Numerical Evidence of Cutoffs in Chaotic Mixing by the Standard Map

T.-C. Liang and M. West

February 26, 2011

Abstract

We numerically study the decay of the variance of a passive scalar function being mixed by the diffusive Standard Map. An efficient and parallelizable Markov Chain model is used to approximate the Perron-Frobenius and Koopman operators of the chaotic map with diffusion. The limit of zero diffusion is approximated by constructing a sequence of Markov Chains with ever higher resolution (up to 6.4×10^9 states). In the limit of diffusion going to zero we show that scalar variance exhibits a sharp decay and present numerical evidence to suggest that this is a cutoff in the sense of finite Markov Chains.

1 Introduction

The question of how chaotic advection mixes a passive scalar function has attracted much research effort in recent years [18]. The main issues in this field are: how to measure the thoroughness of the mixing, how the mixing process changes qualitatively and quantitatively when the diffusion is close to zero, and how to enhance the overall mixing process by designing the map which produces chaotic advection. Unfortunately, we have only partial understanding for most of these topics. In spite of the fact that the detailed mechanism of mixing is unclear, non-trivial mixing processes have been observed in experiments [21, 25] and can be simulated by large-scale computations [15, 24]. In this article we build a linear Markov Chain model to simulate chaotic mixing with small diffusion. Similar approaches for nonlinear dynamical systems can be found in, for example, [5, 6, 10, 11, 12] . This simple and parallelizable linear model not only captures the multi-stage feature of a chaotic mixing process, but also generates a series of finite Markov Chains through which we can observe the multi-stage feature of the mixing trajectory near the zero-diffusion limit.

The multi-stage feature of chaotic mixing processes. A widely observed phenomenon in the chaotic mixing process when small diffusion exists is the two or three-stage transition [23, 9, 1, 16]. The map does not mix the scalar function with a constant rate in general. When the variance of the scalar function is

measured during the mixing process, one can in general observe a relatively flat decay initially, followed by a super-exponential change, and then finally it tends to a exponential decay. We are interested in when these transitions happen, why they happen, and how to predict the slope of the exponential region. A good review and physical interpretation can be found in [22].

Thiffeault and Childress [23] study these properties for a modified Arnold's cat map. Analytical formulas are given to predict the transitions as well as the slopes. Because the linear part of this map has an eigenvalue approximately equal to 2.618, which stretches very fast, and the chaotic part is relatively small, the three phases are separated clearly. The same analytical procedure cannot be applied to, for example, the Standard Map, although the only difference between the Standard Map and the modified Arnold's cat map is in the linear part.

As for the exponential decay part, there is still debate about whether the decay rate goes to zero in the zero diffusivity limit or whether it tends to a constant independent of the diffusion [22, 24]. Theoretical analysis shows both of these possibilities can occur for different chaotic flows [13].

Difficulties typically arise in studying the above problems numerically, because the small diffusion usually means fine grids are required in the solution of the advection-diffusion equation or the simulation of the map. Some studies and numerical results conclude that a proportional relation exists between the stationary decay rate and the diffusion [3, 20]. However, this is only true for certain diffusion ranges.

In [24], the author uses a simple and parallelizable numerical strategy which can simulate up to 6×10^4 by 6×10^4 grids to show the decay rate of a certain chaotic map tends to a constant. However, general numerical results for most other chaotic maps have not been found so far.

MW (old): I reworded the sentence starting ``Some simulation results ...''. Is this true? Did some people really incorrectly conclude that a proportional relationship exists between the stationary decay rate and the diffusion? And did they incorrectly conclude this only because they didn't use enough resolution? Are [3, 20] the people who made such an error or are they the people who pointed out the error?

Todo

TC: In both [3, 20], there are analytical results to compare with their numerical simulations. So I believe their conclusions are correct--the proportional relation between diffusion and stationary decay rate. However, they deal with the cases that the diffusions are relatively large ($D = 10^{-2} \sim 10^{-4}$ in [3], and 500×500 grids ($D \approx 10^{-4}$ when adding diffusion) in [20]).
In [3] the author mentions that when D goes to zero, decay rate should be a constant. In [20], they simply say the resolution is not enough so they cannot conclude the zero-diffusion decay rate.
Hence, to be safe I reword the above paragraph again.

TODO

In this paper we focus on the mixing process of the Standard Map in the near-zero diffusion limit. We present a numerical strategy to simulate the map with very high resolution (up to 8×10^4 by 8×10^4 grids) and hence very low numerical diffusion. This numerical strategy is realized by a Markov Chain simulation. To characterize the evolution of the scalar variance in the near-zero diffusion limit, we use the concept of cutoff from the study of finite Markov Chains. We present numerical evidence to suggest that the sequence of models presents a cutoff, which qualitatively characterizes the mixing process when the diffusion goes to zero. The main contribution of this paper is to build a bridge between finite Markov Chain theory and 2-D chaotic maps with small diffusion. Related analytical results for 1-D chaotic maps clarifying their relationship to the cutoff phenomenon can be found in [14].

This paper is organized as the follows: in Section 2 we briefly review the background of the operators we use and the cutoff phenomenon. Section 3 provides a model reduction view of building the Markov Chain model of chaotic maps. Numerical evidence of Standard Map cutoff is given in Section 4 and conclusions are presented in Section 5.

2 Background

2.1 The measure space and operators

We work on the probability space $(X, \mathcal{A}, \mu)$. We take $S : X \to X$ to be a transformation (or map) that is non-singular and measurable. We choose μ to be the Borel measure. In the measure space $(X, \mathcal{A}, \mu)$ we define the following operators.

Definition 1 (Perron-Frobenius operator)**.** *The Perron-Frobenius operator* $P : L^1(X) \to L^1(X)$ *associated with* S *satisfies*

$$\int_A (P\omega)(x)\mu(dx) = \int_{S^{-1}(A)} \omega(x)\mu(dx) \tag{1}$$

for every $\omega \in L^1(X)$ *and* $A \in \mathcal{A}$.

The Perron-Frobenius operator is linear. Because of our choice of measure space, the Perron-Frobenius operator can be interpreted as a map that evolves probability density functions. Also, suppose that $\bar{\omega}$ is an invariant measure of S, so that

$$\bar{\omega}(S^{-1}(A)) = \bar{\omega}(A) \text{ for all } A \in \mathcal{A}. \tag{2}$$

Then we have (omitting x)

$$P\bar{\omega} = \bar{\omega}. \tag{3}$$

MW (old): For consistency we should always write $P(\omega)$ or $P\omega$. I tend to value the latter to suggest its linearity.

TODO

Definition 2 (Koopman operator)**.** *Let $f \in L^\infty(X)$. The operator $U : L^\infty(X) \to L^\infty(X)$ defined by*

$$Uf(x) = f(S(x)) \tag{4}$$

is called the Koopman operator associated with S.

The Koopman operator is adjoint to Perron-Frobenius operator, which we write as $U = P^*$.

MW (old): I added the domain to the function spaces above, so we have $L^\infty(X)$ rather than just L^∞. Should we also add the range to make it completely clear? That is, $L^\infty(X, \mathbb{R})$?

TODO

2.2 Notion of a cutoff

In some Markov Chains, certain probability distributions converge to an equilibrium via a sharp transition, which becomes sharper for larger chains. This phenomenon is referred to as *cutoff* in the finite Markov chain literature [8, 4]. Here we extend the usual definition slightly to accommodate converge to non-zero distance values.

MW (old): Why don't we compute the variance of the f_n^k after subtracting the invariant distribution, or something, so that we really converge to zero?
TC: We discussed this before. When the diffusion is not zero, the stationary distribution is always uniform. However, Standard Map has non-chaotic regions, so as the diffusion decreases, each of the trajectories in Figure 2 goes to zero with decreasing exponential decay rate (the second largest eigenvalue). And even in the normalized plot, the decay rate is still decreasing, so for fixed k/t_n (say, $k/t_n = 3$), we can expect var(f_n^k) to converge to a constant, but it is not stationary. It is the eigenvector connrsponding to the second largest eigenvalue.

TODO

MW (old): I think we should name this d so that we can refer to it below in the definition of cutoff. Does it have a standard name? If not, maybe we should call it a distance function?

TODO

To any finite set Ω and any pair of probability measures ω, $\bar{\omega}$ on Ω we associate a real number $d(\omega, \bar{\omega})$ such that

$$d(\omega, \bar{\omega}) \in [0, 1] \tag{5a}$$

$$d(\omega, \bar{\omega}) = 0 \text{ if and only if } \bar{\omega} = \omega \tag{5b}$$

$$\max_{\Omega, \omega, \bar{\omega}} d(\omega, \bar{\omega}) = M_d. \tag{5c}$$

Note that d need not satisfy the triangle inequality and so is not a metric.

MW (old): Should the max above really be over all Ω as well? It seems to be we only have one Ω.

TODO

MW (old): Should the max above be a sup? Why do we know the max is achieved?

TODO

MW (old): Check that the not-a-metric comment above is correct.

TODO

Consider a sequence of finite probability spaces (Ω_n) for $n = 1, 2, \ldots$. We think of n as the size of the space. Each space is equipped with a probability measure $\bar{\omega}_n$ which we think of as the unique invariant measure of a Markov

chain on Ω_n. For each n we now take a sequence of probability measures ω_n^k for $k = 0, 1, 2, \ldots$ such that

$$\lim_{k\to\infty} d(\omega_n, \bar{\omega}_n) = 0. \tag{6}$$

The ω_n^k should be thought of as an initial condition ω_n^0 and then iterates of the distribution under the evolution of a Markov chain.

MW (old): The above paragraphs are a bit of a mess, but somehow I wanted to explain what the objects are. Do we want the ω to be pdfs, or do we rather want them to be functions on Ω?

TODO

MW (old): Use ldots (...) rather than ... to get better spacing.

TODO

The definition of a cutoff follows.

Definition 3 (Cutoff)**.** *Take a family $(\Omega_n, \bar{\omega}_n, (\omega_n^k)_{k=0}^\infty)_{n=1}^\infty$ of finite probability spaces Ω_n and probability measures $\bar{\omega}_n$ and ω_n^k. This family presents a d-cutoff if there exists a sequence $(t_n)_{n=1}^\infty$ of positive reals such that, for any $\epsilon \in (0, 1)$,*

$$\lim_{n\to\infty} d(\omega_n^{k_n}, \bar{\omega}_n) = m \text{ if } k_n > (1+\epsilon)t_n \tag{7a}$$

$$\lim_{n\to\infty} d(\omega_n^{k_n}, \bar{\omega}_n) = M \text{ if } k_n < (1-\epsilon)t_n \tag{7b}$$

MW (old): Should we define the t_n to be *cutoff times*?

TODO

MW (old): In the cutoff definition we don't say what the k_n are. This needs to be clarified.

TODO

MW (old): Remind me why we need the ϵ in the cutoff definition? Why can't we just say $k_n > t_n$ and $k_n < t_n$?

TODO

This definition is taken from [8] with the change that m and M are 0 and M_d in the original. The reason for this modification will be clear when we present the results of Standard Map simulation.

The definition of cutoff implies that the change of $d(\omega_n^k, \bar{\omega}_n)$ from M to m happens ever more rapidly as n increases, but only in relation to the cutoff times t_n. We can think of this as rescaling the each trajectory $(\omega_n^k)_{k=0}^\infty$ in time by t_n and seeing cutoff as the limit of these rescaled trajectories to a step function.

3 A model reduction view and the numerical strategy

3.1 A model reduction view

In this section we explain the numerical strategy to perform the simulation in detail. It is based on a model reduction view of Perron-Frobenius and Koopman operators, and the advection-diffusion process is done by performing a Markov Chain/linear system simulation in a finite-dimensional space. Although we make this numerical strategy extremely simple in order to scale to very fine resolution, it is still worth explaining how we make the simplification step by step.

MW: Cite the CDC paper here, also Chorin et al. TODO

Firstly, in the measure space $(X, \mathcal{A}, \mu)$, given an invertible map S, let P_S and U_S be the Perron-Frobenius and the Koopman operators of a map S. We have the following relations:

	forward in time	backward in time
probability density	P_S	$P_{S^{-1}}$
scalar function	$U_{S^{-1}} = P^*_{S^{-1}}$	$U_S = P^*_S$

MW (old): What are A and B here? The language below about ‘‘infinite dimensional matrices’’ is not mathematically meaningful and should be cut. It’s fine to talk about P_S and U_S being dual, so long as they really are (check this). TODO

Note the above interpretation of the operators comes from the choice of μ to be the Borel measure. One can have different interpretation if the choice is different. Moreover, since U_S, $U_{S^{-1}}$, P_S, and $P_{S^{-1}}$ are all real linear operators, one can think them as matrices, and so U_S is simply the transpose of P_S, so do $U_{S^{-1}}$ and $P_{S^{-1}}$. We can also find the relation between P_S and $U_{S^{-1}}$:

MW (old): Under what conditions does (8) actually hold? TODO

$$U_{S^{-1}} = [\bar{\omega}^{-1}] \circ P_S \circ [\bar{\omega}], \tag{8}$$

where $[\bar{\omega}]f(x) = f(x)\bar{\omega}(x)$ and $[\bar{\omega}^{-1}]f(x) = f(x)/\bar{\omega}(x)$ for any function $f : X \to \mathbb{R}$. This result can be derived directly from the definition of the two operators. Therefore once we have obtained any one of the four operators, it is in general easy to find the other three.

The goal of this paper is to study how a scalar function is advected by a chaotic map in the near-zero diffusion limit. From the above table, the operator one needs in this situation is $U_{S^{-1}}$. For a given scalar function $f^0 : X \to \mathbb{R}$,

$$f^{k+1} = U_{S^{-1}} f^k, \text{ for all } k. \tag{9}$$

The way we actuall perform the iteration is we approximate $U_{S^{-1}}$ by a finite-dimensional Markov transition matrix $B_n \in \mathbb{R}^{n\times n}$, and use it to iterate $f_n^0\mathbb{R}^n$, an approximation of the given initial function. Thus,

$$f_n^{k+1} = B_n f_n^k, \text{ for all } k. \tag{10}$$

The difference between the approximate process (10) and the actual one (9) is treated as the numerical diffusion, and this numerical diffusion can be reduced by increasing n. Thus we can have a sequence of irreducible Markov transition matrices B_n (and similarly A_n) which converge to $U_{S^{-1}}$ (and U_S) when n goes to infinity. Each of the $B_n \in \mathbb{R}^{n\times n}$ represents a finite-dimensional approximation of $U_{S^{-1}}$. Because we treat the error between (10) and (9) as the added numerical diffusion, when n goes to infinity, the numerical diffusion decreases to zero. The way to find B_n for a given $U_{S^{-1}}$ is through the procedure of model reduction.

Begin with B_n. It evolves $f_n \in \mathbb{R}^n$. We first define a map (an observer) $g_n : f(x) \mapsto f_n$ such that

$$(f_n)_i = (g_n(f(x)))_i = \int_{a_i} f(x)\mu(dx) \text{ for } i = 1, \ldots, n. \tag{11}$$

The sample space X is discretized into n grids $a_1, a_2, \ldots, a_n$ and each grid represents one state in the new sample space. There are various methods to find B_n such that $\lim_{n\to\infty} B_n = U_{S^{-1}}$. One example is the lattice method [19]. It approximates $U_{S^{-1}}$ by a permutation matrix, and then adds a smoothing step to make the Markov chain irreducible. In fact an optimal B_n can be found by the techniques of optimal model reduction [2, 12, 11]: for any f and $f_n = g_n(f)$, the optimal reduced model of $U_{S^{-1}}$ is the B_n such that

$$B_n f_n = \underset{f_n'}{\operatorname{argmin}} ||f_n' - g_n(U_{S^{-1}} f)||_{\mathrm{diag}(\sqrt{\bar{\omega}})}. \tag{12}$$

That is, it minimize the iteraiton error in the weighed norm space. Suppose the grids are numbered by $a_1, a_2, \ldots, a_n$. The optimal $A_n \in \mathbb{R}^{n\times n}$ satisfying (12) can be calculated explicitly as

$$(B_n)_{ij} = \frac{\bar{\omega}(S(a_j) \cap a_i)}{\bar{\omega}(a_j)}. \tag{13}$$

To evaluate (13) requires the calculation of $S(a_j)$, finding the area intersection, and integrating over a non-uniform measure $\bar{\omega}$. For most of the maps, these steps can only be done numerically. Similarly, the optimal A_n, which is the best approximation of U_S, has the explicit form

$$(A_n)_{ij} = \frac{\bar{\omega}(S^{-1}(a_j) \cap a_i)}{\bar{\omega}(a_j)}. \tag{14}$$

One can check that the pair $\{A_n, B_n\}$ satisfies

$$B_n = \mathrm{diag}(\bar{\omega}_n^{-1}) A_n^T \mathrm{diag}(\bar{\omega}_n), \tag{15}$$

which is a discrete version of (8). Apparently, how good the approximation is for a certain n depends on the choice of mesh grids. In this article, we simply use regular grids: for $X = [0,1] \times [0,1]$, each grid has size $h = 1/\sqrt{n}$ in each dimension and area $1/n$, $A_n \in \mathbb{R}^{n \times n}$.

MW (old): The following example seems again to be irrelevant for the main point of the paper. I think reviewers will be confused by it. We should just directly define the method that we will use to generate the numerical results and cut the example below, I think. Just to make sure I understand things here, the numerical method described below and used for ?? is not the method used to generate the final numerical results, right?

TODO

The proposed numerical strategy can be easily applied to 1-D maps, where $\bar{\omega}(S^{-1}(a_j) \cap a_i)$ is calculated by linear integration with measure $\bar{\omega}$ over the overlapping region of grids $S^{-1}(a_j)$ and a_i. In 2-D, it is already hard to decide whether $S(a_j) \cap a_i$ is non-empty, let alone to calculating the areas. Therefore further simplification needs to be done for high resolution. In fact, when the grid gets smaller, the variation of $\bar{\omega}$ over one grid also becomes smaller, and one can safely approximate B_n by

$$(B_n)_{ij} \approx \frac{\mu(S(a_j) \cap a_i)}{\mu(a_j)}. \tag{16}$$

To go up to extremely high resolution, we make the approximation procedure even simpler: suppose $X = T^2$ and $S : T^2 \to T^2$, and let the grid size be h in both dimensions. We number the grids by $a_1, a_2, \ldots, a_{n^2}$. Let $\mathbf{x}_i = (x_{1i}, x_{2i})$ be the center of grid a_i; then

$$(B_n)_{ij} \approx \begin{cases} \frac{1}{4} & \text{if } S(x_{1j} \pm \frac{h}{2}, x_{2j} \pm \frac{h}{2}) \in a_i, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

To evaluate (17) we need only map points on the grid corners. The approximation of B_n has only 4 non-zeros in each row. Compared with (13), this approximation is far from optimal, but when n goes to infinity, it still converges to $U_{S^{-1}}$. During the simulation, the matrix B_n is never explicitly formed. We need only store a size n state vector during the iteration. This approximation ensures that the operation count of evolving the system is always a constant times n. Equation (17) will serve as our main tool in Section 4.

We want to stress again that although the numerical strategy we use is just equation (17), the above simplification procedure is still necessary for understanding the insight of it. We apply a similar numerical strategy in [15] for the

mixing process of a microfluidic mixing channel. The flow map generated by the microfluidic mixing channel does not have a uniform invariant distribution. In fact, the physical intrepretation of the invariant distribution is the normal velocity profile on the cross-section of the channel. Hence the choice of the operators to assure the correct physical conservation is thus crucial for the accuracy of the result. Refer to [15] for more detail.

MW (old): We should have a figure like ?? but for the standard map and using the numerical method that we actually use for the numerical results section.

TODO

3.2 Additional smoothing steps

Using the numerical strategy from the previous section, we can evolve a function or a probability distribution by the map with some small numerical diffusion. The effect of numerical diffusion is similar to physical diffusion on large scales, but their behaviors can be quite different on small scales, and small-scale phenomena can be important for a chaotic map to form its stationary eigenfunction. Therefore, to simulate the physical diffusion correctly, we need to simulate the map with far higher resolution with some additional diffusion added. The additional diffusion can be added either in the spatial or frequency domains. In the spatial domain, we adopt the method of adding a smoothing step used in [24]:

$$f^{k+1}_{(p,q)} = \sum_{|r|,|s|\leq 2} C_{|r|}C_{|s|}f^{k}_{(p,q)} \tag{18}$$

with $C_0 = 1/8, C_1 = 1/4$, and $C_2 = 3/16$ and (p, q) the two-dimensional index in the grid. This creates a large-scale diffusion $D \approx h^2$, which is several times larger than the numerical diffusion we originally have [24]. We use a smoothing operator M_n and $f^{k+1} = M_n(f^k)$ to denote the above smoothing step. Alternatively, in frequency domain a two-dimensional FFT/IFFT with multiplication by a wave-number-dependent constant in between can be applied to simulate physical diffusion. This procedure is denoted by an operator F_n and $f^{k+1} = F_n(f^k)$. Note that the FFT/IFFT scheme is much more expensive when n is large, because a large data transfer is required between the FFT/IFFT in the first and second dimensions.

4 Numerical results

4.1 Small diffusion

MW: Do we have a citation that proves the standard map is chaotic below?

TODO

MW: What is g_n below. Is n supposed to be the iteration number here?

TODO

We study the Standard Map on T^2:

$$\begin{aligned} x_1' &= x_1 + x_2 + \epsilon \sin 2\pi x_1 \pmod 1, \\ x_2' &= x_2 + \epsilon \sin 2\pi x_1 \pmod 1. \end{aligned} \tag{19}$$

This map is known to be chaotic, where ϵ is a parameter that can be adjusted to change the behavior of the map. Various studies of how a point is advected by the map can be found, for example, in [17]. Here we mainly focus on how a scalar function is evolved by the map with the presence of small diffusion. A simulation of the Standard Map using $f^0 = \cos(2\pi x_2)$, $f_n^0 = g_n(f^0)$ and $n = 500 \times 500$ is shown in 1. Because this map is volume preserving, its invariant measure is uniform, and hence by equation (8), $U_{S^{-1}} = P_S$. Thus there is no difference between evolving a scalar function and a probability distribution. We make the following remarks about the notion of cutoff:

1. The quantity of interests in chaotic mixing is the variance of the function, and in the study of cutoff phenomenon, total variaion distance is the object that is discussed the most. There is no difficulty in setting the distance function $d(\omega, \bar{\omega})$ to be the 2-norm as follows:

$$d(\omega_n, \bar{\omega}_n) = \left(\sum_{i=1}^{n} \left(\frac{(\omega_n)_i}{(\bar{\omega}_n)_i} - 1 \right)^2 (\bar{\omega}_n)_i \right)^{1/2}. \tag{20}$$

   This corresponds to the study of L_2 cutoff in the cutoff terminology. When talking about Standard Map, due to the uniformity of the invariant distribution, the above distance is the same as the standard deviation of a function defined as $(f_n)_i = (\omega_n)_i/(\bar{\omega}_n)_i$ with mean 1. Thus in all of our simulations, we scale the mean of the initial functions to be 1 to fit the definition.

2. It is clear that the maximal value of the 2-norm distance is ∞, so we need to set $M_d = \infty$ (instead of $M_d = 1$ for total variation distance). In the original cutoff definition, (M, m) always equals $(M_d, 0)$, but the f_n we are interested in are sinusoidal functions in the x_1 and x_2 directions, which do not maximize the distance function d. Hence we scale the standard deviation of f^0 to be 1 in all the simulations and set M to be 1.

3. The Standard Map is known to have some non-chaotic regions when ϵ is not zero, so the distance d converges to a value $m \neq 0$.

The above points explain why we need to slightly modify the definition of cutoff phenomenon from the standard definition given by Diaconis.

Using the numerical discretization (17), we can generate a sequence of $B_n = A_n^T$ to approximate the Koopman operator $U_{S^{-1}}$ of the inverse map by changing

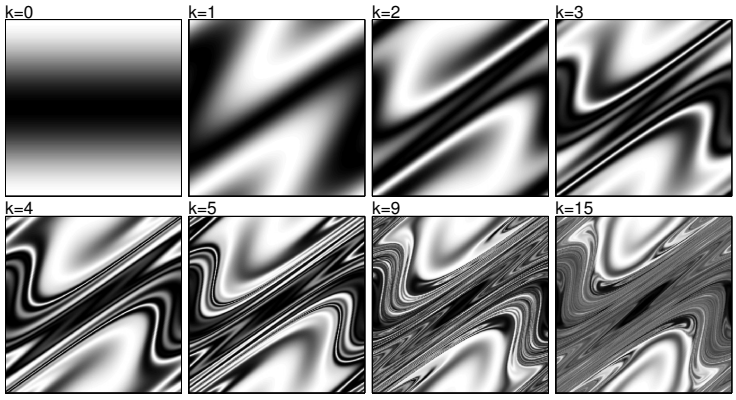

Figure 1: For Standard Map, $\epsilon = 0.3$, B_n with $n = 500^2$ is applied to simulate the system with initial condition $f^0 = \cos(2\pi x_2)$. The first eight iterations are plotted.

the number of mesh grids on T^2, and have $\lim_{n\to\infty} B_n = U_{S^{-1}}$. To simulate the more physical diffusion, we define $\bar{B}_n = M_n \circ B_n$ and $\tilde{B}_n = F_n \circ B_n$. The diffusion of F_n is chosen to be equal to one smoothing step ($D = 1/n = h^2$) for comparison.

For $\epsilon = 0.3$, $n = 40000^2$, $f^0 = \cos(2\pi x_2)$, $f_n^0 = g_n(f^0)$. We simulate $f_n^{k+1} = B_n f_n^k$ for $k = 1$ to 20. Each f_n^k is then applied a 2-D Fourier transform to obtain the frequency domain $\hat{f}_n^k$. The magnitude versus wave number plot is shown in the left of Figure 2. From this plot one can clearly observe how the chaotic map maps low frequency components to high frequency ones (it also does the converse). Since the number of grids is 4×10^4 by 4×10^4, the largest wave number it can catch is around 2×10^4 and it has large numerical diffusion. A major portion of components are being pushed to high wave numbers in just a few iterations, and where diffusion works well to smooth out the corner. Once the effect of pushing and smoothing is balanced, an eigenfunction is thus formed.

To compare the difference between operators B_n, $\bar{B}_n$, and $\tilde{B}_n$, we do the same simulation for all three operators. The result is plotted in the right of Figure 2 for $k = 3, 7$, and 20. The three trajectories agree well in the low frequency part because all kinds of diffusion have a tiny effect in low frequency terms. They begin to have discrepancy at around wave number 2×10^3. The smoothing operator M_n actually has larger diffusion than the FFT/IFFT operator F_n.

The above results also explain why it is hard to simulate the function evolved by a chaotic map correctly: very fine grids are needed to capture the high frequency terms generated in just a few iterations.

All three operators B_n, $\bar{B}_n$, and $\tilde{B}_n$ satisfying

$$\lim_{n\to\infty} B_n = U_{S^{-1}}, \quad \lim_{n\to\infty} \bar{B}_n = U_{S^{-1}}, \quad \lim_{n\to\infty} \tilde{B}_n = U_{S^{-1}}$$

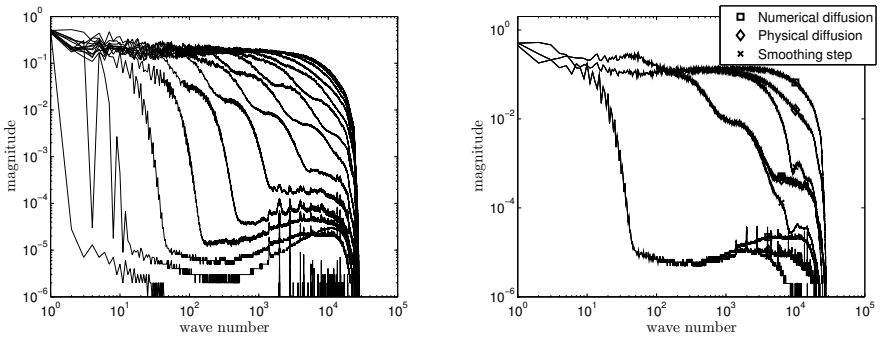

Figure 2: For Standard Map, $\epsilon = 0.3$, B_n with $n = 40000^2$ is applied to simulate the system with initial condition $f^0 = \cos(2\pi x_2)$. The magnitude versus wave number plot for the first 20 iterations is shown in the left plot. In the right plot, B_n, $\bar{B}_n$, and $\tilde{B}_n$ are applied to simulate the same system and the magnitude versus wave number curves are plotted for iteration 3, 7, and 20.

can be applied to study the behavior of near-zero diffusion limit of Standard Map. We choose B_n and $\bar{B}_n$ for the following studies because the computation of FFT/IFFT diffusion of $\tilde{B}_n$ is too expensive when n is very large. Since their relation in the magnitude of diffusion is $B_n < \tilde{B}_n < \bar{B}_n$, if we can show both B_n and $\bar{B}_n$ have similar behaviors, it is reasonable to believe that $\tilde{B}_n$ has, too.

4.2 Cutoff phenomenon

We present the main results of this article in this section. For n from 2500^2 to 80000^2, the variance of f_n^k versus iteration plots using B_n and $\bar{B}_n$ as the Koopman operators are shown in the left of Figures 3 and 4. The tendency is clear: when n gets larger, the variances stay high ($M_B = M_{\bar{B}} = 1$) for more iterations and then drop rapidly to $m_B = 0.4521$ and $m_{\bar{B}} = 0.4498$, respectively—they do not drop to zero because there are unmixed "islands". The slope of the rapid dropping also becomes slightly milder when n increases. To justify whether they satisfy the definition of cutoff given in Section 2, we let t_n equal the point where each trajectory passes through $(M_\star + m_\star)/2$, where $\star = \{B, \bar{B}\}$ and we normalize all trajectories by rescaling t_n to 1. The results are plotted in the right of Figure 3 and 4. Although the normalized trajectories are very similar, one can still see that when n gets larger, the trajectory becomes sharper. In the left of Figure 5 we plot t_n versus $1/n$ in log scale and see two straight lines. Note that for both cases, we have $D \sim O(1/n)$. Hence this plot shows that the cutoff time is inversely proportional to $\log(D)$. To see how fast the trajectories converge to their limits, we define the interpolating functions for the trajectories in the normalized plots of Figures 3 and 4 to be $\beta_{B_n}(x)$ and $\beta_{\bar{B}_n}(x)$, where x represents the normalized iteration. The function they should

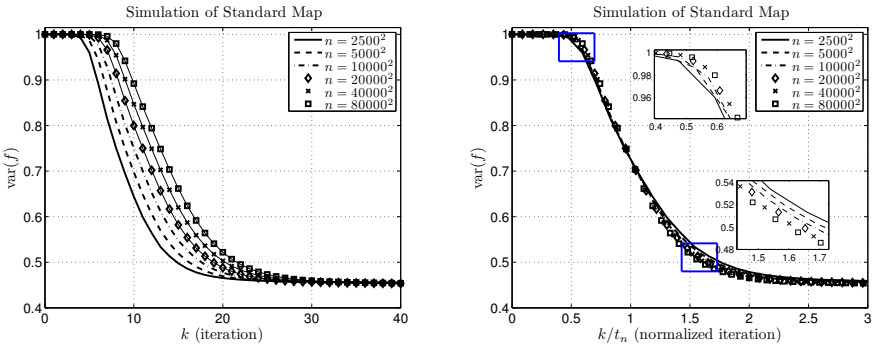

Figure 3: The left figure shows the variance of f_n^k versus iteration trajectories of Standard Map simulation. We use $\epsilon = 0.3$, $f^0 = \cos(2\pi x_2)$, and number of grids varies from 2500^2 to 80000^2. The normalized plot is shown on the right. We rescale the iteration axis and make all trajectories pass through the point $(1, 0.7260)$ to observe the cutoff. The two small plots show the detailed view of the corners when trajectories have sharp change.

converge to when $n \to \infty$ is

$$\nu_{\star_\infty}(x) = \begin{cases} M_\star & \text{if } x < 1, \\ m_\star & \text{otherwise,} \end{cases} \tag{21}$$

where $\star = \{B, \bar{B}\}$. Define the distance between $\nu_{\star_n}(x)$ and $\nu_{\star_\infty}(x)$ to be

$$\Delta_\star^l = \int_0^l |\nu_{\star_\infty}(x) - \nu_{\star_n}(x)| dx \tag{22}$$

for $\star = \{B, \bar{B}\}$. We calculate Δ_B^3 and $\Delta_{\bar{B}}^3$ for all $\nu_{B_n}(x)$ and $\nu_{\bar{B}_n}(x)$, and plot them versus t_n in the right of Figure 5. This figure shows that when t_n increases, $\Delta_\star^3$ decays slightly slower than linear, but clearly when $n \to \infty$, Δ_B^3 and $\Delta_{\bar{B}}^3$ are both going down, which strongly suggests that both sequences of Markov Chains present cutoffs.

4.3 Chaotic parameter ϵ

We also study how the parameter ϵ affects the mixing trajectory when the diffusion is small. We use $f^0 = \cos(2\pi x_1)$ and $f^0 = \cos(2\pi x_2)$ as the initial functions and vary ϵ from 0.1 to 0.9. The simulation is done by $n = 40000^2$ with operator B_n. The results are shown in Figure 6. For both initial functions and $\epsilon \geq 0.3$, the trajectories all have similar tendency. As for $\epsilon = 0.1$, the trajectories show no sign of dropping in the first 50 iterations. Hence in Figure 7 we simulate $\epsilon = 0$ and 0.1 for both initial functions for 800 iterations. One can see that for $f^0 = \cos(2\pi x_2), \epsilon = 0$, the trajectory still stay almost 1 for the first 800 iterations, and in the case $\epsilon = 0.1$ it has a small drop but we have no

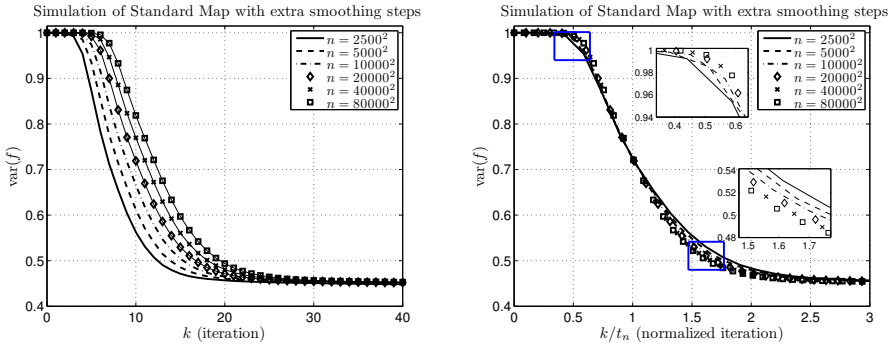

Figure 4: The left figure shows the variance of f_n^k versus iteration trajectories of Standard Map simulation when a smoothing step is added after each iteration. We use $\epsilon = 0.3$, $f^0 = \cos(2\pi x_2)$, and the number of grids varies from 2500^2 to 80000^2. The normalized plot is shown in the right. We rescale the iteration axis and make all trajectories pass through the point $(1, 0.7249)$ to observe the cutoff. The two small plots show the detailed view of the corners when trajectories have sharp change.

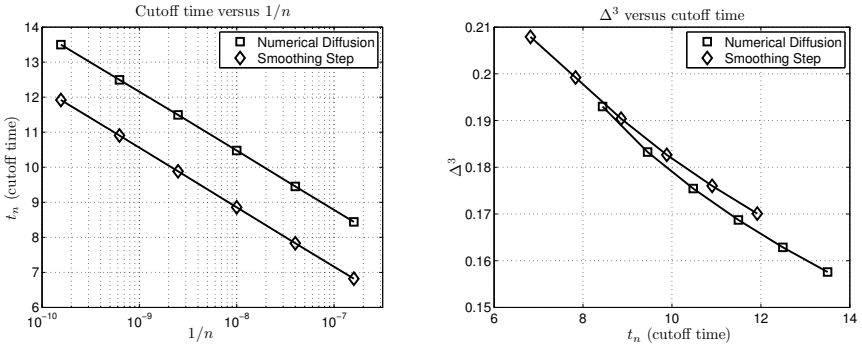

Figure 5: The left plot shows how cutoff time t_n relates to $1/n$ in log scale. Since we have $D \sim O(1/n)$ for our simulation strategy, one can conclude that t_n is inversely proportional to $\log(D)$. The right plot shows how the normalized trajectories converge to their limit. Δ^3 is defined in equation (22). Both curves predict that when t_n is very large, Δ^3 goes down, and the normalized trajectory would probably become a step function.

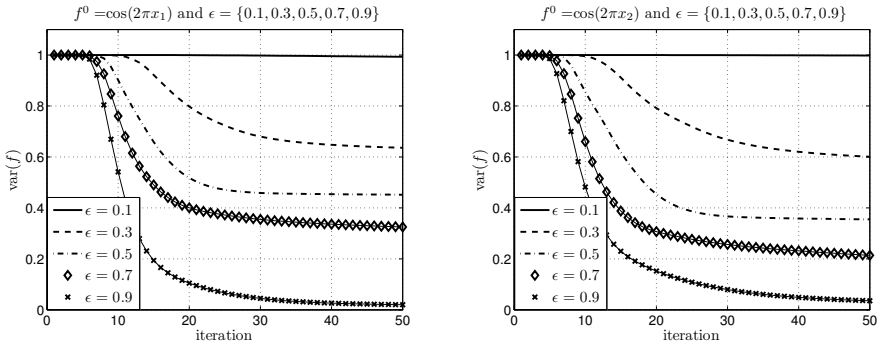

Figure 6: The two figures show the Standard Map simulations using B_n, $n = 40000^2$ with different ϵ and initial function f^0. For $\epsilon = \{0.3, 0.5, 0.7, 0.9\}$ and $f^0 = \{\cos(2\pi x_1), \cos(2\pi x_2)\}$, the variance trajectories all have sharp changes.

evidence to say it presents a cutoff or not. A more interesting thing is observed when $f^0 = \cos(2\pi x_1)$. In this case, the trajectory of $\epsilon = 0$ crosses over the one of $\epsilon = 0.1$ at around iteration 627. The phenomenon that the $\epsilon = 0$ map mixes faster than some other trajectories with higher ϵ is also observed in [16]. Clearly this is because when $\epsilon = 0$, there is no unmixed region (no "islands") and so the decreasing rate of variance would not slow down by forming the eigenfunction. Hence when $\epsilon = 0$ the mixing rate only depends on whether the map is good in mixing the initial functions.

All the numerical results are generated by a 72-node cluster with gigabit ethernet connection. Each CPU is equipped with 1GB RAM, and the largest simulation (80000×80000 number of grids) needs 51.2GB of memory to store a state vector.

4.4 Zero diffusion and norm issues

In all above simulations we observe how the variance of a scalar function is evolved by Standard Map with small diffusion. As we have mentioned, this is equivalent to the study of L_2 cutoff. One may also be interested in how other norms are evolved. In the left plot of Figure 8 we simulate Standard Map, $\epsilon = 0.3$ and $f^0 = \cos(2\pi x_1)$ by B_n with $n = 500^2$ and measure the following norms: L_1, L_2, $H_{-0.5}$, H_{-1}, and mix-norm [16]. All norms are normalized to equal 1 at iteration 0. Comparing with previous simulations, this one is done by very coarse grids, so one would not expect to see a very clear cutoff tendency. Nonetheless, these norms do form two groups of different behaviors: for L_1 and L_2 norms, one observes concave curves in the first few iterations, and for other norms, they drop immediately to around 0.5 and then oscillate.

To explain what causes this, let us consider the zero diffusion case: Standard Map is volume preserved, so without diffusion, the L_1 norm of the scalar function should stay constant when the function is evolved by the Koopman operator

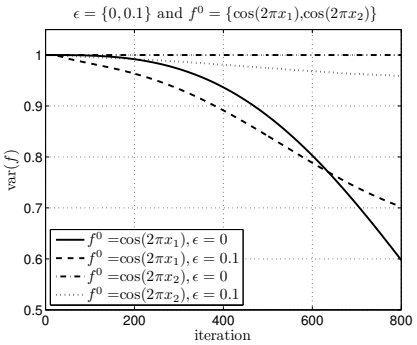

Figure 7: This figure shows the Standard Map simulation with $\epsilon = \{0, 0.1\}$ and $f^0 = \{\cos(2\pi x_1), \cos(2\pi x_2)\}$. For $f^0 = \cos(2\pi x_2)$, the limiting behavior of $\epsilon \to 0$ is quite clear. However, for $f^0 = \cos(2\pi x_1)$, the $\epsilon = 0$ trajectory crosses over the $\epsilon = 0.1$ one near iteration 627.

of Standard Map. Also, L_2 norm is invariant in the spatial and frequency domain. It is clear that the L_2 norm would also stay constant if there is no diffusion. For $H_{-0.5}$, H_{-1}, and mix-norm, they have a simple representation in frequency domain. Just the way diffusion works, to evaluate these norms, one multiplies each wave number term by a constant weight, and the weight is decreasing when the wave number increases. For 2-D domain, we plot the weights for each wave number in the right plot of Figure 8 for all the norms. The diffusion weights when $D = 0.01$ is also plotted on the same figure for comparison. One can conclude that the way $H_{-0.5}$, H_{-1}, and mix-norm work in frequency domain is very similar to diffusing the function according to a large diffusion and then evaluating their L_2 norms: the diffusion is very large ($D \sim 0.01$) and far from entering the region where we can observe cutoffs. One important feature of all 2-D chaotic maps is the frequency cascade: it generates high wave number terms in just a few iterations, as we have seen in Figure 2. The three norms $H_{-0.5}$, H_{-1}, and mix-norm suppress these terms by multiplying by small weights. The interesting cutoff phenomenon happens when diffusion is very small: $t_n \sim O(-\log(D))$. Thus these three norms lack the ability to capture things happening in small scales.

Finally we want to stress that there is no one norm that is better than another. They simply give us different information about the mixing process. The concave trajectories of L_1 and L_2 norms tell us the "irreversibility" feature of the chaotic system: beyond a certain iteration, it is not possible to recover the system's original state. The fast dropping trajectories of other norms show the increasing complexity of the function evolved by the chaotic map in the first few iterations.

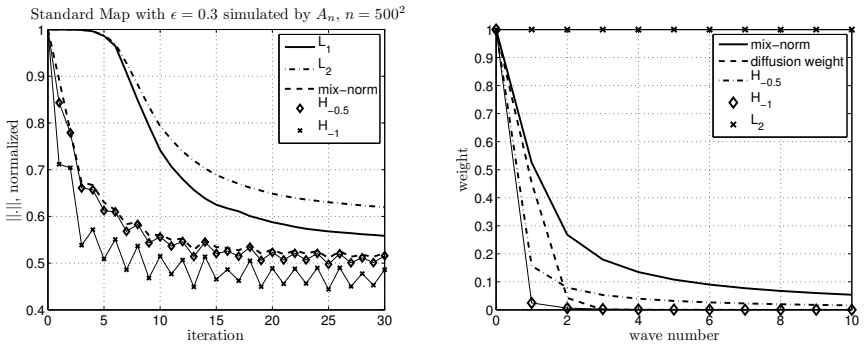

Figure 8: The left figure shows the Standard Map simulation with $\epsilon = 0.3$ and $f^0 = \cos(2\pi x_1)$. We use B_n with $n = 500^2$ and measure the following norms: L_1, L_2, $H_{-0.5}$, H_{-1}, and mix-norm. The norm trajectories form two groups: L_1 and L_2 norms show the tendency of presenting cutoffs, and the other three norms do not. In the right figure we plot the weights of L_2, $H_{-0.5}$, H_{-1}, and mix-norm at each wave number in the frequency domain. L_2 norm has a constant weight 1 for all wave numbers. The other three norms have much smaller weights in large wave numbers. We also plot the diffusion weights when $D = 0.01$ for comparison. One can see why these norms fail to measure the small-scale phenomena that cause cutoffs.

5 Conclusion

In our numerical simulations, we provide the evidence of cutoffs for the sequence of Markov chains generated by approximating the Koopman operator of Standard Map. The notion of a cutoff is first brought from finite Markov Chain studies to the study of chaos, and we find that not only is it suitable for characterizing the behavior of chaotic map in near zero-diffusion limit, but also it builds a bridge between large finite Markov Chain studies and chaotic maps. It is still unclear whether the Markov Chains that present cutoffs have direct link to chaos: are they discretizations or approximations of some chaotic map? We offer no good answer to this question so far. However, in [14] another attempt is made: we show that by choosing suitable initial distributions, a 1-D chaotic map with symbolic dynamics can have the same limiting behavior as the cutoff of random walk on n-dimensional hypercube problem [7], and thus it demonstrates another link between these two fields.

References

[1] T. M. Antonsen, Jr., Z. Fan, E. Ott, and E. Garcia-Lopez. The role of chaotic orbits in the determination of power spectra of passive scalars. *Physical Fluids*, 8(11):3094–3104, November 1996.

[2] C. L. Beck, S. Lall, and M. West. Model reduction, optimal prediction, and the Mori-Zwanzig representation of Markov Chain. (In preparation), 2008.

[3] S. Cerbelli, A. Adrover, and M. Giona. Enhanced diffusion regimes in bounded chaotic flows. *Physics Letters A*, 312(5):355–362, June 2003.

[4] G.-Y. Chen. *The Cutoff Phenomenon for Finite Markov Chains*. PhD thesis, Cornell University, August 2006.

[5] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, 36(2):491–515, February 1999.

[6] M. Dellnitz and O. Junge. Set oriented numerical methods for dynamical systems. In *Handbook of Dynamical Systems II: Towards Applications*, pages 221–264. World Scientific Publishing, 2002.

[7] P. Diaconis, R. L. Graham, and J. A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures and Algorithms*, 1(1):51–72, 1990.

[8] P. Diaconis and L. Saloff-Coste. Separation cut-offs for death and birth chain. *Annals of Applied Probability*, 16(4):2098–2122, 2006.

[9] D. R. Fereday, P. H. Haynes, A. Wonhas, and J. C. Vassilicos. Scalar variance decay in chaotic advection and Batchelor-regime turbulence. *Physical Review E*, 65(3):035301, February 2002.

[10] G. Froyland. Approximating physical invariant measures of mixing dynamical systems in higher dimensions. *Nonlinear Analysis, Theory, Methods and Applications*, 32(7):831–860, 1998.

[11] G. Froyland. Markov modelling for random dynamical systems. In *Proceedings of Equadiff 99, Berlin, Germany*, August 1999.

[12] G. Froyland. Extracting dynamical behaviour via Markov models. In *Nonlinear Dynamics and Statistics*, pages 283–324. Birkhäuser, Basel, 2001.

[13] P. H. Haynes and J. Vanneste. What controls the decay of passive scalars in smooth flows? *Physics of Fluids*, 17:097103, September 2005.

[14] T.-C. Liang and M. West. Cutoff phenomenon and the mixing properties of chaotic maps. (In preparation), 2008.

[15] T.-C. Liang and M. West. Optimized mixing in microfluidic channels. (In preparation), 2008.

[16] G. Mathew, I. Mezić, and L. Petzold. A multiscale measure for mixing. *Physica D*, 211(1-2):23–46, November 2005.

[17] E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, 2002. ISBN 0521010845.

[18] J. M. Ottino and S. Wiggins. Introduction: Mixing in microfluidics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 362(1818):923–935, March 2004.

[19] R. T. Pierrehumbert. Lattice models of advection-diffusion. *Chaos*, 10(1):61–74, March 2000.

[20] A. Pikovsky and O. Popovych. Persistent patterns in deterministic mixing flows. *Europhysics Letters*, 61(5):625–631, March 2003.

[21] D. Rothstein, E. Henry, and J. P. Gollub. Persistent patterns in transient chaotic fluid mixing. *Letters to Nature*, 401:770–772, October 1999.

[22] J.-L. Thiffeault. Scalar decay in chaotic mixing. In *Transport in Geophysical Flows: Ten Years After, Proceedings of the Grand Combin Summer School*, June 2004.

[23] J.-L. Thiffeault and S. Childress. Chaotic mixing in a torus map. *CHAOS*, 13(2):502–507, June 2003.

[24] Y.-K. Tsang, T. M. Antonsen, Jr., and E. Ott. Exponential decay of chaotically advected passive scalars in the zero diffusivity limit. *Physical Review E*, 71:066301, June 2005.

[25] G. A. Voth, G. Haller, and J. P. Gollub. Experimental measurements of stretching fields in fluid mixing. *Physical Review Letters*, 88(25), June 2002.