A global study of enhanced stretching and diffusion in chaotic tangles

Darin Beigie,[a] Anthony Leonard,[b] and Stephen Wiggins[c]
California Institute of Technology, Pasadena, California 91125

(Received 28 August 1990; accepted 7 December 1990)

A global, finite-time study is made of stretching and diffusion in a class of chaotic tangles associated with fluids described by periodically forced two-dimensional dynamical systems. Invariant lobe structures formed by intersecting global stable and unstable manifolds of persisting invariant hyperbolic sets provide the geometrical framework for studying stretching of interfaces and diffusion of passive scalars across these interfaces. In particular, the present study focuses on the material curve that initially lies on the unstable manifold segment of the boundary of the *entraining turnstile* lobe. A knowledge of the stretch profile of a corresponding curve that evolves according to the unperturbed flow, coupled with an appreciation of a symbolic dynamics that applies to the *entire* original material curve in the perturbed flow, provides the framework for understanding the mechanism for, and topology of, enhanced stretching in chaotic tangles. Secondary intersection points (SIP's) of the stable and unstable manifolds are particularly relevant to the topology, and the perturbed stretch profile is understood in terms of the unperturbed stretch profile approximately repeating itself on smaller and smaller scales. For sufficiently thin diffusion zones, diffusion of passive scalars across interfaces can be treated as a one-dimensional process, and diffusion rates across interfaces are directly related to the stretch history of the interface. An understanding of interface stretching thus directly translates to an understanding of diffusion across interfaces. However, a notable exception to the thin diffusion zone approximation occurs when an interface folds on top of itself so that neighboring diffusion zones overlap. An analysis which takes into account the overlap of nearest neighbor diffusion zones is presented, which is sufficient to capture new phenomena relevant to *efficiency of mixing*. The analysis adds to the concentration profile a saturation term that depends on the distance between neighboring segments of the interface. Efficiency of diffusion thus depends not only on efficiency of stretching along the interface, but on how this stretching is distributed relative to the distance between neighboring segments of the interface.

I. INTRODUCTION

The study of fluid mixing under chaotic advection encourages a global, finite-time analysis of dynamics. Such analysis requires recasting some standard methods of investigation. For example:

(i) Much of dynamical systems analysis, e.g., the definition of chaos and many of the tools for measurement such as Lyapunov exponents, involves infinite-time concepts; however, infinite-time measures may have little relevance to fluid mixing on finite time scales.

(ii) Typically one establishes analytically the presence of chaos in two-dimensional flows via the horseshoe map construction. The chaotic motion, and the symbolic dynamics used to describe the motion, thus applies to an invariant Cantor set in phase space. To study fluid mixing, however, one is less interested in the dynamics of a Cantor set of points than in the motion of entire interfaces or material blobs.

(iii) Some common measures of mixing, such as efficiency of stretching (see Ottino[1,2]), are local in nature. The study of enhanced stretching is greatly facilitated by a consideration of the global geometry underlying the chaotic zones.

We present results of a global, finite-time study of stretching and diffusion in a class of chaotic flows: two-dimensional fluids that evolve according to time-periodic velocity fields. We specialize to near-integrable systems whose unperturbed solution contains a hyperbolic fixed point connected to itself by one or more homoclinic orbits, or a pair of hyperbolic fixed points connected to each other by two or more heteroclinic orbits. In particular, we consider two fluid flows induced by oscillating vortex pairs as prototypes of fluid mixing in open and closed flows. Though we consider a highly specialized class of velocity fields, we wish to emphasize the robustness of the analysis. For example, the framework for studying global dynamics within chaotic tangles via invariant manifolds extends to velocity fields with more complicated time dependences,[3] to certain classes of higher-dimensional flows,[4] and to cases with an arbitrary number of invariant hyperbolic sets of flows that need not be near integrable[5] (simple examples of these extensions include two-frequency quasiperiodic velocity fields, and three-dimensional fluids with normally hyperbolic invariant 1-tori rather than hyperbolic fixed points). For concreteness, however, we focus on a simple class of velocity fields.

Consider the two-dimensional fluid flows induced by (i) a pair of equal and opposite point vortices and (ii) a pair of identical point vortices oscillating periodically in response to a time-periodic strain-rate field (see Rom-Kedar *et al.*[6]

[a] Department of Physics.
[b] Department of Aeronautics.
[c] Department of Applied Mechanics.

 0899-8213/91/051039-12$02.00

for the first flow, which should make the equations of motion for the second flow apparent). We henceforth refer to these two flows as the open and closed flow, respectively. The streamlines in the comoving and corotating frames, respectively, of the unperturbed systems are shown in Fig. 1(a), and the perturbing strain-rate field is described in the comoving and corotating frames of the perturbed system by the streamfunction $\psi(x,y) = \epsilon xy \sin(2\pi t)$, where ϵ is the strain rate. Periodic forcing drastically alters the dynamics near the unperturbed separatrices, and it is convenient to study the dynamics by sampling trajectories periodically in time (with period equal to that of the strain-rate field, $\Delta t = 1$), which defines a Poincaré map, denoted by P_ϵ for the perturbed system. For small enough perturbations the hyperbolic fixed points of the map persist, and if their global stable and unstable manifolds intersect once they intersect in a countable infinity of points to produce the boundary of a complicated two-dimensional lobe structure (the so-called heteroclinic/homoclinic tangle) that is invariant under the Poincaré map [see Fig. 1(b)]. With each application of P_ϵ (i.e., from one time sample to the next), lobes map from one to another within the invariant lobe structure in an orientation-preserving manner, and this provides a global picture of the dynamics, i.e., of *transport in phase space* (which is identical to transport in *physical* space in the case of chaotic advection). There is a wealth of features of chaotic dynamics that can be studied within this framework; we concentrate here on the stretching of interfaces and the diffusion of passive scalars across these interfaces.

II. INTERFACE DYNAMICS

A. Lobe dynamics

We first provide a more detailed description of the global dynamics in chaotic tangles with the help of Fig. 2. An invariant core boundary, approximately equal to the unperturbed core boundary formed by the separatrices, is defined by segments of the stable and unstable manifolds, as shown by the dashed lines. The stable and unstable manifolds intersect in two classes of points: *primary intersection points* (PIP's) (marked by dots) and *secondary intersection points* (SIP's) (marked by crosses). The nature of these two types of intersection points should be clear from the figure (see Wiggins[7] for a discussion of intersection points); both the PIP's and SIP's will be seen to be relevant to understanding the topology of enhanced stretching. There are two types of lobes, which we distinguish as *entraining* and *detraining*: the mth ($m \geq 1$) entraining lobe $E(m)$ is *entrained into* the core

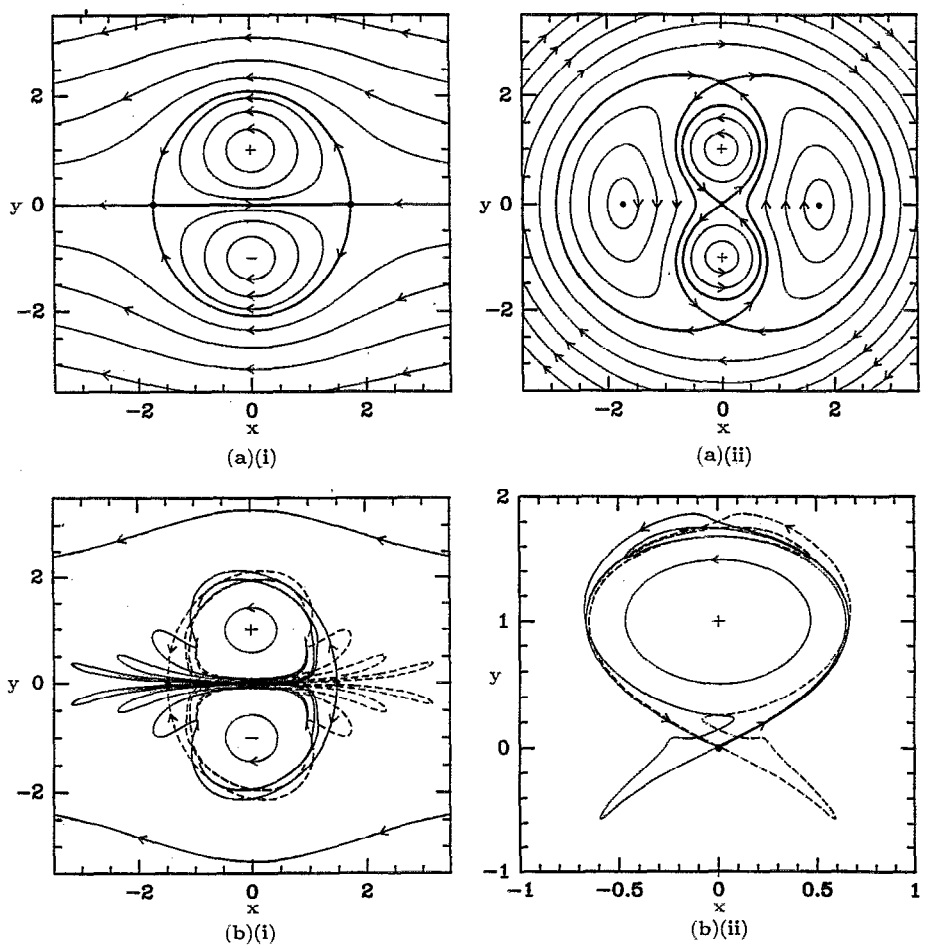

FIG. 1. (a) The streamlines of the unperturbed (i) open and (ii) closed flows with the vortices at $(0, \pm 1)$. The fixed points are marked by dots and the separatrices are shown in boldface. (b) The (i) heteroclinic and (ii) upper homoclinic tangles of the perturbed flow. The persisting hyperbolic fixed points are marked by dots, and their stable and unstable manifolds are marked by dashed and solid lines, respectively; other invariant curves are marked by solid lines. The closed flow has several tangles, and we focus on the homoclinic tangle associated with the innermost separatrix. See Fig. 5 for the perturbed flow parameters.

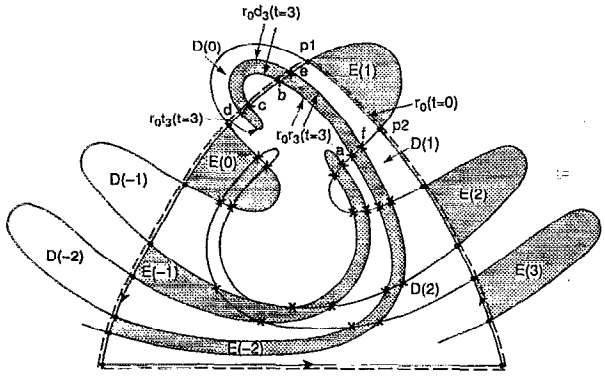

FIG. 2. The upper ($y \geq 0$) invariant lobe structure of the perturbed open flow (the system is symmetric about $y = 0$).

(mapped from outside to inside the core) upon the mth iterate of P_ϵ; the mth ($m > 1$) detraining lobe $D(m)$ is *detrained from* the core (mapped from inside to outside) upon the mth iterate of P_ϵ (in a similar manner one defines the lobes for $m < 0$, as should be clear from Fig. 2). One pictures transport in phase space in Fig. 2 by recognizing that the lobes map according to

$$P_\epsilon(E(m)) = E(m-1)$$
$$P_\epsilon(D(m)) = D(m-1) \tag{1}$$

(these transport equations follow from the invariance of the manifolds, the orientation-preserving nature of the Poincaré map, and the choice of sampling, as explained in Rom-Kedar *et al.*[6]). We refer to $E(1)$ and $D(1)$ as *turnstile lobes*, and these lobes are the *only* mechanism for transport in and out of the core under P_ϵ. The segments of the lobe boundaries formed by the stable manifold behave asymptotically in a simple manner: they asymptote to the left hyperbolic fixed point and contract asymptotically exponentially in time. The segments of the lobe boundaries formed by the unstable manifold behave in a complicated manner, and it is these interfaces whose dynamics we wish to study. Indeed, an understanding of the stretch properties of these interfaces provides a good understanding of stretching throughout the tangle region, since the unstable manifold acts somewhat like an attractor.[3,6] Of course, since the dynamical system describing the fluid motion is Hamiltonian (the fluid is incompressible), the unstable manifold is not truly an attractor; however, the strong stretching and contraction of the fluid lobes implies that, though area elements do not shrink, they tend to be stretched in one direction and contracted in another such that the unstable manifold dominates the evolution of material curves in the tangle region.

B. Stretch profiles and a symbolic dynamics for the entire interface

From the way the lobes map from one to another, it suffices to consider the segment of the unstable manifold along the boundary of the *turnstile* lobes; for simplicity we consider the material curve that starts off ($t = 0$) on the segment of the unstable manifold that bounds the entraining turnstile lobe, which we denote by $r_0(t)$ (the end points of this line segment at $t = 0$ are the PIP's $p1$ and $p2$, as marked in Fig. 2). The behavior of r_0 is best understood by first making a comparison with an unperturbed case; for example, consider the evolution according to the *unperturbed* flow of a material curve $\tilde{r}_0(t)$ which at $t = 1$ lies on the segment of the *perturbed* unstable manifold that is essentially $P_\epsilon[r_0(t=0)]$, except that its end points $\tilde{p}1$ and $\tilde{p}2$ are the points which asymptote to the unperturbed fixed point rather than the perturbed fixed point (see Fig. 3). This material curve winds around the core *ad infinitum*, never intersecting the unperturbed core boundary, except at $\tilde{p}1$ and $\tilde{p}2$. Since the end points asymptote to a hyperbolic fixed point, the infinitesimal line elements of $\tilde{r}_0$ associated with the end points stretch asymptotically exponentially in time; however, overall $\tilde{r}_0$ stretches only asymptotically linearly in time. Hence as time progresses, the stretch profiles in Fig. 3 will be peaked more and more toward the end points, with an overall dip in the middle corresponding to poor stretching. In the perturbed case, r_0 is entrained into the core at $t = 1$, and then for near-integrable flows starts to wind around the core not much unlike $\tilde{r}_0$ in the unperturbed case, except that an arbitrarily small time-periodic modulation of the velocity field suffices to cause r_0 to intersect the stable manifold segment of the perturbed core boundary (these intersection points are SIP's), as shown heuristically in Fig. 2 for the open flow (numerical simulations are shown presently). The SIP's have a twofold significance: first, like the PIP's they asymptote to a hyperbolic fixed point and thus the infinitesimal line elements of r_0 associated with the SIP's stretch asymptotically exponentially in time; second, they are the dividing points between segments of r_0 that behave in a qualitatively different fashion, as shown in Fig. 2.

(i) A "returned" part $r_0 r_3$, at $t = 3$ between SIP's a,b and e,f and which at $t = 3$ is close to $r_0(t = 0)$, and hence *will evolve from time $t = 3$ approximately the way r_0 does from time $t = 0$ (the SIP's playing the role that the PIP's did originally for r_0).* The subscript on the second r specifies the number of time samples it takes for $r_0 r_3(t = 0)$ to "return" to $r_0(t = 0)$ (the subscript on future symbols should be clear from this comment).

(ii) A detrained part $r_0 d_3$, at $t = 3$ between SIP's b,c and d,e, which will remain outside the core and stretch asymptotically linearly in time.

(iii) A tail, $r_0 t_3$, at $t = 3$ between SIP's c,d, which will wind around the core as its SIP's asymptote to the left fixed point, but in a way that is qualitatively different from r_0. [Though a general scheme needs to include a tail, for some system parameters there will be no tail, as can occur, for example, if in Fig. 2 the tip of $E(-2)$ lies in $D(0)$.]

(iv) The remainder of r_0, to which we add no further labels for the moment.

In this manner we build up a symbolic dynamics, so to speak, that applies to the *entire material curve*, as shown heuristically in Fig. 4. To understand how this is built up, one needs to recognize the following.

(i) With each additional application of P_ϵ, a portion of the remainder of r_0 will be "returned" and another portion detrained. Hence, there will be an $r_0 r_4$ and an $r_0 d_4$ after the fourth iterate of P_ϵ, then an $r_0 r_5$ and an $r_0 d_5$ after the fifth

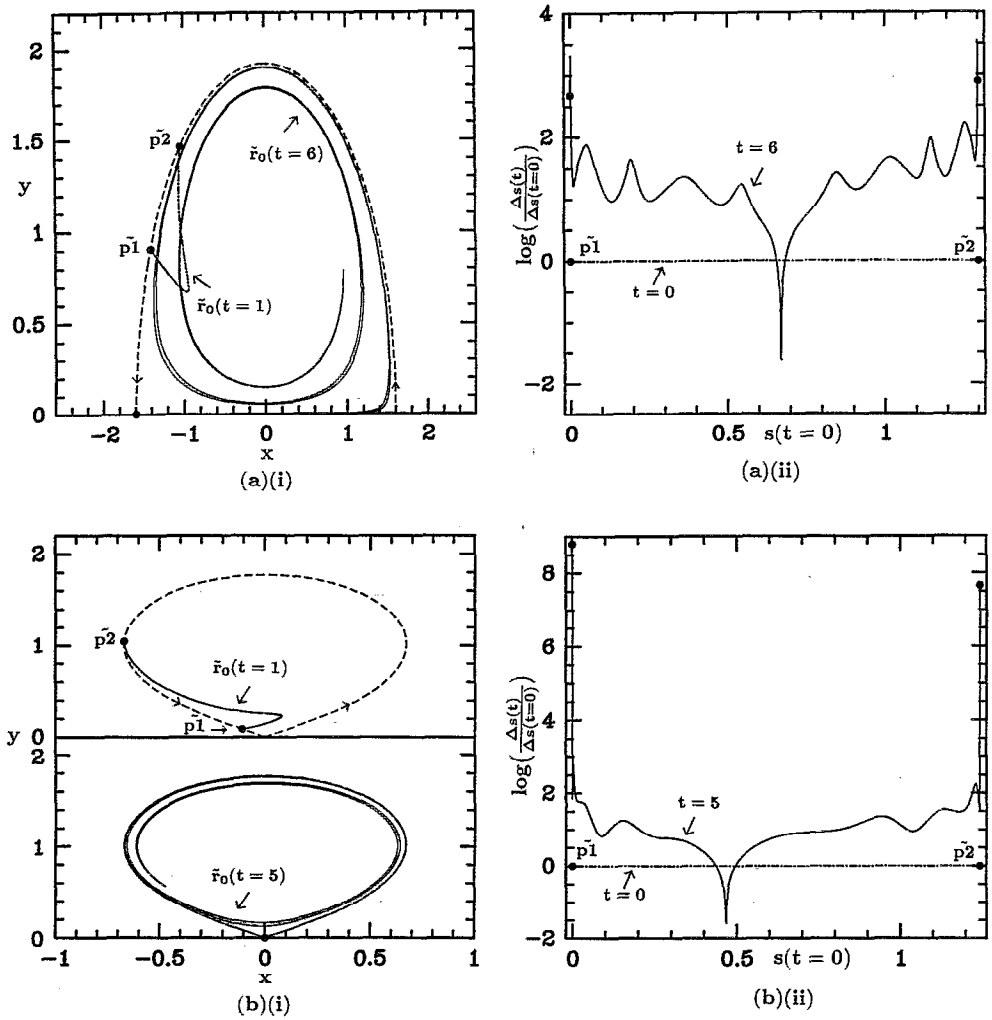

FIG. 3. Evolution, according to the unperturbed (a) open and (b) closed flows, of the material curve $\tilde{r}_0(t)$. The material curves are shown in (i) (the dashed line represents the unperturbed separatrix), and the associated stretch profiles $\log[\Delta s(t)/\Delta s(t=0)]$ are shown in (ii) (s denotes arclength along the material curve). The magnitude of the circulation associated with each vortex is $\Gamma = 0.4(2\pi)^2$. For the stretch profiles, note that $\tilde{r}_0(t=0) = P_\epsilon^{-1}[\tilde{r}_0(t=1)]$, i.e., we evolve $\tilde{r}_0(t)$ from $t=0$ to $t=1$ according to the *perturbed* flow (see Fig. 5 for the parameters) and from then on according to the *unperturbed* flow. For the perturbed flows, the vortices are at $(0, \pm 1)$ at $t=0$ (following the convention of Rom-Kedar *et al.*[6]); for the unperturbed flows the vortices are set at the time average of the perturbed vortex coordinates, i.e., at $(0, \pm 0.921\,83)$ for the open flow and $(0, \pm 0.980\,39)$ for the closed flow.

iterate of P_ϵ, and so on. Referring to Fig. 2, it is clear from the lobe dynamics that at $t = 3$ $r_0 r_n$ ($n \geqslant 3$) is a portion of the boundary of $E(-2)$ between $D(n-3)$ and $D(n-2)$, and $r_0 d_n$ is a portion of the boundary of $E(-2)$ inside $D(n-3)$. Hence the $r_0 r_n$'s and $r_0 d_n$'s join contiguously to cover r_0, as shown by the upper line in Fig. 4. These pairs of symbols on the upper line pertain to the first revolution around the core of successive pieces of r_0.

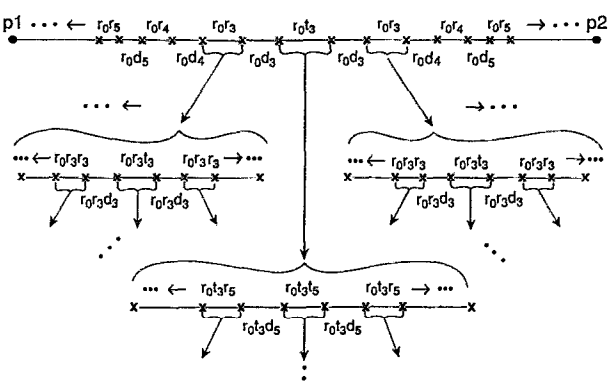

FIG. 4. The symbolic dynamics that applies to r_0 in the open flow. The upper line corresponds to r_0, and the lower lines to enlargements of pieces of r_0.

(ii) The tail and returned parts will further subdivide qualitatively in the same way as r_0 starting at $t=0$; the detrained parts do not keep subdividing. When we say that further subdivisions are qualitatively similar, we should note that successive revolutions around the core may take a different number of iterates of P_ϵ for the first return to occur, and that these numbers may in fact differ (by at most one) on either side of the tail (when this occurs we split the tail in two and have the t subscript on each side agree with that of the adjacent detrained part). Also, for large enough perturbations, though the tail subdivisions will for large enough n settle down to the alternating sequence of $r_0 \cdots t_m r_n$, $r_0 \cdots t_m d_n$, $r_0 \cdots t_m r_{n+1}$, $r_0 \cdots t_m d_{n+1}$, etc., for small n the ordering may be a bit erratic, due to the possibly complicated behavior of the tail. These minor complications do not detract from the overall usefulness of the symbolic dynamics in understanding the topology of enhanced stretching.

The result is a dense labeling of the interface, consisting of a string of symbols that terminate if a "d" is reached. For

example, the string $r_0 r_7 r_{20} d_5$ denotes the segment of r_0 which takes seven time samples to wind around the core and "return" to r_0, then another twenty time samples wind around the core once more and "return," and then five time samples to wind around the core one last time to be detrained. The dynamics is thus described by a string of symbols, with three types of entries: r, d, and t. The differences from conventional symbolic dynamics are that the strings may be finite (if they contain a d), and that there is a subscript on each entry that specifies the time interval between successive entries. These differences in time scales, as we shall later discuss, are highly relevant to understanding the dynamics within the tangle regions. One can construct a similar symbolic dynamics for the closed flow, with two differences: (i) detrained segments can be reentrained, so detrained segments will further subdivide, and the string of symbols will not terminate at a d; (ii) the segment r_0 can pass through both the upper and lower core, which can be kept track of by using, say, small letters for the upper core and capital letters for the lower core.

The symbolic dynamics, coupled with a knowledge of the unperturbed stretch profile, provides a framework for studying the mechanism for, and topology of, enhanced stretching in chaotic tangles. As described earlier, the "returned" segments evolve approximately like $r_0(t=0)$, so one understands the enhancement of the stretch profile in terms of the unperturbed profile approximately repeating itself on smaller and smaller scales, the topology of this repetition understood via the symbolic dynamics (refer back to Fig. 4). For the open flow, the detrained segments stretch asymptotically linearly in time, so the stretch profile will contain gaps of poor stretching, the ordering of which is again understood by the symbolic dynamics. We present in Fig. 5 the stretch profiles of a short-time simulation of the open and closed flow, and in Fig. 6 we compare the total length of the interfaces in the perturbed and corresponding unperturbed case (as described earlier). Rather than give a detailed discussion of the topology of the profiles in Fig. 5 and the relation to symbolic dynamics, let us instead make several observations about the stretching.

(i) Notice the relation between the unperturbed stretch profiles and the perturbed profiles. For example, the unperturbed closed flow profile, for the parameters chosen, has good stretching highly localized around the end points, with a large dip of poor stretching between the end points; hence the perturbed profile has good stretching highly localized around the SIP's, with significant dips of poor stretching between SIP's. In contrast, the unperturbed open flow profile is more even (due in part to the presence of the other hyperbolic fixed point); hence the perturbed profile has a more even distribution of good stretching. We see then that *the unperturbed stretch profile provides qualitative and yet basic information about the stretch distribution in the perturbed profile*, which is highly relevant to efficiency of mixing in the context of diffusion across interfaces, as we will soon describe.

(ii) The mechanism for enhanced stretching can be understood partially in a variety of ways, and the different flows highlight these different ways. For example, consider *residence time near the hyperbolic fixed points.* The creation of SIP's by external forcing entails a creation of a countable infinity of points along the interface that eventually reside near the fixed point for all time, and hence entails an overall increase in the interface's residence time near the hyperbolic fixed points. This mechanism is dominant in the closed flow, where the stretching is highly localized around SIP's. Consider also *the repeated stretch, fold, and return mechanism*, which need not be tied to residence near the hyperbolic fixed point. For example, referring back to Fig. 2, after three perturbation periods r_0 has been stretched and folded. The portion $r_0 r_3$ will repeat the stretch and fold process, as will then $r_0 r_3 r_3$, and so on. As long as the concerned segment stretches by a factor greater than 1 with each fold and return, it will be undergoing exponential stretching. Here, then, exponential stretching is not necessarily due to good local stretching near a hyperbolic fixed point, but to the repeated stretch and fold mechanism. This mechanism is highlighted by the open flow, where stretching is more evenly distributed between the SIP's. Note that the stretch profile of $r_0 r_3$ $(t=3)$ need not be uniformly greater than 1, so that one will have to compute the profile (and then further profiles) to determine which portions of r_0 are candidates for exponential stretching. In such a manner one determines "where the horseshoe maps are," so to speak. A third explanation is that turnstile lobes act in a certain sense as *reorientation* lobes. As segments of r_0 wind around the core, their stretch efficiency decays not much unlike the unperturbed case; however, when they reach the turnstile lobes they are then mapped by P_ϵ in such a way as to reorient many of the line elements to increase stretch efficiency. All of the above-mentioned mechanisms for stretch enhancement are really part of the same overall mechanism we previously mentioned, that is the *approximate self-similar behavior of r_0 as it evolves.*

(iii) Notice the great variation of stretch histories (the vertical scale of the profiles in Fig. 5 is logarithmic) and the spatial complexity of the interface produced in a very short time by a very simple velocity field. As we shall discuss, mixing in the context of diffusion of passive scalars across an interface depends on the stretch *histories* of the interface, and, as the simulations in Fig. 5 suggest, there may be little obvious connection between a velocity field's properties and the corresponding advection properties. In particular, simple velocity fields can of course produce complicated advection, and more generally the scaling properties of turbulent velocity fields may have little to say about the mixing properties. In addition, the presence of hyperbolicity in the chaotic tangles implies substantial temporal variation of the strain experienced by an infinitesimal line element, as well as great spatial variation of strain history along a material curve. In studies of turbulent mixing, sometimes an *assumption* of constant strain is made at some point (see, for example, Batchelor[8] for one of the early well-known works using this assumption); such an assumption certainly would not apply in chaotic tangles, and we see no particular reason for it to be valid for turbulent flows.

(iv) Though previous investigators have commonly referred to chaotic zones as regions of "enhanced" stretching, notice how in both the open and *closed* flows there are gaps of

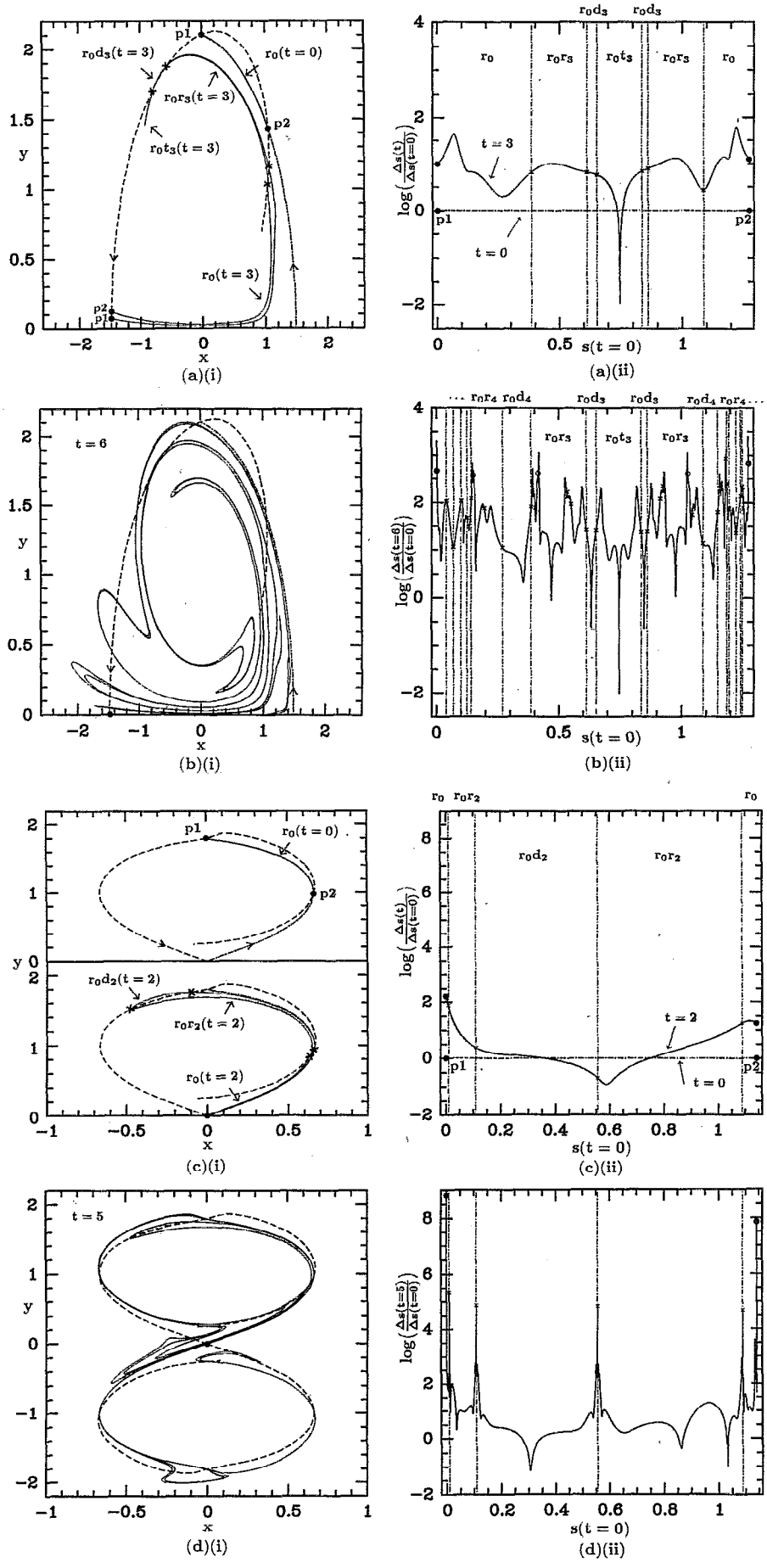

FIG. 5. Plots of (i) r_0 and (ii) associated stretch profiles for (a) the open flow at $t = 0,3$, (b) the open flow at $t = 6$, (c) the closed flow at $t = 0,2$, and (d) the closed flow at $t = 5$. The stable manifolds are shown by dashed lines, and the unstable manifolds by dash-dotted lines. In plots (b) and (d) we do not explicitly label all the subdivisions of r_0, since the labeling would become much too crowded on the scale shown [plot (b) does indicate additional SIP's by crosses, and points which are "almost SIP's" (i.e., almost intersect the stable manifold segment of the core) by open dots]. The flows are specified by the parameters $\Gamma = 0.4(2\pi)^2$ and (a) $\epsilon = 0.085 \times 2\pi$, (c) $\epsilon = 0.02 \times 2\pi$. For these and all other simulations, interfaces are represented by a weighted distribution of passive marker particles. The weighting is determined by applying dynamic point insertion to an initially uniform distribution (starting the simulation over with the weighted distribution then eliminates any point insertion error). Particle trajectories are found by integrating the governing ODE system with a fourth-order Runge–Kutta scheme.

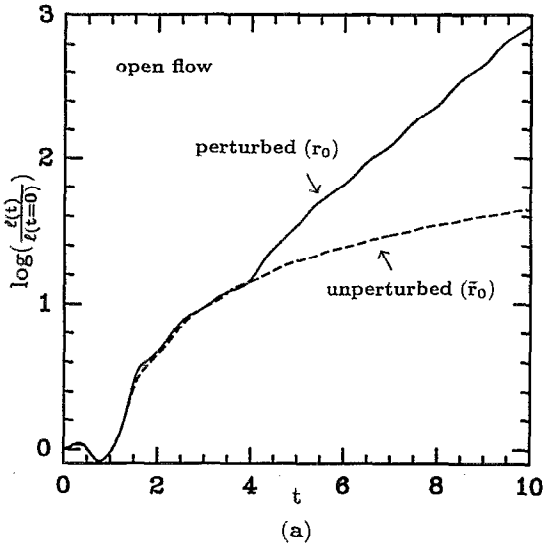

FIG. 6. Total length $l(t)$ of r_0 in the perturbed case and of $\bar{r}_0$ in the unperturbed case for (a) the open flow and (b) the closed flow.

poor stretching on finite time scales. As previously mentioned, these gaps in the open flow are partially explained by detrainment. For both flows, however, the additional explanation is offered by the previously described dip in the unperturbed stretch profile, which carries through to the perturbed profile by the previously described approximate self-similar enhancement of the stretch profile. Note how the regions of poor stretching typically correspond to "bulbous" parts of the lobe, and turning points such as the "tip" of the interface. These bulbous parts can rotate throughout the chaotic zones with ineffectual stretching on finite time scales, not coming close to the fixed point, and not undergoing the repeated stretch and fold mechanism. We should stress here a basic difference between the open and closed flow: in the closed flow the chaotic zone associated with the homoclinic tangle is finite, reentrainment occurs, and poor stretching eventually gives way to enhanced stretching; in the open flow the chaotic zone associated with the heteroclinic tangle is infinite, and detrainment leads to advection out to $x = -\infty$ in an asymptotically linear manner, so that poor stretching can remain so for all time.

(v) Consider the countable infinity of time scales involved in the interface stretching. Basically r_0 is undergoing a repeated stretch and fold process, with the magnitude of stretching depending on the location relative to the fixed point, *and this stretch, fold, and return process takes place on all time scales (above some minimum time), as monitored by the symbolic dynamics*: for example, $r_0 r_n(t=0)$, $n > n_{\min}$, takes n time samples for the first stretch, fold, and return. Hence horseshoe maps occur on all time scales above some minimum time. We feel these different time scales are relevant to the study of many basic elements of the dynamics in the tangle regions, such as statistical relaxation under chaotic advection, but we will not elaborate upon this here.

(vi) Consider the total length of the interfaces, as shown in Fig. 6. Up until the first "returned" portion of r_0 wraps around $E(0)$ (i.e., a portion of r_0 begins its second wind around the core), the total length of r_0 does not differ significantly from that of $\bar{r}_0$ in the unperturbed problem; however, thereafter (after $t \approx 4$ for the open flow and $t \approx 3$ for the closed flow) the length of r_0 quickly converges to mild oscillation about exponential stretching. In this context, great regularity underlies the approximate self-similar behavior of r_0.

The above observations highlight some of the relevant features of stretching as understood in the geometrical setting of invariant manifolds. With these observations as a backdrop, we wish to convey the utility of the geometrical framework offered by the invariant manifolds. Heretofore, the horseshoe map construction and its associated symbolic dynamics have been the central paradigm for enhanced stretching in the chaotic tangles of two-dimensional maps. Though establishing exponential stretching on a nontrivial subset of phase space, this construction offers little global understanding of stretching, and the study of global stable and unstable manifolds provides an extension of this analysis. The conventional horseshoe map analysis applied to chaotic tangles considers the repetition of an essentially uniform stretch and fold process in a domain that is a small subset of phase space near a hyperbolic fixed point, and ignores all pieces of the domain that map outside the domain; the analysis based on invariant manifolds considers the *approximate* repetition of highly *nonuniform* stretching and folding of turnstile lobe boundaries *throughout* the chaotic zone, and monitors the *entire* material curve. For near-integrable systems, it is the unperturbed stretch profile that is approximately repeated on smaller and smaller scales, and the topology of this repetition is understood by the symbolic dynamics. The utility of both the conventional horseshoe map analysis and the invariant manifold analysis is primarily conceptual, but the latter offers a framework for more detailed quantitative analysis. Applications to realistic flows can in practice be difficult, which motivates first a study of model systems. Past models for the stretch processes under chaotic flows have typically been fairly elementary. For example, in their study of the kinematic dynamo problem un-

der chaotic advection, Ott and Antonsen[9] assume simple enough models for the stretch processes, such as a baker's map, that stretch distributions are described by standard multifractals from a binomial multiplicative process. The invariant manifold analysis of this paper provides a framework for much more realistic models, in which the topology of, and multiple time scales involved in, fluid flows are respected. We are presently studying some models, and from the previous discussion it should not be too difficult to anticipate the nature of these models: one chooses a reasonable analytical model for the unperturbed stretch profile, one whose nonuniformity can be adjusted by a parameter, and have the profile repeat on smaller and smaller scales according to the symbolic dynamics (with the SIP locations chosen by some reasonable model). The result is a multifractal with a much richer construction, one that is truer to what one would expect in realistic flows. To study realistic flows rather than models can involve intensive computation; however, the invariant manifolds provide a framework for such computation. The goal of course, as done in Fig. 5, is to identify the hyperbolic fixed points and local stable and unstable manifolds (for near-integrable systems there is a standard and easily implemented procedure for doing this); one then computes the global stable and unstable manifolds by evolving the local stable and unstable manifolds backward and forward in time, respectively, from which one can determine the turnstile lobes and hence $r_0\,(t=0)$ and $\bar{r}_0\,(t=0)$. Simulations of $\bar{r}_0$ illustrate the stretch profile that is approximately repeated on smaller and smaller scales in the perturbed case, and one can study and alter this profile (by changing system parameters) before going on to the perturbed simulations. Of course, to determine the topology for any given perturbed flow, such as finding the SIP's and calculating stretch profiles, one must resort to explicit numerical simulation. Such computations can become arduous for long time scales, but *it is nevertheless crucial to appreciate that an underlying geometrical framework does exist in the chaotic tangles.* At present, the studies of chaotic advection rarely, if ever, acknowledge the presence of an underlying topology in chaotic zones: typically a blob or set of points are initialized in a fairly random fashion, and the fact that finite-time stretching can vary greatly within the chaotic zones is not addressed. *A main goal of this paper then is to encourage future studies to address the underlying geometries within chaotic tangles and the great variation of the dynamics on finite time scales.* There are a number of examples where such a consideration would be helpful. For example, in their study of separation under flow reversal, Aref and Jones[10] present a computational example illustrating that separation of a material blob is better in chaotic zones than in regular zones, and argue that this enhancement is due to the enhanced stretching in chaotic zones. A consideration of the topology of stretching on finite time scales would be useful in maximizing this effect.

We should point out that, though long-time computations can be arduous, they are not so difficult as one might first think. For example, suppose one wished to determine $\mu(r_0 r_i \cdots r_j d_k\,(t=n))$ for all possible combinations of $i+\cdots+j+k=n$, where $\mu(\,)$ denotes the total stretch from $t=0$. As $n\to\infty$, both the value each indice can take and the number of possible r entries in the string go to infinity. However, it is reasonable to expect that $\mu(r_0 r_{i+1}\,(t=i+1))/\mu(r_0 r_i\,(t=i))$ asymptotes to a simple relation as $i\to\infty$ and that for sufficiently small perturbations one can find upper and lower bounds on how the first revolution around the core [i.e., $\mu(r_0 r_i\,(t=i))$] describes later revolutions (due to the approximate self-similar behavior of r_0). With these two properties, a finite set of quantities [$\mu(r_0 r_i\,(t=i))$ for a finite number of i's] provides a reasonable characterization of the stretching. We are presently studying quantitative issues such as these; however, as with the conventional horseshoe map construction and other paradigms of dynamical systems theory, the goal of our stretch analysis is not to provide a large amount of data, but rather to provide a conceptual framework with which to understand the order that underlies the chaotic dynamics.

We end this section by stressing that for classes of three-dimensional flows and flows with more general velocity field time dependences one can define from invariant manifolds sets of lobes which map from one to another and obtain from this a global picture of enhanced stretching. To do so requires recent and fairly technical developments in dynamical systems theory, involves much more complicated geometries, and is beyond the scope of this paper. We will briefly discuss, however, the two simple cases mentioned in the Introduction. For a two-frequency velocity field,[3] one needs to generalize from *maps* to *sequences of maps.* Transport in phase space is described in terms of lobes mapping within a *sequence* of lobe structures that are derived from invariant manifolds embedded in a higher-dimensional Poincaré section. For a three-dimensional system with normally hyperbolic invariant 1-tori[4] (i.e., the hyperbolicity is in a plane "normal" to the 1-torus, and there is no exponential stretching "along" the torus), the extension from two dimensions is fairly robust, with three-dimensional lobes stretching and folding in two dimensions, and negligible stretching in the third dimension.

III. DIFFUSION ACROSS INTERFACES

Let us now consider mixing in the context of diffusion of passive scalars across an interface (the analysis extends to passive vectors, which will be the subject of a future paper). For example, a natural problem starts off with fluid A outside the perturbed core and fluid B inside the perturbed core; we consider a simpler scenario that starts off with fluid A in the entraining turnstile lobe and fluid B everywhere else (this allows us to concentrate on a single lobe, and from the lobe dynamics it should be clear how the latter problem relates to the former). As these fluids advect according to the open or closed flow, let them diffuse, according to the standard diffusion equation, across the interface defined by the material curve that initially lies on the entraining turnstile lobe boundary. This curve thus consists of r_0 and a segment of the stable manifold; the stable manifold part of the curve, as seen in Fig. 5, quickly shrinks to negligible length relative to r_0, and so with negligible error we ignore the diffusion across the stable manifold segment and focus only on r_0. For

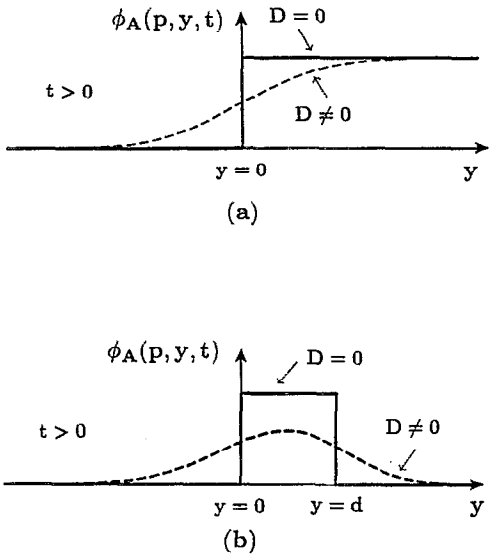

FIG. 7. The concentration profile of fluid A near the interface at $\mathbf{p}$ for (a) the nonoverlapping thin diffusion zone approximation and (b) the overlapping thin diffusion zone approximation.

thin enough diffusion zones and sufficiently steep concentration gradients across the interface, which is valid for a given diffusion coefficient for small enough times, the diffusion process is essentially normal to the interface, and can be treated as one dimensional. The amount diffused across the interface can then be related to the stretch history of the interface.[1,2,11-13] The concentration of fluid A (and similarly fluid B) spreads with increasing time like an error function [see Fig. 7(a)]:

$$\Phi_A(\mathbf{p},y,t) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{S(\mathbf{p},t)y}{[4D\tau(\mathbf{p},t)]^{1/2}}\right)\right], \tag{2}$$

where $\mathbf{p}$ specifies a point on the interface, y is a relative coordinate normal to the interface at $\mathbf{p}$ and pointing into the lobe, D is the diffusion coefficient, $S(\mathbf{p},t) = \delta s(\mathbf{p},t)/\delta s(\mathbf{p},t=0)$ is the interface stretch at $\mathbf{p}$, and

$$\tau(\mathbf{p},t) = \int_0^t S^2(\mathbf{p},\tilde{t})\,d\tilde{t}. \tag{3}$$

The diffusion rate per unit initial arclength across the interface at $\mathbf{p}$, $C(\mathbf{p},t)$, is given by

$$C(\mathbf{p},t) = SD\left.\frac{\partial \Phi_A}{\partial y}\right|_{y=0} = \frac{1}{2}\left(\frac{D}{\pi\tau}\right)^{1/2} S^2. \tag{4}$$

The amount of fluid A diffused across the interface at $\mathbf{p}$ per unit initial arclength is then

$$\mathcal{A}(\mathbf{p},t) = \int_0^t C(\mathbf{p},\tilde{t})\,d\tilde{t} = \left(\frac{D\tau}{\pi}\right)^{1/2}. \tag{5}$$

Hence, the local time-dependent diffusion rate and the amount diffused across the interface is completely determined by the local stretch history of the interface. The total amount of fluid diffused out of the lobe is found by integrating Eq. (5) over all $\mathbf{p}$. Enhanced interface stretching corresponds to enhanced diffusion across the interface, and a global understanding of stretching directly translates into an understanding of diffusion. However, this analysis is valid only for sufficiently thin diffusion zones, and a notable exception to this case occurs when the interface folds back onto itself so that the diffusion zones overlap.[13] As should be clear from the plots in Fig. 5, in chaotic zones interfaces quickly fold and wrap around themselves in a violent manner, and one would thus like to employ an analysis that addresses overlap of neighboring diffusion zones. A full analysis that takes into account the overlap of all neighboring diffusion zones would be quite laborious. Instead we employ an analysis that takes into account the overlap of *nearest neighbor* diffusion zones, which is sufficient to capture additional phenomena relevant to the notion of *efficiency of mixing*. The key to such an analysis is to notice, as seen in Fig. 5, that much of the neighboring sections of the interface are essentially parallel to one another and have small curvature (typically these portions of the interface have good stretching); in addition, the parts of high curvature tend to be bulbous and hence have less of a tendency for overlap of diffusion zones (typically these portions have poor stretching). Under the assumption that overlap of neighboring diffusion zones is not significant until the neighboring segments of the interface are essentially parallel, which is valid for a small enough diffusion coefficient, the diffusion process can still be treated as one dimensional. We address overlap of diffusion zones when it is due to the lobe being sufficiently thin; we do not worry about overlap due to the lobe folding on top of itself. When there is overlap at $\mathbf{p}$, Eq. (2) is replaced by a superposition of two error functions

$$\Phi_A(\mathbf{p},y,t) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{Sy}{(4D\tau)^{1/2}}\right) - \operatorname{erf}\left(\frac{S_d(y-d)}{(4D\tau_d)^{1/2}}\right)\right], \tag{6}$$

where $d = d(\mathbf{p},t)$ is the normal distance at $\mathbf{p}$ between the two neighboring segments of the interface whose diffusion zones overlap, and S is understood to represent $S(\mathbf{p},t)$, while $S_d \equiv S(\mathbf{p} + d\hat{y},t)$ (and similarly for τ and τ_d). Note that in the limit $d \to \infty$ (with y fixed) the second error function goes to -1, and hence Eq. (5) recovers Eq. (2). The concentration spreads with increasing time like a sum of two error functions, one associated with each interface [see Fig. 7(b)]. By a calculation similar to that in Eq. (4), except that one has to now take into account the presence of the other interface, the diffusion rate out of the lobe per unit initial length of neighboring segments of the lobe boundary is

$$C(\mathbf{p},t) = \frac{1}{2}\left(\frac{D}{\pi}\right)^{1/2}\left\{\frac{S^2}{\tau^{1/2}}\left[1 - \exp\left(\frac{-(Sd)^2}{4D\tau}\right)\right] + \frac{S_d^2}{\tau_d^{1/2}}\left[1 - \exp\left(\frac{-(S_d d)^2}{4D\tau_d}\right)\right]\right\}. \tag{7}$$

The presence of a neighboring interface thus adds to the diffusion process a saturation term of the form $\{1 - \exp[-(Sd)^2/4D\tau]\}$, so that the separation distance d is a factor in the diffusion process. To find the total amount of fluid diffused out of the lobe, a convenient trick, with physical meaning, is to separate the two terms in Eq. (7) and

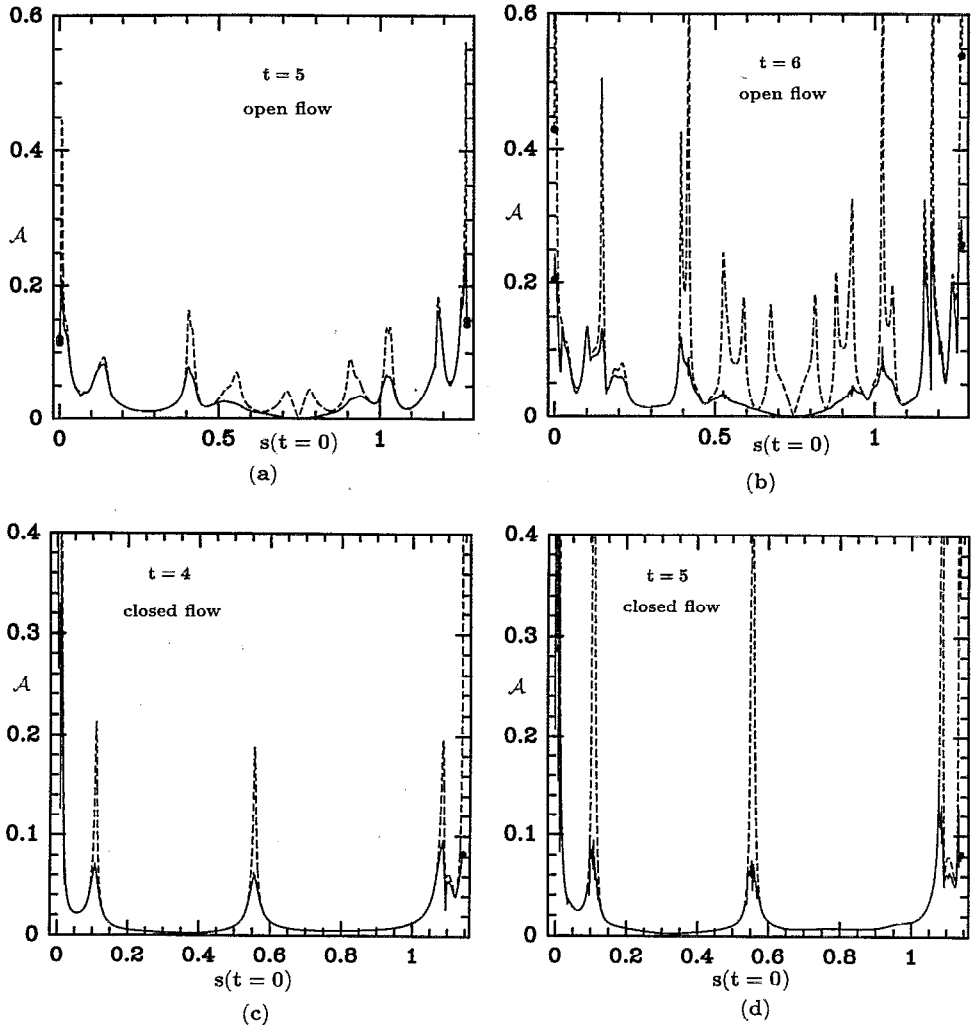

FIG. 8. Comparisons of $\mathscr{A}(\mathbf{p},t)$ when r_0 advects under the same velocity fields as considered in the examples of Fig. 5 ($D = 2\pi \times 10^{-6}$ for both flows). The dashed line corresponds to the nonoverlapping thin diffusion zone theory and the solid line to the overlapping thin diffusion zone theory.

then integrate over *all* $\mathbf{p}$ in the interface, rather than integrate (7) over half the $\mathbf{p}$ of neighboring segments of the interface whose diffusion zones overlap. Hence for all $\mathbf{p}$ we write

$$C(\mathbf{p},t) = \frac{1}{2}\left(\frac{D}{\pi\tau}\right)^{1/2} S^2\left[1 - \exp\left(\frac{-(Sd)^2}{4D\tau}\right)\right], \qquad (8)$$

where it is understood that when there is negligible overlap at $\mathbf{p}$ we send $d \to \infty$. The amount of fluid A diffused across the interface at $\mathbf{p}$ per unit initial arclength, $\mathscr{A}(\mathbf{p},t)$, is then given by integrating $C(\mathbf{p},t)$ in time as done in Eq. (5), and then the total amount of fluid A that leaks out of the lobe is given by integrating $\mathscr{A}(\mathbf{p},t)$ over *all* $\mathbf{p}$. Note that, to determine $\mathscr{A}(\mathbf{p},t)$ at a particular time, one can of course also deal with the concentration $\Phi_A(\mathbf{p},y,t)$ rather than a time integral of $C(\mathbf{p},t)$; the advantage of the former approach is that one need calculate $d(\mathbf{p},t)$ only at that given time.

We present in Fig. 8 some plots of $\mathscr{A}(\mathbf{p},t)$ when r_0 advects under the same velocity fields as considered in the examples of Fig. 5. Figure 9 shows the total amounts of fluid A that diffuse out of the lobe. The saturation term is clearly relevant to the efficiency of the diffusion process. For example, in the closed flow simulation, though the interface stretches more than in the open flow, this stretching is highly localized in a few regions where the lobe is extremely thin, and the diffusion process quickly saturates since there is little fluid available to diffuse out of these parts of the lobe.

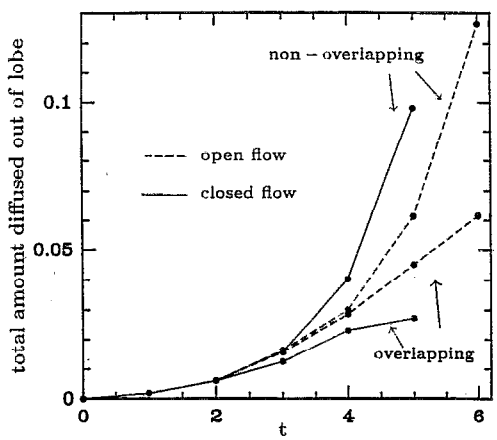

FIG. 9. The total amount of fluid A that diffuses out of the lobe in the nonoverlapping and overlapping theory. The entraining turnstile lobe area is 0.1388 and 0.0931 for the open and closed flow, respectively.

Hence, we see from $t = 4$ to $t = 5$ there is little enhancement of diffusion since, though there is a great amount of stretching going on, the stretching is localized in regions where the diffusion process has saturated (note that as more SIP's asymptote to the hyperbolic fixed point, the diffusion process will again be enhanced, so that the overall enhancement is staggered in time). This is in contrast to the open flow example, where the stretching is more evenly distributed, and saturation is less of a factor. *One sees then that, in the context of diffusion, the efficiency of mixing depends not only on the efficiency of stretching along the interface, but on how this stretching is distributed relative to the separation of neighboring segments of the interface.* This is highlighted in Fig. 9, where we see that, even though r_0 stretches more in the closed flow, the open flow has a greater amount of diffusion on the time scale shown (the greater saturation in the closed flow is due in part to the smaller turnstile lobe area, but more importantly to the greater nonuniformity of the stretch distribution). On longer time scales the differences between a finite and infinite mixing region (the tangle) of the closed and open flow, respectively, will become more apparent. As mentioned earlier, regions of poor stretching in the closed flow, due to reentrainment, eventually give way to enhanced stretching, which is not necessarily the case in the open flow. However, the finiteness of the closed flow mixing region implies that diffusion zone overlap, and hence saturation, will become more of a factor than in the open flow (especially when one considers more than just nearest neighbor overlap). Also, in the original problem of fluid B in the core and fluid A outside the core, with successive time samples the open flow entraining turnstile lobe will always contain only fluid A, except for the amount diffused out, while the closed flow entraining turnstile lobe will contain less and less of fluid A due to more and more reentrainment of fluid B via this lobe. In this context, open flow mixing involves more fluid and lasts longer.

With Figs. 8 and 9 in hand, we emphasize the spirit of our investigation. Certainly the approach to studying diffusion is an approximate one, with several sources of error: we consider only nearest neighbor overlap, we ignore diffusion across the stable manifold boundary, and the assumption of parallel neighboring segments of the interface and a one-dimensional diffusion process break down at sharp turning points of the interface, such as the tip of the lobe. However, the errors regarding the stable manifold and sharp turning points involve a truly negligible amount of fluid, and nearest neighbor overlap is sufficient to approximately capture the effect of saturation. Our main goal is not to focus on a computational prescription for exact numerical results, but rather to take a first step past the nonoverlapping thin diffusion zone theory, and to show how the efficiency of diffusion is affected by this extension. Until now our discussion of "efficiency" has been informal; though there is more than one candidate for a formal definition of efficiency of diffusion in this context, none seem adequate for a *global* definition. For example, it seems natural to address the nonuniformity of stretching and interface separation, and hence to compare the amount of fluid diffused out of the lobe with an idealized case that has the same lobe area and lobe boundary length, but spatially uniform stretching and manifold separation. Dividing the nonuniform result by the uniform result could define an "efficiency" of diffusion, except that the result is not guaranteed to be less than one: though $\Delta A_o(t)/\Delta A_{no}(t)$ can be *decreased* by nonuniformity, where $\Delta A_o, \Delta A_{no}$ denote the total amount of A diffused out of the lobe in the overlapping and nonoverlapping theory, respectively, $\Delta A_{no}(t)$ can be *increased* by nonuniformity, so that the above "efficiency" can be greater than 1 or less than 1 (indeed for Fig. 9 the above "efficiency" is found to be greater than 1 for the open flow and less than 1 for the closed flow for $t \geqslant 3$). For a global consideration of diffusion, it is not clear that there is a "best case" on which to base a definition of efficiency. Rather than focusing on a formal definition, and exact quantification, of efficiency, however, we would rather highlight the qualitative, practical aspects of the problem. For example, it is clear that the efficiency of diffusion across an interface is affected by the nonuniformity of stretching in the perturbed case, which, as discussed before, is related to the unperturbed stretch profile, which in turn is affected by the geometry of the unperturbed phase portrait (for example, the presence of an additional hyperbolic fixed point in the open flow tended to make the stretch distribution more even), and by the unperturbed system parameters (such as the strength of the vorticity). Thus one can alter the efficiency of the diffusion process by changing the geometry of the unperturbed phase portrait or the system parameters. In addition, we see from Figs. 5 and 8 that the local stretch and diffusion rates vary greatly with initial conditions, and by straightforward calculations one can identify the regions of good stretching and mixing; for example, in the closed flow simulation good stretching and mixing is isolated near the SIP's, so identifying these points locates in phase space the "seeds" of good stretching and mixing. Though exact results can only come from explicit numerical simulation, we feel that a knowledge of the unperturbed stretch profile, coupled with an appreciation of the symbolic dynamics that applies to material interfaces, provides the framework for studying many aspects of stretching and mixing in chaotic tangles.

ACKNOWLEDGMENTS

This material is based upon work supported by a National Science Foundation Graduate Fellowship, the National Science Foundation Presidential Young Investigator Program, the Office of Naval Research Young Investigator Program, and Caltech's Program in Advanced Technologies, sponsored by Aerojet General, General Motors, and TRW.

[1] J. M. Ottino, J. Fluid Mech. **114**, 83 (1982).
[2] J. M. Ottino, *The Kinematics of Mixing: Stretching, Chaos, and Transport* (Cambridge U.P., Cambridge, 1989).
[3] D. Beigie, A. Leonard, and S. Wiggins, to appear in Nonlinearity (1991).
[4] S. Wiggins, Physica D **44**, 471 (1990).
[5] V. Rom-Kedar and S. Wiggins, Arch. Rat. Mech. Anal. **109**, 239 (1990).
[6] V. Rom-Kedar, A. Leonard, and S. Wiggins, J. Fluid Mech. **214**, 347 (1990).
[7] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Springer-Verlag, New York, 1990).
[8] G. K. Batchelor, J. Fluid Mech. **5**, 113 (1958).

[9] E. Ott and T. M. Antonsen, Jr., Phys. Rev. A **39**, 3660 (1989).
[10] H. Aref and S. W. Jones, Phys. Fluids A **1**, 470 (1989).
[11] G. F. Carrier, F. E. Fendell, and F. E. Marble, SIAM J. Appl. Math. **28**, 463 (1975).
[12] F. E. Marble, in *Recent Advances in the Aerospace Sciences*, edited by C. Casci (Plenum, New York, 1985), pp. 395–413.
[13] A. Leonard, V. Rom-Kedar, and S. Wiggins, Nucl. Phys. B (Proc. Suppl.) **2**, 179 (1987).