \int_{\mathbb{R}^d} \check{K}\left(\frac{\check{D}^{-1}\check{y}}{\lambda_s \varepsilon}\right) \check{\mathbf{F}}_0(\check{\varphi}^n(\check{x}) - \check{y}) d\check{y} = \lambda_s^d (\check{K}_{\lambda_s \varepsilon} * \check{\mathbf{F}}_0)(\check{\varphi}^n(\check{x})), \end{aligned}$$

where the last equality followed because of (2.5). Using (3.18) (with  $\varepsilon$  replaced by  $\lambda_s \varepsilon$ ), we note

$$(\check{K}_{\lambda_s \varepsilon} * \check{\mathbf{F}}_0) \geq (1 - a(\lambda_s \varepsilon)^\gamma) \check{\mathbf{F}}_0,$$

and hence

$$\check{K}_\varepsilon * \check{\mathbf{F}}_s(\check{x}) \geq \frac{1}{\pi(C_s)} (1 - a(\lambda_s \varepsilon)^\gamma) \check{\mathbf{F}}_0 \circ \check{\varphi}^n(\check{x}) = (1 - a(\lambda_s \varepsilon)^\gamma) \check{\mathbf{F}}_s(\check{x}).$$

This implies

$$K_\varepsilon * \mathbf{F}_s(x) \geq \check{K}_\varepsilon * \check{\mathbf{F}}_s(\check{x}) \geq (1 - a(\lambda_s \varepsilon)^\gamma) \check{\mathbf{F}}_s(\check{x}) = (1 - a(\lambda_s \varepsilon)^\gamma) \mathbf{F}_s(x),$$

yielding (3.5) as claimed.  $\square$

It remains to prove Lemma 3.4. Under Assumption 2.2, we can just choose  $\mathbf{F}_s$  to be the principal eigenfunctions of the associated elliptic operator, and we will obtain (3.5) with  $\gamma = 2$ . We do this next.

*Proof of Lemma 3.1 under Assumption 2.2.* Let  $\check{A} = (\check{a}_{i,j})$  be the covariance matrix of  $\check{K}$ , and define the differential operator  $\mathcal{L}$  by

$$\mathcal{L} = \frac{1}{2} \sum_{i,j=1}^d \check{a}_{i,j} \partial_i \partial_j.$$

Since  $\check{A}$  is non-degenerate, by assumption, the operator  $\mathcal{L}$  is elliptic. Let  $\check{\mathbf{F}}_0$  be the principal eigenfunction of  $\mathcal{L}$  on the unit square  $[0, 1]^d$ , with Dirichlet boundary conditions, normalized so that  $\|\check{\mathbf{F}}_0\|_{L^1} = 1$ . Since  $\mathcal{L}$  is elliptic we know that  $\check{\mathbf{F}}_0$  can be chosen to be strictly positive on  $(0, 1)^d$ , and the associated principal eigenvalue  $\check{\lambda}_0 > 0$ .

Define the functions  $\theta, \underline{\theta}$  by

$$\theta(t, x) = \check{K}_{\varepsilon\sqrt{t}} * \check{\mathbf{F}}_0(x), \quad \text{and} \quad \underline{\theta}(t, x) = e^{-\check{\lambda}_0 \varepsilon^2 t} \check{\mathbf{F}}_0(x).$$

Since  $\check{K}_\varepsilon$  is a Gaussian we know the function  $\theta$  satisfies the diffusion equation

$$(3.19) \quad \partial_t \theta = \varepsilon^2 \mathcal{L} \theta \quad \text{in } \mathbb{R}^d,$$

with initial data  $\theta(0, x) = \check{\mathbf{F}}_0$  (extended by 0 outside the cube  $[0, 1]^d$ ). Since

$$-\mathcal{L} \check{\mathbf{F}}_0 = \check{\lambda}_0 \check{\mathbf{F}}_0 \quad \text{in } (0, 1)^d,$$

and the outward normal derivative of  $\check{\mathbf{F}}_0$  is nonpositive on the boundary of the cube  $[0, 1]^d$ , the function  $\underline{\theta}$  is a sub-solution to the diffusion equation (3.19). As a result we must have  $\underline{\theta}_t \leq \theta_t$  for all  $t \geq 0$  and  $x \in [0, 1]^d$ . Setting  $t = 1$  yields

$$\check{K}_\varepsilon * \check{\mathbf{F}}_0 \geq e^{-\check{\lambda}_0 \varepsilon^2} \check{\mathbf{F}}_0.$$

Since  $e^{-t} \geq 1 - t$ , we also obtain (3.18) with  $\gamma = 2$  and  $a = \check{\lambda}_0$ , concluding the proof.  $\square$

*Remark.* When  $\check{K}$  is the standard Gaussian, then

$$\check{\mathbf{F}}_0(\check{x}) = \prod_{k=1}^d \sin(\pi \check{x}_k), \quad \text{and} \quad \lambda_0 = \frac{\pi^2 d}{2}.$$

It remains to prove Lemma 3.4 under Assumption 2.3. This is more involved, and we present the proof in Section 3.5.

**3.5. Constructing  $\check{\mathbf{F}}_0$  under Assumption 2.3 (Lemma 3.4).** In the non-Gaussian case (Assumption 2.3), we will start with  $d = 1$  and choose  $\check{\mathbf{F}}_0$  to be a tent function. This will eventually yield (3.18) with  $\gamma = 1$ . We begin with the calculation in one dimension.

**Lemma 3.5.** *Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be the tent like function defined by*

$$\check{f}(\check{x}) = \begin{cases} \check{x} & 0 \leq \check{x} < \frac{1}{2} \\ 1 - \check{x} & \frac{1}{2} \leq \check{x} \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

*If  $\check{K}: \mathbb{R} \rightarrow \mathbb{R}$  is an even function, then  $\check{\mathbf{F}}_0 = \check{f}/\|f\|_{L^1}$  satisfies (3.18) with  $\gamma = 1$ , and*

$$a = 4 \int_{-\infty}^{\infty} |\check{y}| \check{K}(\check{y}) d\check{y}.$$

*Proof.* Because  $\check{K}$  is even, and  $f$  is symmetric about  $\check{x} = 1/2$ , it suffices to prove the bound for  $\check{x} \in (0, 1/2]$ . For  $0 < \check{x} \leq 1/4$  we note

$$\begin{aligned} \check{f}(\check{x}) - \check{K}_\varepsilon * \check{f}(\check{x}) &= \int_{-\infty}^{\infty} (\check{x} - \check{f}(\check{x} - \check{y})) \check{K}_\varepsilon(\check{y}) d\check{y} \\ &\leq \int_{\check{x}-\frac{1}{2}}^{\check{x}} \check{y} \check{K}_\varepsilon(\check{y}) d\check{y} + \check{x} \int_{\check{y} \notin (\check{x}-\frac{1}{2}, \check{x})} \check{K}_\varepsilon(\check{y}) d\check{y} \end{aligned}$$



$$\leq \int_{-\check{x}}^{\check{x}} \check{y} \check{K}_\varepsilon(\check{y}) d\check{y} + \check{x} \int_{|\check{y}| \geq \frac{1}{4}} \check{K}_\varepsilon(\check{y}) d\check{y} \leq 4\varepsilon f(\check{x}) \int_{-\infty}^{\infty} |\check{y}| \check{K}(\check{y}) d\check{y},$$

since  $\check{K}$  is even. For  $1/4 < \check{x} \leq 1/2$  we note  $f(\check{x}) \geq 1/4$  and so

$$\begin{aligned} \check{f}(\check{x}) - \check{K}_\varepsilon * \check{f}(\check{x}) &\leq \int_{-\infty}^{\infty} |\check{x} - \check{f}(\check{x} - \check{y})| \check{K}_\varepsilon(\check{y}) d\check{y} \leq \int_{-\infty}^{\infty} |\check{y}| \check{K}_\varepsilon(\check{y}) d\check{y} \\ &\leq 4\varepsilon f(\check{x}) \int_{-\infty}^{\infty} |\check{y}| \check{K}(\check{y}) d\check{y}. \end{aligned}$$

Thus whenever  $\check{f} > 0$  we have the estimate

$$\check{f} - \check{K}_\varepsilon * \check{f} \leq a\varepsilon \check{f},$$

which immediately implies (3.18) with  $\gamma = 1$ .  $\square$

Next, in arbitrary dimension  $d \geq 1$ , we construct  $\check{F}_0$  in the case that  $\check{K}$  is separated.

**Lemma 3.6.** *Suppose  $\check{K} : \mathbb{R}^d \rightarrow \mathbb{R}$  is of the form*

$$\check{K}(\check{x}) = \prod_{i=1}^d \check{k}_i(\check{x}_i),$$

for some even functions  $\check{k}_i : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$\int_{\mathbb{R}} \check{k}_i(\check{x}_i) d\check{x}_i = 1, \quad \int_{\mathbb{R}} |\check{x}_i| \check{k}_i(\check{x}_i) d\check{x}_i = A, \quad \int_0^{\frac{1}{2}} \check{x}_i \check{k}_i(\check{x}_i) d\check{x}_i = \underline{A},$$

for constants  $A < \infty$  and  $\underline{A} > 0$ . Let  $\check{f}$  be the tent function from Lemma 3.5, and define

$$(3.20) \quad \check{F}(\check{x}) = \prod_{i=1}^d \check{f}(\check{x}_i), \quad \text{and} \quad \check{F}_0 = \frac{\check{F}}{\|\check{F}\|_{L^1}}.$$

There exists a constant  $a = a(d, A, \underline{A})$  such that (3.18) holds with  $\gamma = 1$  and all sufficiently small  $\varepsilon > 0$ .

*Proof.* Note first for any  $\check{x} \in (0, 1)^d$ ,

$$\check{K}_\varepsilon * \check{F}(x) \geq \check{K}_\varepsilon * \check{F}(0) \geq \prod_{i=1}^d \int_0^{1/2} \check{x}_i \check{k}_{i,\varepsilon}(\check{x}_i) d\check{x}_i \geq \left(\frac{\underline{A}\varepsilon}{2}\right)^d,$$

for all sufficiently small  $\varepsilon$ . Thus whenever  $\check{F}(\check{x}) < (\underline{A}\varepsilon/2)^d$  we have

$$(3.21) \quad \check{F}(\check{x}) - \check{K}_\varepsilon * \check{F}(\check{x}) \leq 0.$$

Now suppose  $\check{F}(\check{x}) \geq (\underline{A}\varepsilon/2)^d$ . Then for at least one  $i \in \{1, \dots, d\}$  we must have  $\check{f}(\check{x}_i) \geq \underline{A}\varepsilon/2$ . For simplicity, and without loss of generality, we assume  $i = 1$ . We will now use the notation  $\check{x} = (\check{x}_1, \check{x}')$  where  $\check{x}' = (\check{x}_2, \dots, \check{x}_d)$ , and  $\check{F}'(\check{x}') = \prod_{i=2}^d \check{f}(\check{x}_i)$ , etc. By induction on  $d$  we may also assume

$$\check{F}' - \check{K}'_\varepsilon * \check{F}' \leq a\varepsilon \check{F}',$$

for some dimensional constant  $a = a(A, \underline{A})$ . We will subsequently allow  $a$  to increase from line to line, as long as it does not depend on  $\varepsilon$  or  $\check{x}$ .

Now, we compute

$$\begin{aligned} \check{F}(\check{x}) - \check{K}_\varepsilon * \check{F}(\check{x}) &= \int_{\mathbb{R}^d} (\check{f}(\check{x}_1) \check{F}'(\check{x}') - \check{f}(\check{x}_1 - \check{y}_1) \check{F}'(\check{x}' - \check{y}')) \check{K}_\varepsilon(\check{y}) d\check{y} \\ &= \int_{\mathbb{R}^d} \check{F}'(\check{x}') (\check{f}(\check{x}_1) - \check{f}(\check{x}_1 - \check{y}_1)) \check{K}_\varepsilon(\check{y}) d\check{y} \\ &\quad + \int_{\mathbb{R}^d} \check{f}(\check{x}_1 - \check{y}_1) (\check{F}'(\check{x}') - \check{F}'(\check{x}' - \check{y}')) \check{K}_\varepsilon(\check{y}) d\check{y} \\ (3.22) \quad &\leq a\varepsilon \check{F}'(\check{x}') (\check{f}(\check{x}_1) + \check{k}_\varepsilon * \check{f}(\check{x}_1)). \end{aligned}$$

Above we used Lemma 3.5 to bound the first term and the induction hypothesis to bound the second integral.

Observe that

$$|\check{f}(x_1) - \check{k}_\varepsilon * \check{f}(\check{x}_1)| \leq \varepsilon \|\nabla \check{f}\|_{L^\infty} \int_{\mathbb{R}} |\check{x}_1| \check{k}_1(\check{x}_1) d\check{x}_1.$$

Using this and the assumption  $\check{f}(\check{x}_1) \geq \underline{A}\varepsilon/2$  in (3.22) yields

$$\check{F}(\check{x}) - \check{K}_\varepsilon * \check{F}(\check{x}) \leq a\varepsilon \check{F}'(\check{x}') (\check{f}(\check{x}_1) + \varepsilon) \leq a\varepsilon \check{F}'(\check{x}') \left( \check{f}(\check{x}_1) + \frac{2\check{f}(\check{x}_1)}{\underline{A}} \right) \leq a\varepsilon \check{F}(\check{x}).$$

Combining this with (3.21) concludes the proof.  $\square$

Given Lemma 3.6, a standard approximation argument can be used to deduce Lemma 3.4.

*Proof of Lemma 3.4 under Assumption 2.3.* Let  $\check{K}_n$  be as in (2.6), and define

$$\check{K}'_{n,i}(x) = \prod_{j=1}^d \check{k}_{n,i}(\check{x}_j).$$

Let  $\check{F}_0$  be the function defined in (3.20). By Lemma 3.6 we know there exists  $a = a(A, 1/A, d)$ , independent of  $n, i$  such that

$$\check{K}'_{n,i} * \check{F}_0 \geq (1 - a\varepsilon) \check{F}_0.$$

Multiplying by  $b_{n,i}$  (which are nonnegative by assumption) and summing yields

$$\check{K}'_n * \check{F}_0 \geq (1 - a\varepsilon) \check{F}_0 \sum_{i=1}^n b_{n,i}.$$

Since  $\check{K}_n \rightarrow \check{K}$  in  $L^1$  taking the limit as  $n \rightarrow \infty$  yields (3.18) as desired.  $\square$

## 4. Lower Bounds on the Mixing Time.

**4.1. Proof of the lower bound (Theorem 1.2).** The main idea behind the proof of Theorem 1.2 is to choose  $X_0$  to be uniformly distributed on a cylinder set, and track the distribution of  $X_n$  carefully. The action of  $\varphi$  pushes the distribution to be uniform on a cylinder set that is one order lower. If the noise does not change this

too much, then the mixing time can be bounded below by the order of the cylinder set.

To make this idea rigorous, we need to estimate  $\|T_*^n f - U_*^n f\|_{L^1}$  for functions which are piecewise constant on cylinder sets (recall  $T_*$  and  $U_*$  are defined in (3.1) and (3.2) respectively). For notational convenience given  $s \in \mathcal{T}$  we define the normalized indicator function  $\mathbf{I}_s$  by

$$\mathbf{I}_s \stackrel{\text{def}}{=} \frac{1}{\pi(\mathcal{C}_s)} \mathbf{1}_{\mathcal{C}_s}.$$

The action of  $\varphi$  on such functions is explicit. Indeed, for any  $s = (s_0, s_1, \dots, s_n) \in \mathcal{S}$  we see

$$U_* \mathbf{1}_{\mathcal{C}_s} = \sum_{i=1}^M \mathbf{1}_{\mathcal{C}_s} \circ \varphi_i^{-1} p_i = \mathbf{1}_{\mathcal{C}_s} \circ \varphi_{s_0}^{-1} p_{s_0} = p_{s_0} \mathbf{1}_{\mathcal{C}_{\sigma s}},$$

and hence

$$(4.1) \quad U_* \mathbf{I}_s = \frac{1}{\pi(\mathcal{C}_s)} U_* \mathbf{1}_{\mathcal{C}_s} = \frac{p_{s_0}}{\pi(\mathcal{C}_s)} \mathbf{1}_{\mathcal{C}_{\sigma s}} = \mathbf{I}_{\sigma s}.$$

Thus if  $n < |s|$ , then  $U_*^n \mathbf{I}_s = \mathbf{I}_{\sigma^n s}$ , which is not mixed and can be used to give a lower bound on the mixing time.

In order to bound  $\|T_*^n f - U_*^n f\|_{L^1}$  we need to control the amount of mass that leaks out of cylinder sets due to the action of the noise. We will shortly see that this is bounded by the *perimeter to volume ratio*, which we denote by  $H$ . Explicitly, if  $\mathcal{S} \subseteq \mathcal{T}$ , we define

$$H(\mathcal{S}) \stackrel{\text{def}}{=} \max_{s \in \mathcal{S} \setminus \{\mathbf{0}\}} \frac{\text{Per}(\mathcal{C}_s)}{\pi(\mathcal{C}_s)}, \quad \text{and by convention } H(\mathbf{0}) = 0.$$

In our case, all the cylinder sets are cubes by assumption, in which case  $H(\mathcal{S})$  can be expressed in terms of  $\lambda_s$  (which we recall is defined in (3.3)). Indeed, if  $\mathcal{S} \neq \{\mathbf{0}\}$ , then

$$(4.2) \quad H(\mathcal{S}) = \max_{s \in \mathcal{S} \setminus \{\mathbf{0}\}} \frac{2d}{\ell_s} = \max_{s \in \mathcal{S} \setminus \{\mathbf{0}\}} 2d\lambda_s.$$

We now present a lemma controlling the error  $\|T_*^n f - U_*^n f\|_{L^p}$  for piecewise constant functions and any  $p \in [1, \infty)$ . To prove Theorem 1.2 we only need  $p = 1$ . However, in order to prove Theorem 1.5 we will need  $p = 2$ .

**Lemma 4.1.** *Suppose that either  $p = 1$  and  $K_\varepsilon$  satisfies (1.5), or  $p > 1$  and  $K_\varepsilon$  is supported in the ball  $B(0, \varepsilon)$ . Let  $\mathcal{S} \subseteq \mathcal{T}$  be a finite set so that  $\{\mathcal{C}_s \mid s \in \mathcal{S}\}$  partitions the torus  $\mathbb{T}^d$ , and*

$$(4.3) \quad p_{\min}^{1/d} L\varepsilon \leq \ell_s \leq L\varepsilon,$$

for some constant  $L \geq 1$ . Suppose  $f \in L^p$  is of the form

$$(4.4) \quad f_0 = \sum_{s \in \mathcal{S}} c_0(s) \mathbf{1}_{\mathcal{C}_s}.$$

Then for all  $N \in \mathbb{N}$  we have

$$(4.5) \quad \|T_*^N f_0 - U_*^N f_0\|_{L^p} \leq \varepsilon^{1/p} C_1^{1/p'} \sum_{n=1}^N H(\sigma^n \mathcal{S})^{1/p} \|f_0\|_{L^p},$$

where  $p' = p/(p-1)$  is the Hölder conjugate of  $p$ , and

$$(4.6) \quad C_1 \stackrel{\text{def}}{=} 2d \left( 2 + \frac{1}{p_{\min}^{1/d}} \right)^{d-1}.$$

In our case the sum on the right of (4.5) can be bounded explicitly. Indeed, if  $s = (s_0, \dots, s_n) \in \mathcal{T}$ , we note

$$\ell_{\sigma s} = \pi(\mathcal{C}_{\sigma s})^{1/d} = \left( \frac{\pi(\mathcal{C}_s)}{p_{s_0}} \right)^{1/d} \geq \frac{\ell_s}{p_{\max}^{1/d}},$$

and hence  $H(\sigma \mathcal{S}) \leq p_{\max}^{1/d} H(\mathcal{S})$ . This immediately implies

$$(4.7) \quad \sum_{n=1}^N H(\sigma^n(\mathcal{S}))^{1/p} \leq \frac{H(\mathcal{S})^{1/p}}{1 - p_{\max}^{1/(pd)}}.$$

Thus, we obtain the following corollary.

**Corollary 4.2.** *Suppose either  $p = 1$  and  $K_\varepsilon$  satisfies (1.5), or  $p > 1$  and  $K_\varepsilon$  is supported in  $B(0, \varepsilon)$ . Let  $\delta \in (0, 1)$ , choose*

$$(4.8) \quad \Lambda_{p,\delta} \stackrel{\text{def}}{=} \frac{2dC_1^{p-1}}{\delta^p p_{\min}^{1/d} (1 - p_{\max}^{1/(pd)})^p},$$

and define  $\mathcal{S} = \mathcal{S}_{\varepsilon,\delta} \subseteq \mathcal{T}$  by

$$(4.9) \quad \mathcal{S}_{\varepsilon,\delta} = \{s \in \mathcal{T} \mid \ell_s \leq \varepsilon \Lambda_{p,\delta}, \text{ and } \ell_{\sigma s} > \varepsilon \Lambda_{p,\delta}\}.$$

If  $f_0$  is defined by (4.4), then for every  $N \in \mathbb{N}$  we have

$$(4.10) \quad \|T_*^N f_0 - U_*^N f_0\|_{L^p} \leq \delta.$$

*Proof.* Note  $\{\mathcal{C}_s \mid s \in \mathcal{S}_{\varepsilon,\delta}\}$  partitions the torus, and for every  $s \in \mathcal{S}_{\varepsilon,\delta}$  we have

$$(4.11) \quad p_{\min}^{1/d} \Lambda_{p,\delta} \varepsilon \leq \ell_s \leq \Lambda_{p,\delta} \varepsilon.$$

Moreover, since  $\delta < 2$ ,  $\Lambda_{p,\delta} \geq 1$ , and hence we may apply Lemma 4.1. Using (4.2), (4.7), and (4.8) in (4.5) immediately implies (4.10).  $\square$

Momentarily postponing the proof of Lemma 4.1, we prove Theorem 1.2.

*Proof of Theorem 1.2.* For any  $\delta \in (0, 1/2)$ , define

$$\delta' = \frac{1+2\delta}{2}, \quad \delta'' = \frac{1-2\delta}{2},$$

and let  $\mathcal{S} = \mathcal{S}_{\varepsilon,\delta''}$  be defined by (4.9) (with  $\delta = \delta''$ ). Let  $i \in \mathcal{I}$  be such that  $p_i = p_{\max}$ , and choose

$$N = \left\lceil \frac{d \ln \varepsilon \Lambda_{1,\delta''}}{\ln p_{\max}} \right\rceil, \quad N_1 = \left\lceil \frac{\ln(1-\delta')}{\ln p_{\max}} \right\rceil, \quad s = \underbrace{(i, i, \dots, i)}_{N \text{ times}}, \quad t = \underbrace{(i, i, \dots, i)}_{N_1 \text{ times}}.$$

Note  $s \in \mathcal{S}_{\varepsilon,\delta}$ , and so by Corollary 4.2 we note

$$\|T_*^{N-N_1} \mathbf{I}_s - \mathbf{I}_t\|_{L^1} \leq \|T_*^{N-N_1} \mathbf{I}_s - U_*^{N-N_1} \mathbf{I}_s\|_{L^1} \leq \delta''.$$

Also, by choice of  $N_1$  and  $t$ ,

$$\|\mathbf{I}_t - 1\|_{L^1} \geq 1 - p_{\max}^{|t|} \geq \delta'.$$

Thus, by the triangle inequality

$$\|T_*^{N-N_1} \mathbf{I}_s - 1\|_{L^1} \geq \|\mathbf{I}_t - 1\|_{L^1} - \|T_*^{N-N_1} \mathbf{I}_s - \mathbf{I}_t\|_{L^1} \geq \delta' - \delta'' = 2\delta.$$

Consequently, if we choose  $X_0$  to have density  $\mathbf{I}_s$ , then

$$(4.12) \quad \|\text{dist}(X_{N-N_1}) - \pi\|_{\text{TV}} = \frac{1}{2} \|T_*^{N-N_1} \mathbf{I}_s - 1\|_{L^1} \geq \delta$$

We clarify that the factor 1/2 above arises from the commonly used normalization convention

$$(4.13) \quad \|\mu - \nu\|_{\text{TV}} \stackrel{\text{def}}{=} \sup_{A \subseteq \mathbb{T}^d} \frac{|\mu(A) - \nu(A)|}{2}.$$

The lower bound (4.12) immediately implies  $t_{\text{mix}}(\delta) \geq N - N_1$  and the choice of  $N, N_1$  implies (1.6) as desired.  $\square$

**4.2. Mixing piecewise constant functions (Lemma 4.1).** We begin by estimating the mass that leaks out of cylinder sets due to the action of the noise.

**Lemma 4.3** (Convolution estimates). *Suppose  $K_\varepsilon$  satisfies (1.5). For any  $p \in [1, \infty)$ ,  $\varepsilon > 0$ ,  $s \in \mathcal{T}$  we have*

$$\|\mathbf{I}_s * K_\varepsilon - \mathbf{I}_s\|_{L^p} \leq \frac{(\varepsilon \text{Per}(\mathcal{C}_s))^{1/p}}{\pi(\mathcal{C}_s)}.$$

*Remark.* If  $K_\varepsilon$  is supported in  $B(0, \varepsilon)$ , then certainly  $K_\varepsilon$  satisfies (1.5).

*Proof.* This is a standard result that is true in more generality (see for instance Lemma 3.24 in [AFP00]). For convenience we present a simple proof here. Note that

$$\int_{x \in \mathbb{T}^d} |\mathbf{1}_{\mathcal{C}_s}(x-y) - \mathbf{1}_{\mathcal{C}_s}(x)|^p dx \leq d(0, y) \text{Per}(\mathcal{C}_s),$$

where  $d(0, y)$  denotes the torus distance between  $y$  and 0. Thus, by Jensen's inequality and (1.5) we see

$$\begin{aligned} \|\mathbf{1}_{\mathcal{C}_s} * K_\varepsilon - \mathbf{1}_{\mathcal{C}_s}\|_{L^p}^p &\leq \int_{y \in B(0, \varepsilon)} \int_{x \in \mathbb{T}^d} K_\varepsilon(y) |\mathbf{1}_{\mathcal{C}_s}(x-y) - \mathbf{1}_{\mathcal{C}_s}(x)|^p dx dy \\ &\leq \int_{y \in B(0, \varepsilon)} d(0, y) K_\varepsilon(y) \text{Per}(\mathcal{C}_s) dy = \varepsilon \text{Per}(\mathcal{C}_s). \quad \square \end{aligned}$$

Using the convolution estimates above, we can now prove the desired  $L^p$  estimate.

*Proof of Lemma 4.1.* Using (4.1), we note that if  $f_0$  is given by (4.4), we have

$$U_* f_0 = U_* \sum_{s \in \mathcal{S}} c_0(s) \mathbf{I}_s \pi(\mathcal{C}_s) = \sum_{s \in \mathcal{S}} c_0(s) \mathbf{I}_{\sigma s} \pi(\mathcal{C}_s).$$

This implies

$$f_1 \stackrel{\text{def}}{=} T_* f_0 = \sum_{s \in \mathcal{I}} c_0(s) \mathbf{I}_{\sigma s} \pi(\mathcal{C}_s) + \tilde{f}_1 = U_* f_0 + \tilde{f}_1,$$

where

$$\tilde{f}_1 \stackrel{\text{def}}{=} T_* f_0 - U_* f_0 = \sum_{s \in \mathcal{I}} c_0(s) (\mathbf{I}_{\sigma s} * K_\varepsilon - \mathbf{I}_{\sigma s}) \pi(\mathcal{C}_s).$$

For any  $t \in \mathcal{I}$ , we partition  $\mathcal{S}$  into the sets  $\{\mathcal{S}^t \mid t \in \mathcal{I}\}$ , where

$$\mathcal{S}^t \stackrel{\text{def}}{=} \{s \in \mathcal{S} \mid \mathcal{C}_s \subseteq \mathcal{C}_t\}.$$

Grouping the error terms associated to each  $\mathcal{S}^t$ , we can write

$$(4.14) \quad \tilde{f}_1 = \sum_{t \in \mathcal{I}} \tilde{f}_{1,t} \quad \text{where} \quad \tilde{f}_{1,t} \stackrel{\text{def}}{=} \sum_{s \in \mathcal{S}^t} c_0(s) (\mathbf{I}_{\sigma s} * K_\varepsilon - \mathbf{I}_{\sigma s}) \pi(\mathcal{C}_s).$$

We now claim

$$(4.15) \quad \|\tilde{f}_{1,t}\|_{L^p}^p \leq C_1^{p/p'} \sum_{s \in \mathcal{S}^t} |c_0(s)|^p \|\mathbf{I}_{\sigma s} * K_\varepsilon - \mathbf{I}_{\sigma s}\|_{L^p}^p \pi(\mathcal{C}_s)^p,$$

where  $C_1$  is the constant in (4.6). If  $p = 1$ , then  $C_1^{p/p'} = 1$ , and so clearly (4.15) holds.

If  $p > 1$ , we recall that  $\text{supp}(K_\varepsilon) \subseteq B(0, \varepsilon)$  by assumption. We will use this to show the support of each term in (4.14) intersects the support of at most  $C_1$  other terms. This will imply (4.15) as claimed.

To count the number of terms in (4.14) with intersecting support, note that  $\{\mathcal{C}_s \mid s \in \mathcal{S}^t\}$  partitions  $\mathcal{S}^t$ . Hence the sets  $\{\mathcal{C}_{\sigma s} \mid s \in \mathcal{S}^t\}$  partition the torus. We claim now that for any  $s \in \mathcal{S}^t$ , the number of  $s' \in \mathcal{S}^t$  such that  $B(\mathcal{C}_{\sigma s}, \varepsilon) \cap B(\mathcal{C}_{\sigma s'}, \varepsilon) \neq \emptyset$  is at most the constant  $C_1$  in equation (4.6). To see this, we note that (4.3) and the fact that  $L \geq 1$  imply  $\text{diam}(\mathcal{C}_{\sigma s'}) \geq 2\varepsilon$  for every  $s' \in \mathcal{S}^t$ . Moreover every cylinder set is an axis-aligned cube, and the sets  $\{\mathcal{C}_{\sigma s'} \mid s' \in \mathcal{S}^t\}$  partition the torus. Consequently, if for some  $s, s' \in \mathcal{S}^t$ , the sets  $B(\mathcal{C}_{\sigma s}, \varepsilon)$  and  $B(\mathcal{C}_{\sigma s'}, \varepsilon)$  intersect, the closures of  $\mathcal{C}_{\sigma s}$  and  $\mathcal{C}_{\sigma s'}$  themselves must intersect.

Given (4.3) and the fact that  $\text{diam}(\mathcal{C}_{s'}) \geq p_{\min}^{1/d} \text{diam}(\mathcal{C}_{\sigma s'})$ , the number of disjoint sets intersecting a given face of  $\mathcal{C}_{\sigma s}$  is at most  $(2 + p_{\min}^{-1/d})^{d-1}$ . Since there are  $2d$  faces, there are at most  $2d(2 + p_{\min}^{-1/d})^{d-1} = C_1$  cylinder sets  $\mathcal{C}_{\sigma s'}$  such that the neighborhoods  $B(\mathcal{C}_{\sigma s}, \varepsilon)$  and  $B(\mathcal{C}_{\sigma s'}, \varepsilon)$  intersect.

Since  $\text{supp}(K_\varepsilon) \subseteq B(0, \varepsilon)$  by assumption,  $\text{supp}(\mathbf{I}_{\sigma s} * K_\varepsilon) \subseteq B(\mathcal{C}_{\sigma s}, \varepsilon)$ , and hence the support of each term in (4.14) intersects the support of at most  $C_1$  other terms. This shows (4.15) holds even when  $p > 1$ .

Using (4.15) with Lemma 4.3, we have

$$\begin{aligned} \|\tilde{f}_{1,t}\|_{L^p}^p &\leq C_1^{p/p'} \sum_{s \in \mathcal{S}^t} |c_0(s)|^p \|\mathbf{I}_{\sigma s} * K_\varepsilon - \mathbf{I}_{\sigma s}\|_{L^p}^p \pi(\mathcal{C}_s)^p \\ &\leq \varepsilon C_1^{p/p'} H(\sigma(\mathcal{S}^t)) \sum_{s \in \mathcal{S}^t} |c_0(s)|^p \pi(\mathcal{C}_{\sigma s})^{-p/p'} \pi(\mathcal{C}_s)^p \\ &\leq \varepsilon C_1^{p/p'} H(\sigma(\mathcal{S}^t)) \pi(\mathcal{C}_t)^{p/p'} \sum_{s \in \mathcal{S}^t} |c_0(s)|^p \pi(\mathcal{C}_s), \\ &\leq \varepsilon C_1^{p/p'} H(\sigma(\mathcal{S})) \pi(\mathcal{C}_t)^{p/p'} \sum_{s \in \mathcal{S}^t} |c_0(s)|^p \pi(\mathcal{C}_s). \end{aligned}$$

This implies

$$\sum_{t \in \mathcal{I}} \|\tilde{f}_{1,t}\|_p \leq \varepsilon^{1/p} C_1^{1/p'} H(\sigma(\mathcal{S}))^{1/p} \sum_{t \in \mathcal{I}} \pi(\mathcal{C}_t)^{1/p'} \left( \sum_{s \in \mathcal{S}^t} |c_0(s)|^p \pi(\mathcal{C}_s) \right)^{1/p}$$

$$(4.16) \quad \leq \varepsilon^{1/p} C_1^{1/p'} H(\sigma(\mathcal{S}))^{1/p} \left( \sum_{s \in \mathcal{S}} |c_0(s)|^p \pi(\mathcal{C}_s) \right)^{1/p}.$$

Since

$$\|f_0\|_{L^p}^p = \sum_{s \in \mathcal{S}} |c_0(s)|^p \|\mathbf{1}_{\mathcal{C}_s}\|_{L^p}^p = \sum_{s \in \mathcal{S}} |c_0(s)|^p \pi(\mathcal{C}_s),$$

the inequality (4.16) implies

$$\|\tilde{f}_1\|_{L^p} \leq \varepsilon^{1/p} C_1^{1/p'} H(\sigma(\mathcal{S}))^{1/p} \|f_0\|_{L^p}.$$

Now for  $n \geq 2$  we inductively define the error term  $\tilde{f}_n$  via the identity

$$f_n = T_* f_{n-1} = U_*^n f_0 + \sum_{i=1}^{n-1} T_*^i \tilde{f}_{n-i} + \tilde{f}_n$$

and write

$$\tilde{f}_n \stackrel{\text{def}}{=} \sum_{t \in \mathcal{I}^n} \tilde{f}_{n,t} \quad \text{where} \quad \tilde{f}_{n,t} = \sum_{s \in \mathcal{S}_t} c_0(s) (\mathbf{I}_{\sigma^n s} * K_\varepsilon - \mathbf{I}_{\sigma^n s}) \pi(\mathcal{C}_s).$$

Using the same argument as above, we see

$$\|\tilde{f}_n\|_{L^p} \leq \varepsilon^{1/p} C_1^{1/p'} H(\sigma^n \mathcal{S})^{1/p} \|f_0\|_{L^p},$$

for all  $1 \leq n < \max_{s \in \mathcal{S}} |s|$ .

Since  $T_*$  is an  $L^p$  contraction, we have

$$\|T_*^N f_0 - U_*^N f_0\|_{L^p} \leq \sum_{i=1}^N \|\tilde{f}_i\|_{L^p} \leq \varepsilon^{1/p} C_1^{1/p'} \sum_{n=1}^N H(\sigma^n \mathcal{S})^{1/p} \|f_0\|_{L^p}.$$

for all  $N < \max_{s \in \mathcal{S}} |s|$ . This proves (4.5) for all  $N < \max_{s \in \mathcal{S}} |s|$ .

For  $n \geq \max_{s \in \mathcal{S}} |s|$  we note

$$\mathbf{I}_{\sigma^n s} * K_\varepsilon - \mathbf{I}_{\sigma^n s} = \mathbf{I}_0 * K_\varepsilon - \mathbf{I}_0 = 0,$$

and hence  $\tilde{f}_n = 0$ . Since  $H(\mathbf{0}) = 0$  by convention, this implies (4.5) for all  $n \in \mathbb{N}$ .  $\square$

## 5. Relation between the dissipation time and mixing time.

The upper bound for the dissipation time (1.8) is a general fact and requires nothing but the Markov property. The proof is very similar to the proof of Proposition 2.2 in [IZ22]. Since the proof is short and elementary, we present it here for the readers convenience.

*Proof of the upper bound (1.7) in Proposition 1.3.* Let  $\rho_n(x, y)$  be the transition density of the process  $X$  after  $n$  time steps, and  $\theta_0 \in \dot{L}^2$ . Then we note

$$\theta_n(x) \stackrel{\text{def}}{=} \mathbf{E}^x \theta_0(X_n) = \int_{\mathbb{T}^d} \rho_n(x, y) \theta_0(y) \pi(dy).$$

Since  $\theta_0$  has mean 0, this implies

$$\theta_n(x) = \int_{\mathbb{T}^d} (\rho_n(x, y) - 1) \theta_0(y) \pi(dy),$$

and hence

$$\begin{aligned} \|\theta_n\|_{L^2}^2 &\leq \left( \int_{\mathbb{T}^d \times \mathbb{T}^d} |\rho_n(x, y) - 1| \pi(dy) \pi(dx) \right) \\ &\quad \cdot \left( \int_{\mathbb{T}^d \times \mathbb{T}^d} \theta_0(y)^2 (\rho_n(x, y) + 1) \pi(dx) \pi(dy) \right) \\ &\leq 2 \|\theta_0\|_{L^2}^2 \sup_{x \in \mathbb{T}^d} \int_{\mathbb{T}^d} |\rho_n(x, y) - 1| dy. \end{aligned}$$

The last inequality above followed from the fact that  $\rho_n$  is nonnegative, and leaves the measure  $\pi$  invariant. Notice that the right hand side is at most  $4 \|\theta_0\|_{L^2}^2 \|\text{dist}(X_n) - \pi\|_{\text{TV}}$ , where we recall the total variation norm is normalized by a factor of  $\frac{1}{2}$  as in (4.13). Thus, when  $n \geq t_{\text{mix}}(\delta^2/4)$ , we have  $\|\theta_n\|_{L^2}^2 \leq \delta^2 \|\theta_0\|_{L^2}^2$ , proving (1.7) as claimed.  $\square$

In order to prove the lower bound for the dissipation time (1.8), it is convenient to introduce the ‘‘backward’’ operator  $T$ , defined by

$$(5.1) \quad Tf(x) = \mathbf{E}^x f(X_1) = \int_{\mathbb{T}^d} \rho(x, y) f(y) \pi(dy),$$

where we recall  $\rho$  is the transition kernel of the process  $X$ . From (3.1) we see that for any  $f, g \in L^2$  we have

$$(5.2) \quad \langle Tf, g \rangle = \int_{\mathbb{T}^d \times \mathbb{T}^d} \rho(x, y) f(y) g(x) \pi(dy) \pi(dx) = \langle f, T_* g \rangle,$$

where  $\langle \cdot, \cdot \rangle$  denotes the standard  $L^2$  inner-product.

Moreover, if  $\rho_n$  is the  $n$ -step transition kernel of  $X$ , then by the Markov property and induction we see

$$\begin{aligned} T^n f(x) &= T^{n-1} T f(x) = \int_{\mathbb{T}^d \times \mathbb{T}^d} \rho_{n-1}(x, y) \rho(y, z) f(z) \pi(dz) \pi(dy) \\ &= \int_{\mathbb{T}^d} \rho_n(x, z) f(z) \pi(dz) = \mathbf{E}^x f(X_n). \end{aligned}$$

In light of this, we may rewrite the dissipation time  $t_{\text{dis}}(\delta)$  as

$$t_{\text{dis}}(\delta) = \min\{n \mid \|T^n f\|_{L^2} \leq \delta \|f\|_{L^2} \text{ for all } f \in \dot{L}^2\}.$$

We claim  $T$  can be replaced with  $T_*$  in the above definition. That is, we claim

$$(5.3) \quad t_{\text{dis}}(\delta) = \min\{n \mid \|T_*^n f\|_{L^2} \leq \delta \|f\|_{L^2} \text{ for all } f \in \dot{L}^2\}.$$

To see this let  $m = t_{\text{dis}}^*(\delta)$  be the right hand side of (5.3). Using (5.2) we see

$$\|T^m f\|_{L^2}^2 = \langle T^m f, T^m f \rangle = \langle f, T_*^m T^m f \rangle \leq \|f\|_{L^2} \|T_*^m T^m f\|_{L^2} \leq \delta \|f\|_{L^2} \|T^m f\|_{L^2},$$

and hence

$$\|T^m f\|_{L^2} \leq \delta \|f\|_{L^2}.$$

This implies  $t_{\text{dis}}(\delta) \leq m = t_{\text{dis}}^*(\delta)$ . The reverse inequality follows by symmetry, showing (5.3).

We will now prove the lower bound (1.8) by using the identity (5.3)

*Proof of the lower bound (1.8) in Proposition 1.3.* Since  $K_\varepsilon \in L^1$ , we note that

$$\mathbf{P}(X_1 \in dx) = f_1(x)\pi(dx),$$

for some probability density function  $f_1 \in L^1$ . By Young's inequality,

$$\|T_*(f_1 - 1)\|_{L^2} \leq \|K_\varepsilon\|_{L^2} \|U_*(f_1 - 1)\|_{L^1} \leq \frac{2\mathbf{K}}{\varepsilon^{d/2}}.$$

Thus

$$(5.4) \quad \|\text{dist}(X_{n+2}) - \pi\|_{\text{TV}} = \frac{\|T_*^{n+1}f_1 - 1\|_{L^1}}{2} \leq \frac{\|T_*^{n+1}(f_1 - 1)\|_{L^2}}{2}.$$

Notice for any  $g \in \dot{L}^2$  and  $n = mt_{\text{dis}}(\delta)$  we have

$$\|T_*^n(g)\|_{L^2} \leq \delta^m \|g\|_{L^2}.$$

Thus choosing

$$n \geq \log_\delta \left( \frac{\delta' \varepsilon^{d/2}}{\mathbf{K}} \right) t_{\text{dis}}(\delta)$$

the right hand side of (5.4) is at most  $\delta'$ . This implies (1.8), concluding the proof.  $\square$

## 6. Dissipation Time Bounds.

**6.1. The upper bound on the dissipation time.** Throughout this section we will make the same assumptions as Theorem 1.5. That is, we will assume Assumption 2.1 holds, and  $\text{supp}(K_\varepsilon) \subseteq B(0, \varepsilon)$ .

The main idea behind the proof is to use the noise to dissipate high frequencies, and the dynamics of  $\varphi$  to mix the low frequencies. While dissipating high frequencies with the noise is straightforward, showing that the low frequency data gets mixed requires a little more care. We state this as our first lemma.

**Lemma 6.1.** *Let  $p \in [1, \infty)$ ,  $\delta \in (0, 1)$ . There exists an explicit constant  $B_{p,\delta}$  such that if  $f \in \dot{L}^p$  and*

$$(6.1) \quad \|\nabla f\|_{L^p} \leq \frac{B_{p,\delta}}{\varepsilon} \|f\|_{L^p},$$

then

$$\|T_*^n f\|_{L^p} \leq \frac{\delta}{2} \|f\|_{L^p}, \quad \text{for all } n \geq \frac{d \ln(\varepsilon \Lambda_{p,\delta/4})}{\ln p_{\max}}.$$

*Remark 6.2.* Let  $X_0 \sim \mu$ ,  $f_1 = K_\varepsilon * U_* \mu$ , and note  $\|\nabla f_1 - 1\|_{L^1} \leq 2\|\nabla K_\varepsilon\|_{L^1}$ . If

$$(6.2) \quad 2\|\nabla K_\varepsilon\|_{L^1} < B_{p,\delta},$$

it is possible to use Lemma 6.1 to show that for some  $\delta \in (0, 1)$  we have the expected sharp upper bound stated in (1.4). However, the constant  $B_{p,\delta}$  is related to the Poincaré constant of cylinder sets and can be computed explicitly. Even the indicator function of a ball does not satisfy (6.2), and it may not be possible to find even one kernel (that satisfies (1.5)) for which (6.2) holds.

Here  $\Lambda_{p,\delta/4}$  is defined by (4.8) (in Corollary 4.2). Postponing the proof of Lemma 6.1, we prove the upper bound in Theorem 1.5. While Lemma 6.1 holds for any  $p \in [1, \infty)$ , we are presently only able to apply it when  $p = 2$  as our proof of Theorem 1.5 relies on orthogonal projections in frequency space.

*Proof of the upper bound in Theorem 1.5.* Define the Fourier projections  $\mathcal{P}_L$  and  $\mathcal{P}_H$ , which project functions onto the low and high frequency spaces respectively. Explicitly, define

$$(\mathcal{P}_L f)^\wedge(k) = \sum_{|k| \leq \frac{B}{2\pi\varepsilon}} \hat{f}(k) \quad \text{and} \quad \mathcal{P}_H = I - \mathcal{P}_L,$$

where  $B = B_{2,\delta}$  is the constant appearing in (6.1), and  $\hat{f}(k)$  is the  $k$ -th Fourier coefficient of  $f$ .

The reason for defining  $\mathcal{P}_L$  and  $\mathcal{P}_H$  as above is that functions in the range of  $\mathcal{P}_L$  have well controlled gradients, and can be mixed using Lemma 6.1. On the other hand, functions in the range of  $\mathcal{P}_H$  only have high frequencies and are rapidly mixed by  $\mathcal{P}_H$ . More precisely, for any  $g \in \dot{L}^2$ , we have

$$\|\nabla \mathcal{P}_L g\|_{L^2} \leq \frac{B}{\varepsilon} \|\mathcal{P}_L g\|_{L^2},$$

and hence Lemma 6.1 can be applied to  $\mathcal{P}_L g$ .

For high frequencies,

$$\|K_\varepsilon * \mathcal{P}_H g\|_{L^2} = \left( \sum_{|k| \geq \frac{B}{2\pi\varepsilon}} |\hat{K}_\varepsilon|^2 (\mathcal{P}_H g)^\wedge(k)^2 \right)^{1/2} \leq (1 - \chi) \|\mathcal{P}_H g\|_{L^2},$$

where  $\chi = \chi(\delta)$  is defined by

$$\chi = \sup_{|k| \geq \frac{B}{2\pi\varepsilon}} |\hat{K}_\varepsilon(k)|.$$

Recall that assumption (1.11) guarantees  $\chi < 1$ .

We will now show

$$(6.3) \quad t_{\text{dis}}(\delta) \leq N + N_1 + 1$$

where

$$N = \left\lceil \frac{d \ln(\varepsilon \Lambda_{2,\delta/4})}{\ln p_{\max}} \right\rceil, \quad \text{and} \quad N_1 = \left\lceil \frac{\ln \delta}{\ln \left( 1 - \frac{\delta^2(1-(1-\chi)^2)}{4(1-\chi)^2} \right)} \right\rceil,$$

provided

$$\delta^2 \leq \frac{2(1-\chi)^2}{1-(1-\chi)^2}.$$

We divide the proof into two cases.

*Case I:* Suppose for some  $n \leq N_1$  we have

$$(6.4) \quad \|\mathcal{P}_H U_* f_n\|_{L^2} \leq \frac{\delta}{2(1-\chi)} \|f_n\|_{L^2}.$$

Let  $\mathcal{K}_\varepsilon g = K_\varepsilon * g$  denote the convolution operator, and observe

$$\begin{aligned} \|T_*^{N+1} f_n\|_{L^2} &\leq \|T_*^N \mathcal{K}_\varepsilon \mathcal{P}_L U_* f_n\|_{L^2} + \|T_*^N \mathcal{K}_\varepsilon \mathcal{P}_H U_* f_n\|_{L^2} \\ &\leq \frac{\delta}{2} \|\mathcal{P}_L U_* f_n\|_{L^2} + (1-\chi) \|\mathcal{P}_H U_* f_n\|_{L^2} \leq \delta \|f_n\|. \end{aligned}$$

This implies

$$(6.5) \quad \|f_{N+N_1+1}\|_{L^2} \leq \delta \|f_0\|_{L^2}.$$

*Case II:* Suppose now (6.4) does not hold for any  $n \leq N_1$ . In this case we must have

$$\begin{aligned} \|f_{n+1}\|_{L^2}^2 &= \|\mathcal{K}_\varepsilon \mathcal{P}_H U_* f_n\|_{L^2}^2 + \|\mathcal{K}_\varepsilon \mathcal{P}_L U_* f_n\|_{L^2}^2 \\ &\leq (1-\chi)^2 \|\mathcal{P}_H U_* f_n\|_{L^2}^2 + \|U_* f_n\|_{L^2}^2 - \|\mathcal{P}_H U_* f_n\|_{L^2}^2 \\ &\leq \left(1 - \frac{\delta^2(1-(1-\chi)^2)}{4(1-\chi)^2}\right) \|f_n\|. \end{aligned}$$

Iterating this and using the definition of  $N_1$  immediately implies

$$\|f_{N_1}\|_{L^2} \leq \delta \|f_0\|_{L^2},$$

and hence (6.5) also holds in this case.

Thus in either case (6.5) holds, and using (5.3) this implies (6.3) as claimed. By choice of  $N$  and  $N_1$  this proves the upper bound in (1.12) as claimed.  $\square$

**6.2. Mixing low frequency data (Lemma 6.1).** Note if (6.1) holds, then we can approximate  $f$  by a function that is piecewise constant on small cylinder sets. For such functions we may use Lemma 4.1 (or Corollary 4.2) to show that  $T_*^n f$  is close to  $U_*^n f$ . Of course, if  $f \in \dot{L}^2$  is in the form (4.4), and  $N \geq \max_{s \in \mathcal{S}} |s|$ , then

$$U_*^N f = \sum_{s \in \mathcal{S}} c_0(s) \pi(\mathcal{C}_s) = \int_{\mathbb{T}^d} f d\pi = 0.$$

so the bound (4.10) in Corollary 4.2 will ensure  $\|T_*^N f_0\|_{L^2}$  is small. This is the strategy we will use to prove Lemma 6.1.

*Proof of Lemma 6.1.* Let  $\Lambda = \Lambda_{p,\delta/4}$  be defined by (4.8), and  $\mathcal{S} = \mathcal{S}_{\varepsilon,\delta}$  be as in (4.9). Recall, for every  $s \in \mathcal{S}$  the side lengths  $\ell_s$  are bounded by (4.11). Moreover since  $p_{\min}^{1/d} \ell_{\sigma s} \leq \ell_s \leq p_{\max}^{1/d} \ell_{\sigma s}$  we must have

$$(6.6) \quad \frac{d \ln(\varepsilon \Lambda_{p,\delta/4})}{\ln p_{\min}} \leq |s| \leq \frac{d \ln(\varepsilon \Lambda_{p,\delta/4})}{\ln p_{\max}} \quad \text{for all } s \in \mathcal{S}.$$

Choose

$$N \stackrel{\text{def}}{=} \max_{s \in \mathcal{S}} |s| \leq \frac{d \ln(\varepsilon \Lambda_{p,\delta/4})}{\ln p_{\max}}.$$

Now let  $B_{p,\delta}$  be a constant that will be chosen shortly, and suppose  $f \in \dot{L}^p$  satisfies (6.1). Define

$$c_s = \frac{1}{\pi(\mathcal{C}_s)} \int_{\mathcal{C}_s} f d\pi, \quad f_0 = \sum_{s \in \mathcal{S}} c_s \mathbf{1}_{\mathcal{C}_s}, \quad \tilde{f}_0 = f - f_0.$$

Since  $U_*^N f_0 = 0$  (by choice of  $N$ ), Corollary 4.2 implies

$$\|T_*^N f_0\|_{L^p} = \|T_*^N f_0 - U_*^N f_0\|_{L^p} \leq \frac{\delta}{4} \|f_0\|_{L^p} \leq \frac{\delta}{4} \|f\|_{L^p}.$$

To bound  $T_*^N \tilde{f}_0$  observe that the Poincaré inequality,

$$\begin{aligned} \|\tilde{f}_0\|_{L^p}^p &= \left\| f - \sum_{s \in \mathcal{S}} c_s \mathbf{1}_{\mathcal{C}_s} \right\|_{L^p}^p = \sum_{s \in \mathcal{S}} \|f - c_s\|_{L^p(\mathcal{C}_s)}^p \\ &\leq \text{diam}(\mathcal{C}_s)^p \sum_{s \in \mathcal{S}} \|\nabla f\|_{L^p(\mathcal{C}_s)}^p \leq (\varepsilon \Lambda_{p,\delta/4} \sqrt{d})^p \|\nabla f\|_{L^p}^p. \end{aligned}$$

Thus if we choose

$$B_{p,\delta} \stackrel{\text{def}}{=} \frac{\delta}{4 \Lambda_{p,\delta/4} \sqrt{d}},$$

the assumption (6.1) will imply

$$\|T_*^N \tilde{f}_0\|_{L^p} \leq \|\tilde{f}_0\|_{L^p} \leq \frac{\delta}{4} \|f\|_{L^p}.$$

Consequently, for any  $n \geq N$ ,

$$\|T_*^n f\|_{L^p} \leq \|T_*^N f\|_{L^p} \leq \|T_*^N f_0\|_{L^p} + \|T_*^N \tilde{f}_0\|_{L^p} \leq \frac{\delta}{2},$$

concluding the proof.  $\square$

**6.3. Lower bounds on the dissipation time.** The strategy to prove the lower bound in Theorem 1.5 is similar to the strategy used to prove Theorem 1.2. The main difference is that for Theorem 1.2 we could choose our initial distribution to be uniformly distributed on a small cylinder set. We can not do this for Theorem 1.5. If we do, one iteration of  $T_*$  will spread it on a larger cylinder set, and decreases the  $L^p$  norm by a constant factor for every  $p \in (1, \infty)$ . Thus our initial distribution has to be spread over many small cylinder sets. This now requires us to control the smallest cylinder set our initial data is supported on, and results in our bounds being of order  $d |\ln \varepsilon| / p_{\min}$ , and not  $d |\ln \varepsilon| / p_{\max}$  as was the case with Theorem 1.1.

*Proof of the lower bounds in Theorem 1.5.* Let  $\delta' \in (0, 1)$  be a small number that will be chosen shortly, and  $\mathcal{S}$  be the same partition that was used in the proof of Lemma 6.1 with  $\Lambda = \Lambda_{2,\delta'}$ , and set  $N = \min_{s \in \mathcal{S}} |s|$ . By (6.6), we know that

$$N \geq \frac{d \ln(\Lambda \varepsilon)}{\ln p_{\min}} = \frac{d \ln(\varepsilon)}{\ln p_{\min}} + \frac{d \ln(\Lambda)}{\ln p_{\min}} = \frac{d \ln(\varepsilon)}{\ln p_{\min}} - C(1 - \ln \delta'),$$

for some explicit constant  $C > 0$  that can be computed from (4.8).

Consider the function given by

$$f_0(x) = \begin{cases} 1 & \varphi^N(x) \in E_1 \\ p_1/p_2 & \varphi^N(x) \in E_2 \\ 0 & \text{otherwise} \end{cases},$$

which is of the form (4.4) by our choice of  $N$ . Note that

$$U_*^{N-1} f_0(x) = \begin{cases} 1 & \varphi(x) \in E_1 \\ p_1/p_2 & \varphi(x) \in E_2 \\ 0 & \text{otherwise} \end{cases}.$$

Thus  $f_0$  and  $U_*^{N-1} f_0$  have the same distribution function, and hence must have the same  $L^2$  norm. Applying Corollary 4.2, we immediately see that

$$\|T_*^N f_0\|_{L^2} \geq \|U_*^N f_0\|_{L^2} - \|T_*^N f_0 - U_*^N f_0\|_{L^2} \geq (1 - \delta') \|f_0\|_{L^2} = \delta \|f_0\|_{L^2},$$

provided  $\delta' = 1 - \delta \in (0, 1)$ . Using (5.3), this implies

$$t_{\text{dis}}(\delta) \geq N.$$

proving (1.12) as desired.  $\square$

## Appendix A. Double exponential convergence when $\varphi$ is uniformly expanding, or an ergodic toral automorphism.

Throughout this appendix we will assume  $K_\varepsilon$  is the periodized rescaled standard Gaussian. That is, we assume  $\check{K}$  is a standard Gaussian and  $K_\varepsilon$  is defined by (2.4). At the expense of a more technical proof, similar results can be obtained using only spectral assumptions on  $K_\varepsilon$ ; we refer to [FNW04] for details.

**Proposition A.1.** *Let  $N \in \{2, 3, \dots\}$ , and suppose  $\varphi: \mathbb{T}^d \rightarrow \mathbb{T}^d$  is the uniformly expanding map*

$$(A.1) \quad \varphi(x) = Nx \pmod{\mathbb{Z}^d}.$$

*Then there exists a constant  $C$  such that for all  $n \in \mathbb{N}$  and  $\varepsilon > 0$  sufficiently small we have*

$$(A.2) \quad \|\text{dist}(X_{n+1}) - \pi\|_{\text{TV}} \leq \frac{C}{\varepsilon^{d/2}} \exp\left(-\frac{\varepsilon^2 N^{2n}}{C}\right)$$

*Moreover, if  $T_*$  and  $T$  are the evolution operators defined in (3.1) and (5.1) respectively, then*

$$(A.3) \quad \|T^n\|_{\dot{L}^2 \rightarrow \dot{L}^2} = \|T_*^n\|_{\dot{L}^2 \rightarrow \dot{L}^2} \leq \exp\left(-\frac{\varepsilon^2 N^{2n}}{C}\right),$$

As an immediate corollary, we obtain bounds on the mixing time and dissipation time.

**Corollary A.2.** *For the uniformly expanding map (A.1), there exists an explicit constant  $C = C(N, d)$  such that the mixing time and dissipation time are bounded by*

$$(A.4) \quad t_{\text{mix}}(\delta) \leq |\log_N \varepsilon| + \frac{1}{2} \log_N \left( \frac{d}{2} |\ln \varepsilon| + |\ln \delta| + C \right) + C$$

$$(A.5) \quad t_{\text{dis}}(\delta) \leq |\log_N \varepsilon| + \frac{1}{2} \log_N |\ln \delta| + C$$

*Remark A.3.* The uniformly expanding map (A.1) is of the form considered in Section 2.1. Indeed, for the uniformly expanding map,  $M = N^d$ , and each  $E_i$  is a cube of side length  $1/N$  and  $p_{\text{max}} = p_{\text{min}} = 1/N^d$ . From this we see (A.4) attains the conjectured upper bound (1.4) (with an explicit  $\delta$  dependence), and (A.5) improves the  $\delta$  dependence in (1.12) by replacing  $|\ln \delta|/\delta^2$  with a double logarithm.

*Remark A.4.* If  $\varphi$  is an ergodic toral automorphism (with matrix  $A$ ), then the double exponential bounds in Proposition A.1 still hold. The proof is contained in [FW03, FI19], and the main idea is as follows. One can use certain Diophantine approximation results to show that there exists  $\lambda > 1$  such that

$$|A_*^{-n} k| \geq \frac{\lambda^n}{C |k|^{d-1}}, \quad \text{for all } k \in \mathbb{Z}^d - \{0\}.$$

Here  $A_*$  is the transpose of  $A$ , and  $C$  is a dimensional constant. (The proof of this follows from Lemma 4.4 in [FI19], and is the lower bound of  $|B^n k|$  in the proof of Proposition 4.1.) Once this is established, we can follow the proof of Proposition A.1. (The details are carried out in the proof of Theorem 2.12 in [FI19], and a slightly different proof is in [FW03].)

*Proof of Proposition A.1.* Note first for any function  $g$  we have

$$Tg = (K_\varepsilon * g) \circ \varphi,$$

from which the Fourier coefficients of  $Tg$  can be computed explicitly. Namely, if  $k' \notin N\mathbb{Z}^d$ , then  $(Tg)^\wedge(k') = 0$ , and otherwise

$$(A.6) \quad (Tg)^\wedge(Nk) = \hat{K}_\varepsilon(k) \hat{g}(k).$$

Since  $K_\varepsilon$  is a periodized rescaled Gaussian, we know

$$\hat{K}_\varepsilon(k) = e^{-2\pi^2 \varepsilon^2 |k|^2}.$$

Thus iterating (A.6) gives

$$(T^n g)^\wedge(N^n k) = \exp\left(-2\pi^2 \varepsilon^2 \left(\frac{N^{2n} - 1}{N - 1}\right) |k|^2\right) \hat{g}(k),$$

and  $(T^n g)^\wedge(k') = 0$  when  $k' \notin N^n \mathbb{Z}$ .

Now if  $g \in \dot{L}^2$ , then  $\hat{g}(0) = 0$  and Parseval's identity implies

$$\begin{aligned} \|T^n g\|_{L^2} &\leq \|g\|_{L^2} \sup_{k \neq 0} \exp\left(-2\pi^2 \varepsilon^2 \left(\frac{N^{2n} - 1}{N - 1}\right) |k|^2\right) \\ &\leq \|g\|_{L^2} \exp\left(-2\pi^2 \varepsilon^2 \left(\frac{N^{2n} - 1}{N - 1}\right)\right). \end{aligned}$$

Since  $\|T^n\|_{\dot{L}^2 \rightarrow \dot{L}^2} = \|T_*^n\|_{\dot{L}^2 \rightarrow \dot{L}^2}$ , this implies (A.3).

To prove (A.2), suppose  $X_0 \sim \mu$  and observe

$$\begin{aligned} \|\text{dist}(X_{n+1}) - \pi\|_{\text{TV}} &= \frac{\|T_*^{n+1} \mu - 1\|_{L^1}}{2} \leq \frac{\|T_*^{n+1} \mu - 1\|_{L^2}}{2} \\ &\leq \frac{\|T_*^n\|_{L^2 \rightarrow L^2} \|T_* \mu - 1\|_{L^2}}{2} \\ &\leq \|K_\varepsilon\|_{L^2} \exp\left(-\frac{\varepsilon^2 N^{2n}}{C}\right) \leq \frac{C}{\varepsilon^{d/2}} \exp\left(-\frac{\varepsilon^2 N^{2n}}{C}\right), \end{aligned}$$

concluding the proof.  $\square$

## Acknowledgements

We gratefully acknowledge helpful discussions with Albert Fannjiang.

## References

- [AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.
- [BBPS21] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith. Almost-sure enhanced dissipation and uniform-in-diffusivity exponential mixing for advection-diffusion by stochastic Navier-Stokes. *Probab. Theory Related Fields*, 179(3-4):777–834, 2021. doi:[10.1007/s00440-020-01010-8](https://doi.org/10.1007/s00440-020-01010-8).
- [BCZ17] J. Bedrossian and M. Coti Zelati. Enhanced dissipation, hypoellipticity, and anomalous small noise inviscid limits in shear flows. *Arch. Ration. Mech. Anal.*, 224(3):1161–1204, 2017. doi:[10.1007/s00205-017-1099-y](https://doi.org/10.1007/s00205-017-1099-y).
- [BHP17] R. Basu, J. Hermon, and Y. Peres. Characterization of cutoff for reversible Markov chains. *Ann. Probab.*, 45(3):1448–1487, 2017. doi:[10.1214/16-AOP1090](https://doi.org/10.1214/16-AOP1090).

- [CCZW21] M. Colombo, M. Coti Zelati, and K. Widmayer. Mixing and diffusion for rough shear flows. 2021. doi:[10.15781/83FC-J334](https://doi.org/10.15781/83FC-J334).
- [CD20] S. Chatterjee and P. Diaconis. Speeding up Markov chains with deterministic jumps. *Probab. Theory Related Fields*, 178(3-4):1193–1214, 2020. doi:[10.1007/s00440-020-01006-4](https://doi.org/10.1007/s00440-020-01006-4).
- [CKRZ08] P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatoš. Diffusion and mixing in fluid flow. *Ann. of Math. (2)*, 168(2):643–674, 2008. doi:[10.4007/annals.2008.168.643](https://doi.org/10.4007/annals.2008.168.643).
- [Cor51] S. Corrsin. On the spectrum of isotropic temperature fluctuations in an isotropic turbulence. *J. Appl. Phys.*, 22:469–473, 1951.
- [CZDE20] M. Coti Zelati, M. G. Delgadino, and T. M. Elgindi. On the relation between enhanced dissipation timescales and mixing rates. *Comm. Pure Appl. Math.*, 73(6):1205–1244, 2020. doi:[10.1002/cpa.21831](https://doi.org/10.1002/cpa.21831).
- [Dia96] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996. doi:[10.1073/pnas.93.4.1659](https://doi.org/10.1073/pnas.93.4.1659).
- [DKRH21] D. Dolgopyat, A. Kanigowski, and F. Rodriguez-Hertz. Exponential mixing implies bernoulli, 2021. doi:[10.48550/ARXIV.2106.03147](https://doi.org/10.48550/ARXIV.2106.03147).
- [EZ19] T. M. Elgindi and A. Zlatoš. Universal mixers in all dimensions. *Adv. Math.*, 356:106807, 33, 2019. doi:[10.1016/j.aim.2019.106807](https://doi.org/10.1016/j.aim.2019.106807).
- [FFIT20] Y. Feng, Y. Feng, G. Iyer, and J.-L. Thiffeault. Phase separation in the advective Cahn-Hilliard equation. *J. Nonlinear Sci.*, 30(6):2821–2845, 2020. doi:[10.1007/s00332-020-09637-6](https://doi.org/10.1007/s00332-020-09637-6).
- [FI19] Y. Feng and G. Iyer. Dissipation enhancement by mixing. *Nonlinearity*, 32(5):1810–1851, 2019. doi:[10.1088/1361-6544/ab0e56](https://doi.org/10.1088/1361-6544/ab0e56).
- [FM22] Y. Feng and A. L. Mazzucato. Global existence for the two-dimensional Kuramoto-Sivashinsky equation with advection. *Comm. Partial Differential Equations*, 47(2):279–306, 2022. doi:[10.1080/03605302.2021.1975131](https://doi.org/10.1080/03605302.2021.1975131).
- [FNW04] A. Fannjiang, S. Nonnenmacher, and L. Wołowski. Dissipation time and decay of correlations. *Nonlinearity*, 17(4):1481–1508, 2004. doi:[10.1088/0951-7715/17/4/018](https://doi.org/10.1088/0951-7715/17/4/018).
- [FNW06] A. Fannjiang, S. Nonnenmacher, and L. Wołowski. Relaxation time of quantized toral maps. *Ann. Henri Poincaré*, 7(1):161–198, 2006. doi:[10.1007/s00023-005-0246-4](https://doi.org/10.1007/s00023-005-0246-4).
- [FW03] A. Fannjiang and L. Wołowski. Noise induced dissipation in Lebesgue-measure preserving maps on  $d$ -dimensional torus. *J. Statist. Phys.*, 113(1-2):335–378, 2003. doi:[10.1023/A:1025787124437](https://doi.org/10.1023/A:1025787124437).
- [HV05] P. H. Haynes and J. Vanneste. What controls the decay of passive scalars in smooth flows? *Physics of Fluids*, 17(9):097103, 2005. doi:[10.1063/1.2033908](https://doi.org/10.1063/1.2033908).
- [IXZ21] G. Iyer, X. Xu, and A. Zlatoš. Convection-induced singularity suppression in the Keller-Segel and other non-linear PDEs. *Trans. Amer. Math. Soc.*, 374(9):6039–6058, 2021. doi:[10.1090/tran/8195](https://doi.org/10.1090/tran/8195).
- [IZ22] G. Iyer and H. Zhou. Quantifying the dissipation enhancement of cellular flows, 2022. doi:[10.48550/ARXIV.2209.11645](https://doi.org/10.48550/ARXIV.2209.11645).
- [KH95] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. doi:[10.1017/CBO9780511809187](https://doi.org/10.1017/CBO9780511809187). With a supplementary chapter by Katok and Leonardo Mendoza.
- [LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. doi:[10.1090/mbk/058](https://doi.org/10.1090/mbk/058). With a chapter by James G. Propp and David B. Wilson.
- [MD18] C. J. Miles and C. R. Doering. Diffusion-limited mixing by incompressible flows. *Nonlinearity*, 31(5):2346, 2018. doi:[10.1088/1361-6544/aab1c8](https://doi.org/10.1088/1361-6544/aab1c8).
- [Obu49] A. M. Obukhov. Structure of temperature field in turbulent flow. *Izv. Akad. Nauk. SSSR, Geogr. Geofiz.*, 13, 1949.
- [OTD21] B. W. Oakley, J.-L. Thiffeault, and C. R. Doering. On mix-norms and the rate of decay of correlations. *Nonlinearity*, 34(6):3762–3782, 2021. doi:[10.1088/1361-6544/abdbbd](https://doi.org/10.1088/1361-6544/abdbbd).
- [Pie94] R. T. Pierrehumbert. Tracer microstructure in the large-eddy dominated regime. *Chaos, Solitons & Fractals*, 4(6):1091–1110, 1994.
- [Sei22] C. Seis. Bounds on the rate of enhanced dissipation. *Comm. Math. Phys.*, 2022. doi:[10.1007/s00220-022-04588-3](https://doi.org/10.1007/s00220-022-04588-3).
- [SOW06] R. Sturman, J. M. Ottino, and S. Wiggins. *The mathematical foundations of mixing*, volume 22 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2006. doi:[10.1017/CBO9780511618116](https://doi.org/10.1017/CBO9780511618116). The linked twist map as a paradigm in applications: micro to macro, fluids to solids.
- [SS00] B. I. Shraiman and E. D. Siggia. Scalar turbulence. *Nature*, 405(6787):639, 2000. doi:[10.1038/35015000](https://doi.org/10.1038/35015000).
- [TC03] J.-L. Thiffeault and S. Childress. Chaotic mixing in a torus map. *Chaos*, 13(2):502–507, 2003. doi:[10.1063/1.1568833](https://doi.org/10.1063/1.1568833).
- [Thi12] J.-L. Thiffeault. Using multiscale norms to quantify mixing and transport. *Nonlinearity*, 25(2):R1–R44, 2012. doi:[10.1088/0951-7715/25/2/R1](https://doi.org/10.1088/0951-7715/25/2/R1).
- [Wei19] D. Wei. Diffusion and mixing in fluid flow via the resolvent estimate. *Science China Mathematics*, pages 1869–1862, 2019. doi:[10.1007/s11425-018-9461-8](https://doi.org/10.1007/s11425-018-9461-8).
- [Zla10] A. Zlatoš. Diffusion in fluid flow: dissipation enhancement by flows in 2D. *Comm. Partial Differential Equations*, 35(3):496–534, 2010. doi:[10.1080/03605300903362546](https://doi.org/10.1080/03605300903362546).

DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA 15213.

*Email address:* gautam@math.cmu.edu

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, STANFORD, CA 94305.

*Email address:* ethanlu@stanford.edu

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY, DURHAM, NC 27708.

*Email address:* nolen@math.duke.edu