

# Introducing E-tec: Ensemble-based Topological Entropy Calculation

Eric Roberts,<sup>1, a)</sup> Suzanne Sindi,<sup>1, b)</sup> Spencer Smith,<sup>2, c)</sup> and Kevin A. Mitchell<sup>1, d)</sup>

<sup>1)</sup>*School of Natural Sciences, University of California, Merced, California, 95343*

<sup>2)</sup>*Department of Physics, Mount Holyoke College, Massachusetts, 01075*

(Dated: 15 June 2018)

Topological entropy measures the number of distinguishable orbits in a dynamical system, thereby quantifying the complexity of chaotic dynamics. One approach to computing topological entropy in a two-dimensional space is to analyze the collective motion of an ensemble of system trajectories taking into account how trajectories “braid” around one another. In this spirit, we introduce the Ensemble-based Topological Entropy Calculation, or E-tec, a method to derive a lower-bound on topological entropy of two-dimensional systems by considering the evolution of a “rubber band” (piece-wise linear curve) wrapped around the data points and evolving with their trajectories. The topological entropy is bounded below by the exponential growth rate of this band. We use tools from computational geometry to track the evolution of the rubber band as data points strike and deform it. Because we maintain information about the configuration of trajectories with respect to one another, updating the band configuration is performed locally, which allows E-tec to be more computationally efficient than some competing methods. In this work, we validate and illustrate many features of E-tec on a chaotic lid-driven cavity flow. In particular, we demonstrate convergence of E-tec’s approximation with respect to both the number of trajectories (ensemble size) and the duration of trajectories in time.

**From the stirring of dye in viscous fluids to the availability of essential nutrients spreading over the surface of a pond, nature is rife with examples of mixing in two-dimensional fluids. The long-time exponential growth rate of a thin filament of dye stretched by the fluid is a well-known proxy for the quality of mixing in two dimensions. In the real-world study of mixing, this stretching rate may be hard to compute; the velocity field may not be known or may be expensive to recover or approximate, thus limiting our knowledge of the governing system and underlying mechanics driving the mixing. One alternative is to use time-ordered trajectory data, often obtained from tracer particles such as ocean drifters. In this paper, we use the collective motion of such trajectories, along with tools from computational geometry, to develop a lower bound to the stretching rate. The lower bound is obtained by approximating the filament of dye with a piece-wise linear, non-intersecting “rubber band” stretched around the data points. We call our algorithm the Ensemble-Based Topological Entropy Calculation, or E-tec.**

## I. INTRODUCTION

A variety of techniques have been used to quantify complexity in dynamical systems theory. These tools include the finite-time Lyapunov exponent field<sup>1,2</sup>, which measures the time for nearby trajectories to separate; operator-theoretic methods, such as the eigenfunctions and eigenvalues of the Koopman operator<sup>3</sup>; and numerical evolution of a material-curve<sup>4</sup> to compute the topological entropy, which measures the proliferation of distinguishable orbits<sup>5</sup>. Such knowledge aids greatly in a wide variety of natural and industrial fluid systems, including the large-scale dispersion of pollutants in the Earth’s atmosphere and oceans<sup>6</sup> and the rapidly developing field of microfluidics<sup>7</sup>. However, a problem remains for many techniques — the fine-scale structure of a system may not appear without a high point density. A sufficient number of system trajectories and/or the linearizations about these trajectories may simply be too expensive to compute or to measure experimentally. We seek techniques that can accommodate such sparse data.

Our goal is to compute material-line stretching rates using only 2D particle trajectories, like those collected from oceanic floats<sup>8,9</sup> or fluorescent beads in microfluidic systems<sup>10,11</sup>. These data sets may be sparse, and hence may not fully sample all of the 2D space. We are motivated by Budišić, Allshouse, and Thiffeault<sup>9,12,13</sup>, who use braiding theory to compute a lower bound for topological entropy of such data sets. The initially embedded material-curve is thought of as an elastic line whose growth rate is computed using the collective motion of all available trajectories moving through space in concert. In essence, the relative motion of an ensemble of trajectories in space encodes global information that is not contained in any one individual trajectory. That is, extra information is “hiding” in an ensemble of trajectory-

---

<sup>a)</sup>Electronic mail: [eroberts5@ucmerced.edu](mailto:eroberts5@ucmerced.edu)

<sup>b)</sup>Electronic mail: [ssindi@ucmerced.edu](mailto:ssindi@ucmerced.edu)

<sup>c)</sup>Electronic mail: [smiths@mtholyoke.edu](mailto:smiths@mtholyoke.edu)

<sup>d)</sup>Electronic mail: [kmitchell@ucmerced.edu](mailto:kmitchell@ucmerced.edu)

ries, which is not exploited in a trajectory-by-trajectory approach.

In this paper, we focus on these underlying stretching and folding processes that drive mixing in two dimensional fluids. We apply computational geometry techniques to develop a 2D algorithm titled the Ensemble-based Topological Entropy Calculation (E-tec). E-tec achieves three main goals: a) estimation of a lower bound to the topological entropy on data sets, b) convergence to the topological entropy as ensemble size increases, c) linear scaling in runtime with the length of trajectories and  $N \log N$  scaling with the number of trajectories  $N$ , as a worst-case scenario.

The remainder of this paper is broken up into four sections. We first review topological entropy (Sect. II) and then summarize our E-tec algorithm (Sect. III). We evaluate the performance of E-tec on a chaotic, lid-driven cavity flow as a test case (Sect. IV). Finally, we demonstrate E-tec’s robustness and show evidence that the E-tec runtime compares favorably to braiding algorithms (Sect. V). Appendix A contains procedural details of the algorithm.

## II. TOPOLOGICAL ENTROPY

Topological entropy is a measure of the growth rate of the number of distinguishable orbits<sup>14</sup>. More formally, topological entropy is defined by considering equivalence classes of trajectories of duration  $T$  that are only distinguished if they are, at *any* point in time, further than some resolution  $\epsilon > 0$  apart. The number of these  $\epsilon$ -distinct classes of trajectories increases as both  $T \rightarrow \infty$  and  $\epsilon \rightarrow 0$ . Topological entropy measures the growth of all  $\epsilon$ -distinct trajectories as  $T \rightarrow \infty$ . Specifically, the topological entropy  $h$  is the exponential growth rate in time of the number of distinct trajectory classes for arbitrarily small  $\epsilon$ .

In two-dimensional flows, topological entropy  $h$  can be estimated by embedding an initial material-curve, e.g. a line of dye, of length  $L_0$  in the system and estimating its growth under the evolution of the flow<sup>15</sup>. At long times, the length  $L(t)$  of the curve as a function of time  $t$  grows exponentially as

$$L(t) \approx L_0 e^{ht}. \quad (1)$$

Thus, direct computation of the curve’s evolution is troublesome in chaotic flows since the length is expected to grow exponentially fast, which requires an exponentially growing number of trajectories to maintain sufficient point density of the curve. Other techniques for extracting topological entropy operate on a trajectory-by-trajectory basis, i.e. ensemble-averaging some quantity (such as the Jacobian singular values) computed one trajectory at a time. This is the approach taken in recent work on *expansion entropy*<sup>16</sup>, a generalization of topological entropy.

As an alternative approach, a lower bound to the topological entropy may be computed with a finite number of trajectories and no detailed knowledge of the velocity field. The material-curve to be advected is represented by a taut elastic loop that wraps tightly around trajectories that strike it. Since an advected material-curve may be continuously deformed into this taut loop given the same trajectory evolution, the need for maintaining material-curve point density is eliminated. The loop is stretched and folded over itself exponentially many times in a chaotic flow. Its exponential growth rate is a lower bound to the full system’s topological entropy<sup>9</sup>.

In this more topological setting, braiding theory has been used to compute this lower bound. The Finite-Time-Braiding-Entropy (FTBE) method<sup>12</sup> evolves the loop forward using the entanglement of a finite number of trajectories. However, this method scales quadratically in the number of points<sup>12</sup>, rendering braiding exponent calculations unwieldy for systems requiring many points.

To develop a computationally efficient method to estimate a lower bound on the topological entropy of a planar flow, we compute the stretching rate of an advected elastic curve directly. Referring now to the elastic curve or loop as a rubber band, we use the same FTBE idea of trajectories working in concert to stretch and fold the band. The E-tec algorithm achieves this using the same input: i) a set of (typically aperiodic) trajectories  $\{x_i(t), y_i(t)\}$  that are discretized over time  $t_1, t_2, \dots$  and ii) a user-specified, non-self-intersecting elastic band which wraps around a set of trajectories. The output is the number of edge segments in the band as a function of time. However, instead of using a braid representation to compute the stretching of the band, E-tec computes this stretching, and thus the topological entropy, *directly* by using a triangulation to detect all point-band collisions. The idea of using an advected dynamic triangulation to compute the growth rate of a band was first proposed by Marc Lefranc in the context of periodic orbits<sup>17–19</sup>.

## III. OVERVIEW OF E-TEC

We next provide an overview of the E-tec algorithm. (Details are given in Appendix A.) E-tec computes how an initial, closed, piecewise linear, non-self-intersecting rubber band in  $\mathbb{R}^2$  evolves under an ensemble of trajectories. The vertices of the band coincide with trajectories from the ensemble. When trajectories strike the band, they do not penetrate it but stretch it like a piece of elastic (Fig. 1a). In this manner, the band is stretched and folded, typically producing a growing number of edges wrapping around each other. Our algorithm tracks the configuration of the band. Care must also be taken in tracking when a trajectory detaches from an edge. This detachment results in two band edges returning taut (Fig. 1b), in much the same way a tight string will return taut once plucked (stretched) and released (undoing the stretching). Each band edge is assigned

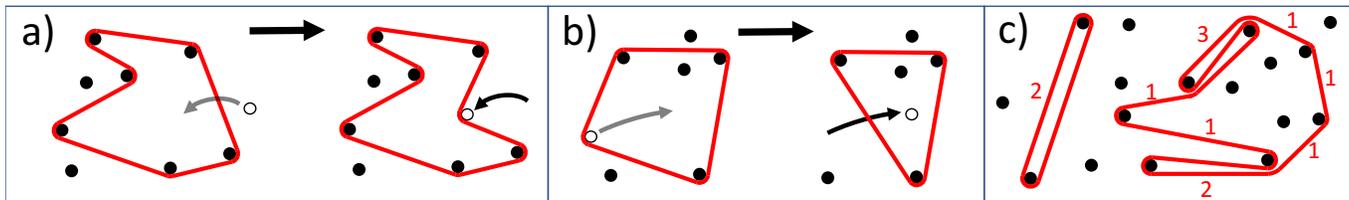


FIG. 1: **Band Deformation.** a) The white point strikes and deforms the band (red). b) The white point detaches from the band. Notice the band edge is taut after detachment. c) An initial rubber band stretched between two points on the left with edge weights displayed. A more complicated band on the right. The edge weights correspond to the number of times the band crosses an edge.

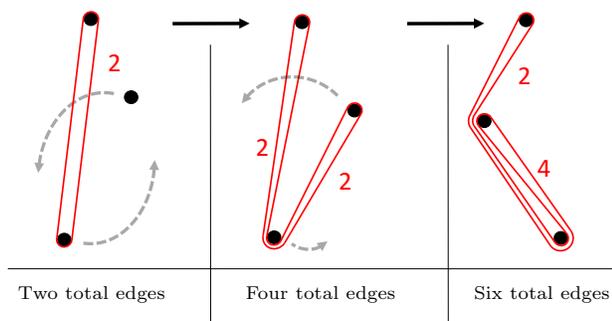


FIG. 2: **Edge Weights.** E-tec counts the number of edges of a rubber band as it is stretched by moving points. As the two bottom points rotate, the red band, initially wrapped around two points, is stretched and folded (left to right). E-tec tracks the growth of this band by assigning a weight to each edge corresponding to the number of times the band passes over this edge.

an integer weight  $\omega$  indicating the number of times the band stretches across it (Fig. 1c and Fig. 2). For chaotic advection, the total weight of the band will grow exponentially, as shown in Sect. IV. This exponential growth rate is a lower bound to the true topological entropy of the dynamical system. Even though the weight of all the edges grows exponentially, the number of unique edges is bounded.

E-tec efficiently tracks band growth by simply shifting edge weights to the appropriate edges when a point collides with the band. A key component of the algorithm is the detection of all relevant point-edge collisions. We achieve this by maintaining a triangulation of all trajectories for all times. These triangles are called *core* triangles. Each edge of the stretched band lies within the triangulation, so that each time a point strikes the band, the orientation of one of the core triangles will be inverted. We refer to this inversion as a triangle collapse. All band deformations will be detected since band edges remain in the core triangulation. The triangulation must be updated upon any triangle collapse. This update is *local* to the detection of each event, resulting in the rearrangement of edges and triangles near the collision only (illustrated in Fig. 3a). Similarly, the only edge weights that are shifted are those involved in the

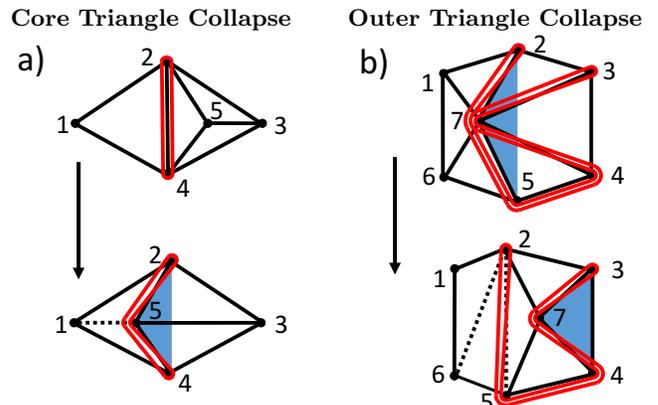


FIG. 3: **Events of E-tec Algorithm.** a) As point 5 moves left, triangle (2, 4, 5) collapses and inverts orientation. Two core triangles are re-triangulated, with the new edge shown as dashed. The initial edge weight of 2 for segment (2, 4) is shifted to segments (2, 5) and (4, 5). The blue-highlighted triangle (2, 4, 5) is the new outer triangle of point 5. It records which triangle collapse would be needed for the band to “snap back” taut, thereby undoing the collision. b) As point 7 moves to the right, outer triangle (2, 5, 7) collapses and the band edges (2, 7) and (5, 7) straighten into (2, 5). The three core triangles within pentagon (1, 2, 7, 5, 6) are reconfigured into three new core triangles (1, 2, 6), (2, 5, 6), and (2, 5, 7). Point 7 is still a candidate for future detachment, with new outer triangle (3, 4, 7), which also happens to be a core triangle.

collision. The update process is independent of both the number of points  $N$  and the number of triangles.

In addition to collisions, we need to detect when a trajectory detaches itself from a band edge. E-tec records which edges of the band are candidates for detachment by storing the triangle made up of the outer-most band edges attached to each point, i.e. the most recent edges to have struck a point. These triangles are called *outer* triangles. Some outer triangles are shown in blue in Fig. 3. Unlike the core triangles, the outer triangles do not form a triangulation of space. Rather, there is simply one outer triangle for each vertex crossed by the band. An outer triangle may coincide with a single core triangle, or it may intersect multiple core triangles. When a point detaches from the band, its corresponding outer triangle collapses and inverts its orientation. Thus, E-tec must track when both core and outer triangles collapse.

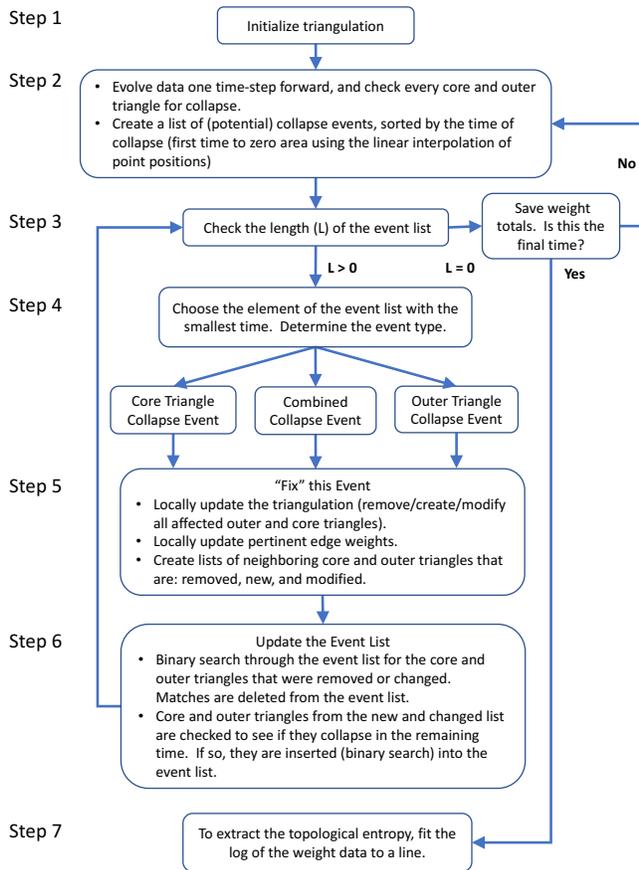


FIG. 4: **E-tec Algorithm Flowchart.** As described in Sec. III, E-tec employs computational geometry techniques for tracking the evolution of a piecewise-linear band. Full details are given in Appendix A.

In summary, there are two kinds of events that must be detected: the collapse of either a core or outer triangle. In the given time interval, these events are detected by finding the time for which their area first goes through zero. This time of first collapse is simply the appropriate root of the area quadratic polynomial, which is formed from the linear interpolation of triangle point positions. Once these events are detected, they are put in a time-sorted list and processed in order. Each event is “fixed” by locally updating the core triangulation, outer triangles, and edge weights. In the course of fixing an event, we may need to add or remove events from the event list. Event lists become large for densely-packed ensembles, though E-tec parses through each event and performs each subsequent triangulation update efficiently, as verified in the next section. A flowchart summarizing the E-tec algorithm is given in Fig. 4.

Finally, a numerical example of E-tec applied to real trajectory data is shown in Fig. 5.

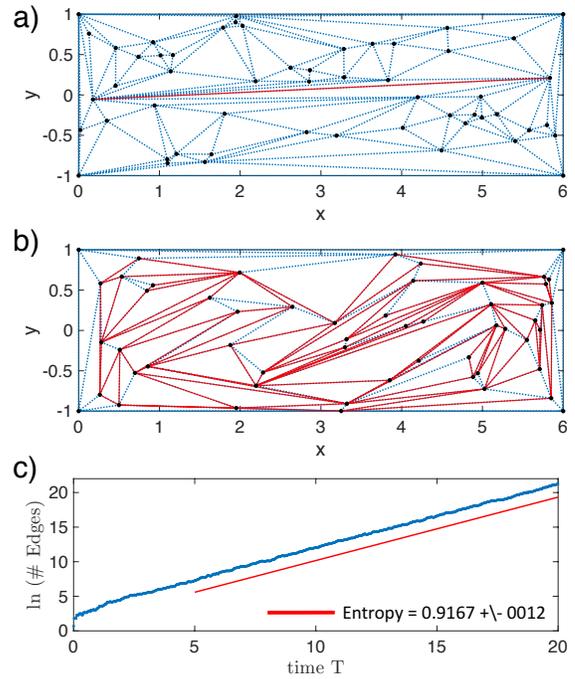


FIG. 5: **Numerical Example of an E-tec Implementation.** a) Initial data points with the band wrapped around two points (in red). The core Delaunay triangulation (in blue, dotted) is constrained to include the red band edge. b) Final data point positions at  $T = 20$ , the triangulation, and the stretched band evolved under the motion of the trajectories. Dynamics is given by model in Sect. IV with  $\tau_f = 0.96$ . c) E-tec output: the number of band edges as a function of time (blue). The slope of the best-fit line (red, dashed) is the topological entropy estimate. Please see online supplemental material for the related movie.

#### IV. E-TEC ALGORITHM VERIFICATION

In this section, we verify the E-tec algorithm by running E-tec on numerical trajectories sampled from a chaotic lid-driven cavity flow used to study chaotic advection<sup>20</sup>. We compare our results to lower bounds on topological entropy computed from two different methods; first, by a direct application of Eq. (1) to a growing material-line, and second, by a technique called homotopic lobe dynamics (HLD), which extracts symbolic dynamics from finite-length pieces of stable and unstable manifolds attached to fixed points of the fluid flow<sup>21–23</sup>.

##### A. Chaotic Lid-Driven Cavity Flow

The chaotic lid-driven cavity model<sup>20,24–26</sup> is a two-dimensional area-preserving flow defined over a 2D vertical cross-section of a rectangular cavity, extending vertically from  $-b \leq y \leq b$  and horizontally from  $0 \leq x \leq a$ . The flow,

$$\mathbf{V}(x, y, t) = \left( \frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x} \right) \quad (2)$$

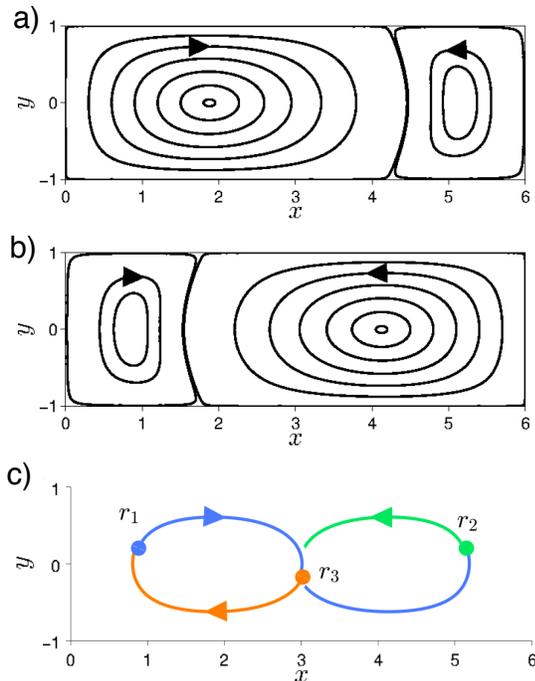


FIG. 6: **Dynamics of Chaotic Lid-Driven Cavity Flow.** We depict streamlines of the flow, Eq. (3). a) Motion under the first half-period,  $n\tau_f \leq t < (n + 1/2)\tau_f$ . b) Motion under the second half-period,  $(n + 1/2)\tau_f \leq t < (n + 1)\tau_f$ . c) Illustration of a period-three orbit  $r_i$ . Each color (blue, green orange) represents the trajectory evolving forward one period.

is defined in terms of a stream function  $\psi(x, y)$ . The stream function is an exact solution of the biharmonic equation  $\nabla^2 \nabla^2 \psi(x, y) = 0$  defined on the rectangular domain. The stream function is time-periodic with period  $\tau_f$  and is given explicitly by

$$\psi(x, y, t) = \begin{cases} U_1 C_1 f_1(y) \sin\left(\frac{\pi x}{a}\right) + U_2 C_2 f_2(y) \sin\left(\frac{2\pi x}{a}\right), & \text{for } n\tau_f \leq t < (n + 1/2)\tau_f, \\ -U_1 C_1 f_1(y) \sin\left(\frac{\pi x}{a}\right) + U_2 C_2 f_2(y) \sin\left(\frac{2\pi x}{a}\right), & \text{for } (n + 1/2)\tau_f \leq t < (n + 1)\tau_f, \end{cases} \quad (3)$$

where

$$f_k(y) = \frac{2\pi y}{a} \cosh\left(\frac{k\pi b}{a}\right) \sinh\left(\frac{k\pi y}{a}\right) - \frac{2k\pi b}{a} \sinh\left(\frac{k\pi b}{a}\right) \cosh\left(\frac{k\pi y}{a}\right), \quad k = 1, 2,$$

and

$$C_k = \frac{a^2}{2k\pi^2 b} \left[ \frac{a}{2k\pi b} \sinh\left(\frac{2k\pi b}{a}\right) + 1 \right]^{-1}, \quad k = 1, 2.$$

We follow Grover et al.<sup>20</sup> and assign  $U_1 = 9.92786$ ,  $U_2 = 8.34932$ ,  $a = 6$ , and  $b = 1$ . Fig. 6a and Fig. 6b show

streamlines for the two steady flows in Eq. (3). Each flow is separately integrable and is asymmetric in  $x$ , with a large vortex on one side and a smaller vortex on the other. The system alternates between each flow for a half-period  $\tau_f/2$ . It is this alternating flow that introduces positive topological entropy into the system.

When  $\tau_f$  is sufficiently large,  $\tau_f \geq \tau_f^* \approx 0.9553$ , there exists a period-three orbit,  $r_i$ ,  $i = 1, 2, 3$ , such that

$$M(r_1) = r_2, \quad M(r_2) = r_3, \quad M(r_3) = r_1, \quad (4)$$

where  $M$  is defined to be the flow map that evolves a point  $(x, y)$  forward to the point  $(x', y') = M(x, y)$  after a single period  $\tau_f$ . Fig. 6c shows the points  $r_i$  and their time evolution over one period. In the first half-period,  $n\tau_f \leq t < (n + 1/2)\tau_f$ , the two trajectories on the left swap positions in a clockwise fashion, while in the second half-period,  $(n + 1/2)\tau_f \leq t < (n + 1)\tau_f$ , the two trajectories on the right swap positions in a counterclockwise fashion. Grover et al.<sup>20</sup> characterize the  $r_i$  as a set of three strands braiding around one another in a nontrivial fashion. The presence of this braid guarantees the topological entropy is at least  $h_{po3} = 0.9624$ , the topological entropy which Boyland et al.<sup>27–29</sup> computed using the Bestvina-Handel train-track algorithm<sup>30</sup>. We note that this period-three orbit lives within a larger coherent set, a period-three island chain<sup>24</sup> when  $\tau_f$  is strictly greater than  $\tau_f^*$ .

## B. Period-Three Orbit and Convergence in Ensemble Size

Here we investigate the convergence of the E-tec algorithm by studying trajectories from the chaotic lid-driven flow with period  $\tau_f = 0.96$ , where we are guaranteed the existence of a period-three island chain<sup>20,31,32</sup>. As illustrated in Fig. 7d, no trajectory starting in an island leaves the island, and no trajectories enter. These islands braid around one another as they swap places in the same fashion depicted in Fig. 6c. In the analysis of Sect. IV A, each trajectory is sampled with time step  $\Delta t = 10^{-2}$  between points; this will be shown to be sufficient in Sect. V C.

First, we run E-tec on a set of three trajectories with the initial condition for each trajectory chosen in a different period-three island (Fig. 7a). We place an initial band around the right two points and observe exponentially growing band weights (Fig. 7b). At  $T = 15$  our estimate for the topological entropy is within 0.1% of the topological entropy guaranteed by the braid (Fig. 7c).

Next, we run E-tec on a set of 75 trajectories consisting of the 3 previously selected trajectories along with 72 randomly chosen ones. We calculate topological entropy by considering the time evolution of the same initial band (Fig. 7b). While the dynamics appear far more complicated than in Fig. 7a, our estimate of topological entropy is within fitting error to  $h_{po3} = 0.9624$  (Fig. 7c). Our results demonstrate that the periodic islands, and their braiding, are what drives most of the system entropy<sup>33,34</sup>. Furthermore, this demonstrates that for certain systems,

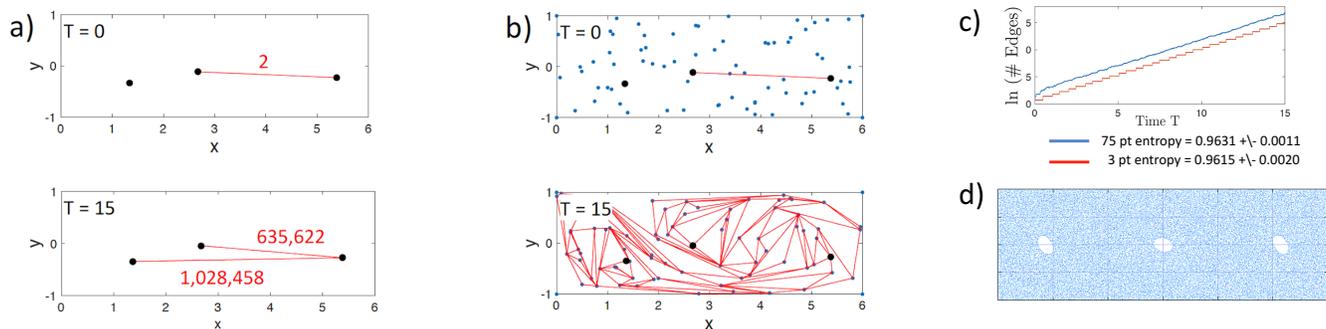


FIG. 7: **E-tec Analysis of the Chaotic Lid-Driven Cavity Flow.** We show E-tec results on trajectories governed by Eq. (3) with  $\tau_f = 0.96$ , guaranteeing the existence of a period-three orbit, seen in Fig 6c. a) We show E-tec results when considering only 3 points close to the period-three orbit and contained in period-three islands. We consider an initial band around the two right points (top). This band evolves (bottom) into a highly stretched band (edge weights in red) around all three points by  $T = 15$ . b) We consider the same 3 initial points, but add 72 random trajectories (top). The dynamics are more complex (bottom, weights omitted). c) The growth rate in the number of edges, i.e. our estimate of the topological entropy, for (a, red) and (b, blue) is the same. This indicates the entropy is driven by the period-three islands as also shown by Ref. 21. d) The coherent period-three islands, noted in Refs. 21, 20, and 24, are clearly seen in the Poincaré return map of a long-lived trajectory.

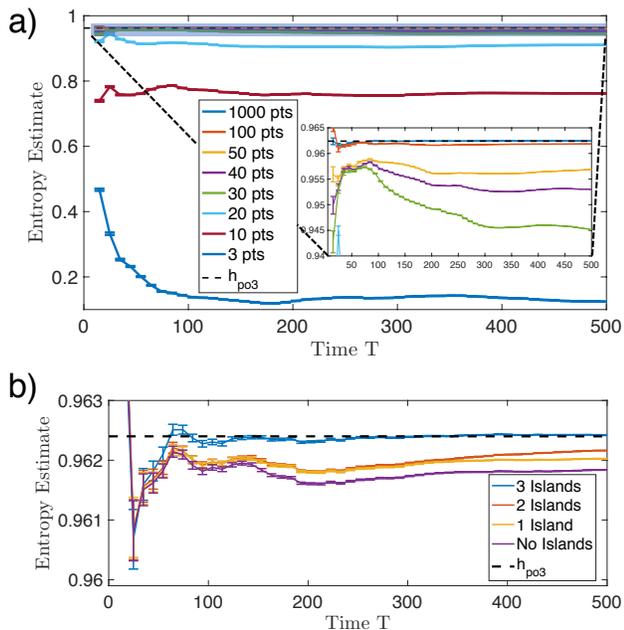


FIG. 8: **Convergence of E-Tec in the Length and Number of Trajectories.** a) We demonstrate convergence of E-tec to  $h_{po3} = 0.9624$  with increasing sample size and trajectory duration. The same initial band is stretched under ensembles of increasing size. All trajectories are sampled from outside the islands in the chaotic lid-driven flow with period-driving parameter  $\tau_f = 0.96$ . The entropy reported at time  $T$  is the fitting slope and 95 percent confidence interval to the log of the total number of edge weights over time  $t$  for the range  $t \in [5, T]$ . b) E-tec output using the 100 point ensemble with a single trajectory added into one, two, and three of the periodic islands.

topological approaches such as E-tec (as well as braiding approaches) are capable of producing accurate estimates of topological entropy with only a small set of carefully chosen trajectories.

Although the coherent sets for our example were straightforward to locate, for other examples and practical applications, coherent sets may be harder to identify. As such, there is no guarantee trajectories from coherent sets, whose dynamics might be governing the topological entropy of the system, will be sampled appropriately. To investigate how E-tec would perform under conditions like this, we examine our ability to accurately recover the topological entropy when randomly sampling initial conditions uniformly in space, but removing any point chosen in the period-three islands. E-tec was run on increasingly larger but nested sets of such trajectories. That is, the points chosen in the 20 trajectory analysis contain all of the points in the 10 trajectory analysis, and so forth. As shown in Fig. 8a, E-tec converges rather quickly in the number of points to the topological entropy lower bound guaranteed by the period-three islands. Convergence results with respect to the length of trajectories may be problematic for many practical applications where long-lived trajectories are not guaranteed. Like any result from a fit in time, estimates may fluctuate based on the interval where fitting is performed.

Finally, in Fig. 8b, we investigate the E-tec convergence using the 100 point ensemble in Fig. 8a by adding additional points in each of the three islands. E-tec performs increasingly better as the island points are added. The result with no island points, given in Fig. 8a, is then taken as a worst-case scenario. This assures our confidence in E-tec results as ensemble sizes are increased in Sect. IV C.

### C. Topological Entropy for Range of Period Driving Parameter $\tau_f$

With confidence in E-tec's ability to characterize topological entropy when  $\tau_f = 0.96$ , we next explore how the

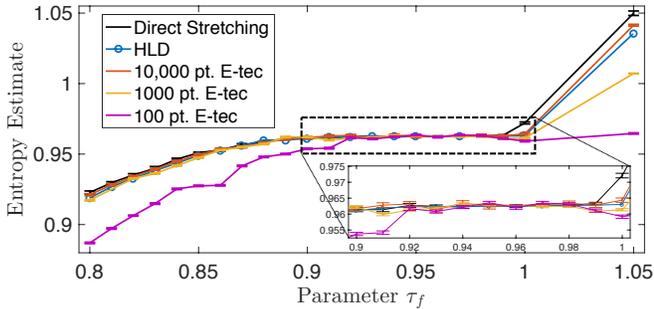


FIG. 9: **Verification of E-tec for Increasing  $\tau_f$ .** E-tec topological entropy results over a range of  $\tau_f$  values using increasing ensemble sizes. We compare to the estimate of topological entropy from directly stretching a material line<sup>21</sup> and through another topological technique, homotopic lobe dynamics<sup>22,23</sup>.

topological entropy changes as  $\tau_f$  varies. As mentioned previously, the period-three orbit is born at  $\tau_f^* \approx 0.9553$  and persists for larger values. Thus, entropy for values  $\tau_f < \tau_f^*$  will be bounded above by the braiding entropy of  $h_{po3} = 0.9624$ , while  $h_{po3}$  remains a lower bound for  $\tau_f > \tau_f^*$ . In all cases, the same initial band is chosen and evolved forward.

As shown in Fig. 9, our estimate of topological entropy using E-tec is within error of the direct calculation of material-line stretching when  $0.85 \leq \tau_f \leq 0.98$  and the number of data points is at least 1000. For  $\tau_f < \tau_f^*$ , there are no known island chains that drive the complexity. Despite this, E-tec performs well here, as shown in Fig. 9. For low values of  $\tau_f$ , when  $\tau_f < 0.85$ , E-tec produces an estimate slightly less than that of direct stretching but consistent with the value produced by HLD. But E-tec’s discrepancy becomes smaller with increasing numbers of samples. For high values of  $\tau_f$ , when  $\tau_f > 0.98$ , both E-tec and HLD produce lower estimates for topological entropy than the calculated direct stretching value. We note that E-tec with 1000 trajectories still produces estimates consistent with HLD, and with 10,000 trajectories E-tec exceeds the HLD estimate but is still below the direct material-line stretching.

To more clearly see what drives the increase in entropy for high values of  $\tau_f$ , we show the band stretched by E-tec for three different values of  $\tau_f$  each computed from a set of 1000 independently chosen trajectories (see Fig. 10). Exponential stretching and folding is present in all tested parameter values, though Fig. 10 shows the band is stretched in a more complex fashion at higher  $\tau_f$  values. Here, additional island chains emerge<sup>24</sup> resulting in secondary folding<sup>35</sup> that seems less “smooth.” This secondary folding results in kinks near the islands that propagate forward, which in turn are further stretched under the dynamics. These small areas with kinks give significant contribution to the topological entropy, but because the entropy estimates (Fig. 9) were generated from uniformly random samples, these highly-kinked regions may remain undersampled. As such, a good por-

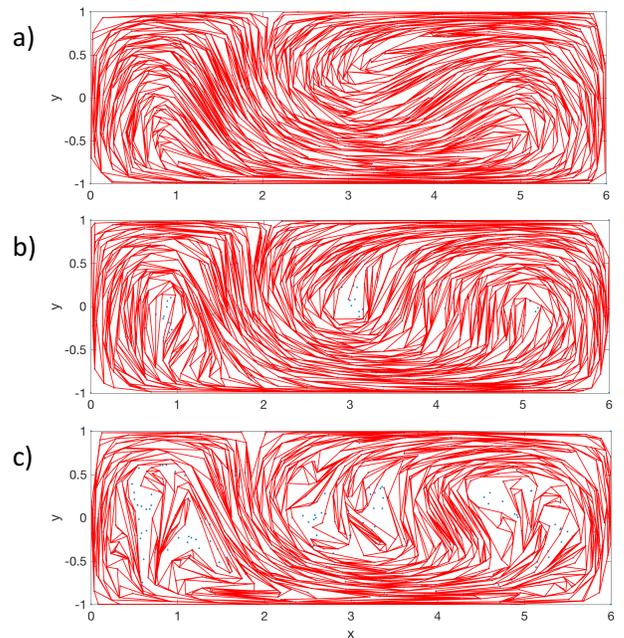


FIG. 10: **Stretched Band Visualization.** E-tec band stretching due to flow advection for period-driving parameters a)  $\tau_f = 0.80$ , b)  $\tau_f = 0.96$ , and c)  $\tau_f = 1.05$ . All bands are stretched by ensembles of 1000 uniformly distributed trajectories.

tion of the stretching may remain undetected by E-tec in Fig. 10c.

## V. E-TEC ROBUSTNESS

In this section, we investigate the robustness of E-tec’s results. More specifically, we examine how E-tec’s ability to correctly estimate topological entropy is impacted by the choice of initial band and the time-step associated with trajectories. Finally, we discuss how the E-tec algorithm’s run-time scales with the duration and number of sampled trajectories.

### A. Robustness to Choice of Initial Band

We make the following conjecture: if all trajectories reside in the same ergodic component then the choice of initial band does not affect the topological entropy computed by E-tec as long as the trajectories are sufficiently long. Figure 11 supports this conjecture. All initial bands eventually become stretched at the same rate despite some differences at early times. Adjacent points may remain close for some time, though the chaotic nature of the flow causes nearby trajectories to eventually diverge, thereby making the band’s deformation inevitable. Thus, as long as it is possible to obtain sufficiently long trajectories within a single ergodic component, E-tec’s topological entropy calculation appears to

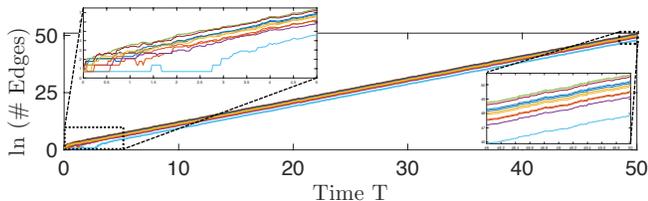


FIG. 11: **Initial Bands.** E-tec output is the logarithm of the sum of edge weights as a function of time. E-tec’s estimate of topological entropy is the best-fit linear slope through this data. Here we show 10 different outputs from E-tec for the same set of 100 trajectories. In each case, we chose a different pair of points around which to stretch our band. Despite some initial differences in the increase in edge weights due to initial adjacent points staying close to one another (*left inset*), eventually all the bands grow at similar rates (*right inset*). When we fit the exponential growth rate, starting at  $T = 5$ , we find the values for each of the 10 bands agree within 5 decimal places and average out to 0.9617.

be invariant to the choice of initial band.

Some chaotic flows have more than one ergodic component, or a mixture of ergodic and non-ergodic regions. This is true of the model flow in Fig. ??d. In such systems, the choice of initial band will impact the topological entropy estimate. For example, a band placed entirely in one of the test flow’s period-three islands (Fig. 7d) will undergo no significant stretching under the flow and thus yield zero topological entropy.

In practice, to make sure all ergodic components are sampled, it is prudent to check that the final band stretches around nearly all of the data points. Alternatively, one could sample many initial bands taking the maximum growth rate of all sampled bands as the best estimate of the entropy<sup>36</sup>. E-tec is fast enough to run multiple bands in ensembles of fewer than  $10^6$  trajectories in a reasonable time<sup>37</sup>.

## B. Algorithm Scaling and FTBE Comparison

The computational runtime of E-tec is linearly proportional to the *duration* of the trajectories. This is because number of edges tracked by E-tec is constant, and it is only the values of the weights that grow exponentially in time. This scaling is the same as the FTBE calculation (which may be calculated using the freely available Matlab package `braidlab`<sup>38</sup>) and stands in contrast to algorithms that precisely evolve a material-curve forward, which requires inserting exponentially more points to maintain sufficient point density<sup>39</sup>.

One advancement we have made over the FTBE calculation is the run-time scaling with respect to the number of trajectories used (see Fig. 12). The FTBE calculation scales quadratically in the number of trajectories  $N$  due to the braid approach requiring  $N^2$  algebraic generators per unit time step<sup>12</sup>. E-tec scales asymptotically as  $O(N \log N)$  due to creating and maintaining a time-sorted list of collapse events, the length of which

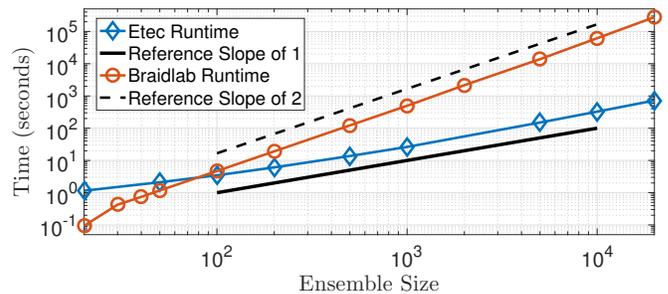


FIG. 12: **E-tec Runtimes.** Runtime comparison of `braidlab`, a freely available Matlab package implementing the FTBE calculation, and E-tec. Both used the same trajectories from the chaotic model flow for  $\tau_f = 0.96$ . All computations were completed using a 2.8 GHz Intel Core i7 processor.

scales linearly with the number of points. As a practical matter, the E-tec runtime for small to moderate ensembles (roughly up to 5,000 trajectories) is dominated by the linear behavior (Fig. 12). One illustrative example highlighting the runtime difference between the two algorithms is rigid rotational flow. While an admittedly special case, there would be no new collapse events (except for ones associated with the boundary) making E-tec very fast, whereas the number of braid generators needed would be proportional to  $N^2$ .

One advantage the braid approach has over E-tec is that once the braid is extracted from the trajectory data, it may be applied to any initial band. E-tec only propagates a single curve forward. However, for topological entropy calculations, a single sufficiently long curve is typically sufficient (as evidenced in Fig. 11).

## C. Robustness to Step Size $\Delta t$

Because E-tec is based on the computational analysis of evolving trajectories, it is necessary to consider discretized time. We next investigate how the trajectory time step  $\Delta t$  affects the entropy calculation and show that E-tec returns trustworthy results even when poorly resolved trajectories are used as input. We use two ensembles of trajectories (of sizes 100 and 1000) sampled at a fine scale using the same reference time step of  $\Delta t^* = 10^{-4}$  to generate two reference topological entropies  $h_t^*$ . We then vary the time step  $\Delta t$  (keeping the trajectories the same) and compute both ensembles’ corresponding  $h_t$ . The effect of time step  $\Delta t$  is quantified by computing the relative error

$$\left| 1 - \frac{h_t}{h_t^*} \right|, \quad (5)$$

which is plotted in Fig. 13. The data shows the relative error grows linearly with the time step  $\Delta t$ . As the trajectory information is input into E-tec using larger step sizes, we detect more events between steps. E-tec detects events individually for all values of  $\Delta t$ , but the order in

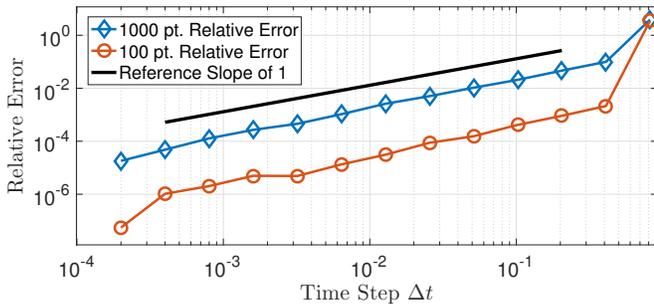


FIG. 13: **Relative Error as a Function of Step Size.** The effect of time step  $\Delta t$  on the relative error in the topological entropy calculation with respect to the reference time step  $\Delta t^* = 10^{-4}$ . Graph displays calculations done on two separate ensembles of size 1000 and 100.

which these events are detected is potentially different as  $\Delta t$  increases, due to the differences in the interpolation of trajectories. In fact, undersampled trajectory data may lead to entirely different events. This explains the larger relative errors for the 1000 trajectory ensemble; at higher point densities, there are simply more events that E-tec must resolve, resulting in more erroneous and misordered event detections. Despite this, Fig. 13 shows that the E-tec error due to step size is still relatively small. It is comparable to (or smaller than) the error due to other sources, such as trajectory length and ensemble size (Fig. 8a), for  $\Delta t < 10^{-2}$ , at least for smaller ensemble sizes.

## VI. CONCLUSION

We introduced the Ensemble-based Topological Entropy Calculation (E-tec), an algorithm that computes topological entropy in a planar flow from an ensemble of system trajectories. We verified E-tec’s convergence to the correct topological entropy with increasing numbers of trajectories on a highly chaotic, lid-driven cavity flow. E-tec’s performance was shown to be robust with respect to the choice of initial band, as well as changes in the time sampling interval ( $\Delta t$ ). Notably, we have shown that E-tec’s runtime scales as  $O(N \log N)$ , where  $N$  is the number of trajectories in the ensemble.

Our work suggests several further directions for the analysis of trajectories with E-tec, which we intend to explore in future studies. First, we shall seek to extend E-tec to three dimensions and higher. Braiding theory, the basis for FTBE calculations, cannot be readily generalized to higher dimensions<sup>40</sup>. The computational geometry framework in which E-tec is based might perhaps be more naturally extended<sup>17–19</sup>. Instead of a rubber band in a planar flow, we would consider a two-dimensional rubber sheet stretched around a collection of points in a three-dimensional flow. A 3D triangulation may still be used to track point-face or edge-edge collisions, and the rubber sheet may be chosen as one of the faces in the

initial triangulation. As the points evolve in time, they carry the sheet along with them, stretching and folding it so that its growth reflects the flow complexity. Though there clearly remain some significant challenges to executing this generalization to three dimensions, we anticipate a host of interesting theoretical opportunities that this route may provide.

Second, by tracking all the trajectories in concert, we believe E-tec’s algorithm may be naturally adapted towards identifying and tracking coherent sets and other emergent structures. Finally, some special challenges arise when working with trajectories derived from experimental observations. These include noise in the position data and trajectories that are only observed for part of the full time interval. We hope to adapt E-tec to some of the challenges facing trajectories derived from experiments.

## ACKNOWLEDGMENTS

This work was supported in part by the US DOD, ARO grant W911NF-14-1-0359 under subcontract C00045065-4. ER also received support from the National Science Foundation under grant number DMS-1331109.

## Appendix A: The E-tec Algorithm Outline

This Appendix outlines our implementation of the E-tec algorithm.

**Input:** The following inputs are required by the algorithm:

1. The precomputed (or experimentally measured) trajectories.
2. An initial, non-self-intersecting rubber band stretched around a sequence of data points, specified by the set of edges connecting pairs of data points. This is represented as a counterclockwise ordering of this set of points. It is often convenient to choose an initial band that encloses two distant points.

**Output:** E-tec tracks the evolution of the band, as we will describe below, and outputs:

1. The state of the stretched rubber band as a function of time, recorded as a (core) triangulation of all data points and a set of edge weights of this triangulation.
2. The sum of all band edge weights  $\omega$  as a function of time.
3. The exponential growth rate of the band (topological entropy), determined by the slope of the best fit line for the  $\ln(\omega)$  vs. time graph.

**Data structures:** E-tec maintains the following data structures as a function of time:

1. A *core* triangulation of all data points in the plane.
2. The weights on each edge in the triangulation. (Non-zero weighted edges constitute the stretched

rubber band.)

- For each relevant data point, the *outer band* triangle (abbreviated *outer triangle*) records the outermost wrapping of the rubber band around that point. (See the blue shaded triangles in Fig. 3.) During the algorithm’s run, the outer triangle represents the piece of rubber band that has struck the point most recently and hence is a candidate for detachment at a future time. For example, upon inspection of vertex 7 in Fig. 3b, we may deduce that of all the red band edges attached to it, the two that created the largest angle would be the ones to snap back and revert to a single edge. Specifically, edge (2, 5) will snap back taut if triangle (2, 5, 7) changes orientation. Notice that outer triangles are not necessarily contained in the set of all core triangles.

**Steps:** We outline the key steps taken by E-tec in tracking the evolution of a rubber band. These steps are summarized in the Fig. 4 flowchart.

- We first initialize the core triangulation using a constrained Delaunay triangulation<sup>41</sup> of the initial points (with the initial placement of the rubber band as the constraint). See Fig. 5a.

In steps (2-6) we evolve the state of the system (core triangulation, weights, and outer triangles) forward using the next time-slice in the trajectory data as input. Notice that E-tec does not need the whole trajectory at once in order to evolve the triangulation forward, and therefore could be used in real-time during experimental data collection.

- For each core and outer triangle in the current state of the system, we use the linear interpolation of point positions to determine if and when a triangle will pass through zero area during this time step. These collapse events are sorted by time into an event list.
- If the event list is non-empty, we go to step 4 and determine the event type of the next collapse event. If the event list is empty, we then add up the weights of every edge to get the current total weight  $\omega$  of the band, and store this value. This acts as a proxy for the length of the band, and grows with the same exponential rate in time. If we are at the final time of the trajectory data, we end by analyzing the accumulated weight data in step 7. Otherwise, we move on to the next trajectory time in step 2.
- A collapse event can be one of three general types: a core triangle collapse (Fig. 3a), an outer triangle collapse (Fig. 3b), or a combined core and outer triangle collapse (see Fig. 14 for an illustration). While the specifics of how the three types of collapse events are handled are different, the broad strokes, as seen in step 5, are the same.
- For each collapse event type, there is a general template for adding, removing, and/or modifying the core and outer triangles that are adjacent to the collapsing triangle. Crucially, this process is local,

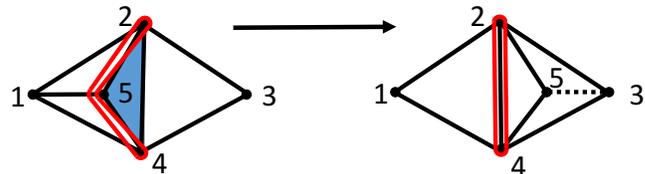


FIG. 14: **Combined Core and Outer Triangle Event.** In blue is a combined core and outer triangle (2,4,5). As point 5 moves to the right and this triangle collapses, the band returns taut around segment (2,4). Three core triangle (1,2,4), (2,3,5), and (3,4,5) are reconfigured, with the new edge shown as dashed. Collapsed triangle (2,4,5) (previously shaded) remains as a core triangle.

and the number of operations is bounded and does not grow with the number of trajectories.

- The local deletions, creations, and modifications of core and outer triangles that result from handling a collapse event potentially affect the overall event list for this time-step. First we consider the deleted and modified core and outer triangles. If, before modification, they have a time-to-zero-area that is in the remaining fraction of the current time-step, then we search for and remove them from the event list. Next we consider the new and modified core and outer triangles. If, after modification, they will collapse in the remaining time-step, we search for the proper position to insert them into the sorted event list. Both searches are binary, and constitute one of the two aspects of the algorithm that give us  $O(N \log N)$  computational complexity ( $O(\log N)$  for binary search and  $O(N)$  searches per time-step). After modifying the event list, we return to step 3.
- Approximate the topological entropy by computing the exponential growth rate for the total weight over time.

- R. Ding and J. Li, *Physics Letters A* **364**, 396 (2007).
- J.-P. Eckmann and D. Ruelle, in *Turbulence, Strange Attractors And Chaos* (World Scientific, 1995) pp. 447–449.
- H. Arbabi and I. Mezić, arXiv preprint arXiv:1611.06664 (2016).
- S. Candelaresi, D. I. Pontin, and G. Hornig, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 093102 (2017).
- R. Bowen, *Transactions of the American Mathematical Society* **154**, 377 (1971).
- H. Aref, J. R. Blake, M. Budišić, S. S. Cardoso, J. H. Cartwright, H. J. Clercx, K. El Omari, U. Feudel, R. Golestanian, E. Gouillart, *et al.*, *Reviews of Modern Physics* **89**, 025007 (2017).
- C.-Y. Lee, C.-L. Chang, Y.-N. Wang, and L.-M. Fu, *International journal of molecular sciences* **12**, 3263 (2011).
- J. R. Ledwell, A. J. Watson, and C. S. Law, *Nature* **364**, 701 (1993).
- J.-L. Thiffeault, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **20**, 017516 (2010).
- D. H. Kelley and N. T. Ouellette, *Nature Physics* **7**, 477 (2011).
- N. T. Ouellette and J. P. Gollub, *Physics of Fluids* **20**, 064104 (2008).
- M. Budišić and J.-L. Thiffeault, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 087407 (2015).

- <sup>13</sup>M. R. Allshouse and J.-L. Thiffeault, *Physica D: Nonlinear Phenomena* **241**, 95 (2012).
- <sup>14</sup>L.-S. Young, in *Entropy* (Princeton University Press Princeton, NJ, 2003) pp. 313–327.
- <sup>15</sup>S. Newhouse and T. Pignataro, *Journal of statistical physics* **72**, 1331 (1993).
- <sup>16</sup>B. R. Hunt and E. Ott, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 097618 (2015).
- <sup>17</sup>M. Lefranc, *Topology and Dynamics of Chaos: In Celebration of Robert Gilmore’s 70th Birthday* **84**, 205 (2013).
- <sup>18</sup>M. Lefranc, *Physical Review E* **74**, 035202 (2006).
- <sup>19</sup>M. Lefranc, P.-E. Morant, and M. Nizette, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **366**, 559 (2008).
- <sup>20</sup>P. Grover, S. D. Ross, M. A. Stremler, and P. Kumar, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **22**, 043135 (2012).
- <sup>21</sup>S. Sattari, Q. Chen, and K. A. Mitchell, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **26**, 033112 (2016).
- <sup>22</sup>K. A. Mitchell, *Physica D: Nonlinear Phenomena* **238**, 737 (2009).
- <sup>23</sup>K. A. Mitchell, *Physica D: Nonlinear Phenomena* **241**, 1718 (2012).
- <sup>24</sup>M. A. Stremler, S. D. Ross, P. Grover, and P. Kumar, *Physical review letters* **106**, 114101 (2011).
- <sup>25</sup>P. Rao, A. Duggleby, and M. A. Stremler, *Journal of Fluids Engineering* **134**, 041203 (2012).
- <sup>26</sup>V. Meleshko and A. Gomilko, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 460 (The Royal Society, 2004) pp. 807–819.
- <sup>27</sup>P. L. Boyland, H. Aref, and M. A. Stremler, *Journal of Fluid Mechanics* **403**, 277 (2000).
- <sup>28</sup>P. Boyland, M. Stremler, and H. Aref, *Physica D: Nonlinear Phenomena* **175**, 69 (2003).
- <sup>29</sup>P. Boyland, *Topology and its Applications* **58**, 223 (1994).
- <sup>30</sup>M. Bestvina and M. Handel, *Annals of Mathematics* **135**, 1 (1992).
- <sup>31</sup>J.-L. Thiffeault and M. D. Finn, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **364**, 3251 (2006).
- <sup>32</sup>E. Guillard, J.-L. Thiffeault, and M. D. Finn, *Physical Review E* **73**, 036311 (2006).
- <sup>33</sup>T.-Y. Li and J. A. Yorke, *The American Mathematical Monthly* **82**, 985 (1975).
- <sup>34</sup>O. Sharkovsky, *Ukrainskij matematicheskij zhurnal* **16**, 61 (1964).
- <sup>35</sup>S. E. Tumas and J.-L. Thiffeault, *Procedia IUTAM* **7**, 117 (2013).
- <sup>36</sup>J.-L. Thiffeault, *Physical review letters* **94**, 084502 (2005).
- <sup>37</sup>An alternative approach to choosing a single initial band is to evolve a “web” of initial bands that covers the entire initial triangulation. This guarantees that all ergodic components sampled by the data will be included.
- <sup>38</sup>J.-L. Thiffeault, arXiv preprint arXiv:1410.0849 (2014).
- <sup>39</sup>Z. You, E. J. Kostelich, and J. A. Yorke, *International Journal of Bifurcation and Chaos* **1**, 605 (1991).
- <sup>40</sup>M. R. Allshouse and T. Peacock, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 097617 (2015).
- <sup>41</sup>C. D. Toth, J. O’Rourke, and J. E. Goodman, *Handbook of discrete and computational geometry* (Chapman and Hall/CRC, 2017).