J. Fluid Mech. (2000), *vol.* 403, *pp.* 277–304. Printed in the United Kingdom

Topological fluid mechanics of stirring

By PHILIP L. BOYLAND[1], HASSAN AREF[2] AND MARK A. STREMLER[2]

[1] Department of Mathematics, University of Florida, Gainesville, FL 32611, USA
[2] Department of Theoretical and Applied Mechanics, University of Illinois, Urbana, IL 61801, USA

(Received 8 April 1997 and in revised form 9 March 1999)

A new approach to regular and chaotic fluid advection is presented that utilizes the Thurston–Nielsen classification theorem. The prototypical two-dimensional problem of stirring by a finite number of stirrers confined to a disk of fluid is considered. The theory shows that for particular 'stirring protocols' a significant increase in complexity of the stirred motion – known as topological chaos – occurs when three or more stirrers are present and are moved about in certain ways. In this sense prior studies of chaotic advection with at most two stirrers, that were, furthermore, usually fixed in place and simply rotated about their axes, have been 'too simple'. We set out the basic theory without proofs and demonstrate the applicability of several topological concepts to fluid stirring. A key role is played by the representation of a given stirring protocol as a braid in a (2+1)-dimensional space–time made up of the flow plane and a time axis perpendicular to it. A simple experiment in which a viscous liquid is stirred by three stirrers has been conducted and is used to illustrate the theory.

1. Introduction

It is well known that even for laminar flow at very low Reynolds numbers stirring can lead to extremely complex flow patterns, and that the emergence of such patterns may be understood by appeal to the theory of dynamical systems applied to the equations describing advection in the Lagrangian representation of fluid kinematics. This general phenomenon is known as *chaotic advection* (Aref 1984). Many studies over the past dozen years have shown the relevance of the notion of chaotic advection to flows of considerable importance for a variety of applications. For review and examples see Ottino (1989, 1990) and Aref (1990, 1991, 1994).

In the case of unsteady two-dimensional flow the complexity is 'explained' by considering the advection equations,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = u(x, y, t), \quad \frac{\mathrm{d}y}{\mathrm{d}t} = v(x, y, t), \tag{1.1}$$

where the two flow components u and v satisfy some set of dynamical equations for fluid motion, and noting that (1.1) will, in many cases, be rich enough to have chaotic solutions. Stokes flow has been a favourite example, since u and v are then given in terms of the motion of the boundaries of the flow region. Thus, chaotic advection has been shown to arise in the Stokes flow between eccentric cylinders that are rotated alternately, a flow that is sometimes called the 'journal bearing flow', although the interest here is in cases where the inner and outer circular boundaries are quite different in size (Aref & Balachandar 1986; Chaiken *et al.* 1986; and others). Other studies have used Stokes flow in a rectangular domain driven by moving belts on the

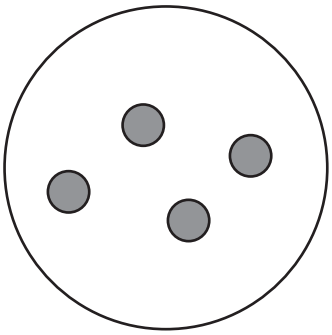

FIGURE 1. Generic 'batch' stirring device considered – a disk of fluid with k embedded stirrers.

top and bottom. In all these applications where certain boundaries move periodically in time the geometrical shape of the fluid region remains fixed. The underlying idea of chaotic particle motion, however, should apply just as well to cases where the geometry is changed, and we shall see in this paper that new insights are obtained when this more general kind of stirring is pursued. Indeed, change in geometry of the fluid region is probably what most people associate with stirring, for example when a spoon moves in a cup of tea.

Figure 1 shows a generic, two-dimensional 'batch' stirring device: an outer boundary specifying the container, shown for convenience as a circle, with a number, k, of internal stirrers, also for convenience shown with equal, circular cross-sections. Thus, the 'journal bearing flow' corresponds to figure 1 with $k = 1$. A device with two fixed internal cylinders (i.e. figure 1 with $k = 2$), rotated alternately, has been explored recently by Jana, Metcalfe & Ottino (1994) as a viscous flow counterpart of the original, inviscid 'blinking vortex flow' of Aref (1984) used to introduce the notion of chaotic advection. In the Stokes flow limit the flow in this two-stirrer device, already considered by Bouasse (1931), has a model representation in terms of 'blinking rotlets' found recently by Meleshko & Aref (1996). It seems clear that pursuing such devices with an ever increasing number of stirrers will not lead to anything qualitatively new. In particular, we would not expect any profound changes if we increased the number of stirrers from two to three. However, if instead of rotating the stirrers in place, one considers 'stirring protocols'† in which the stirrers move about inside the outer boundary, the geometry of the fluid region changes in time and, as we shall see, the topology of how the stirrers are moved relative to one another will determine the efficiency of the stirring process. It then turns out that $k = 3$ is a critical number. For $k \geqslant 3$ it is possible, when the stirrers are moved in a certain way, to produce stirring that is more complex (in a quantifiable sense) than for any motion of $k \leqslant 2$ stirrers. These results hinge only on the continuity of the fluid motion, i.e. not on specific, 'metric' details of the velocity components u and v in (1.1), e.g. whether the flow is a two-dimensional Stokes flow or a flow at higher Reynolds number (although the requirement of a periodically repeatable sequence of flows will arise). The results come almost directly from a theorem in topology that is the cornerstone of so-called Thurston–Nielsen theory (Thurston 1988;‡ Fathi, Laudenbach & Poenaru 1979; Casson & Bleiler 1988; Handel 1985; for a review of

† The term 'stirring protocol' was used in a similar sense by Aref & Balachandar (1986), although in that study the geometry of the fluid region was fixed and the 'stirring protocol' consisted simply in a prescription of how and when to rotate the two eccentric cylinders doing the stirring.

‡ This paper was widely circulated as a preprint in 1975.

dynamical systems applications see Boyland 1994, see also McRobie & Thompson 1993; for a general-audience write-up of the mathematical contributions of Jakob Nielsen see Lundsgaard-Hansen, 1993). Many of the mathematical concepts used in and required by Thurston–Nielsen theory appear, upon proper 'translation', to have both an intuitive immediacy and a potential utility in discussing issues of fluid advection, stirring, quality of mixing, and the like. Since we will not write out formal proofs, much of our paper is devoted to describing precise mathematical ideas and results in qualitative terms that we hope will be found useful by workers in fluid mechanics.

Further motivation comes from the results of a simple laboratory experiment shown in figure 2. Since previous work on chaotic advection has established that low Reynolds number flows provide an ideal match between analytical and experimental results, we set up a system consisting of three stirring rods in a cylindrical container of glycerol. A system of mechanical guides allows us to rotate the two stirring rods on the right, or the two on the left, as we now describe. At the outset the three rods are placed on a diameter. Each step of the stirring protocol consists in transposing either the two rods on the right or the two on the left by moving each through a semi-circle centred halfway between them. At the end of such a step the two rods will have interchanged positions. All motions are performed very slowly so that we may assume the fluid motion to be only due to the motion of the boundaries in the sense of a Stokes flow, i.e. negligible 'secondary fluid motions' are produced by the motion of the stirrers. When two rods have been interchanged and are again stationary, the fluid motion ceases, and the apparatus is ready for the next interchange of stirrers. We start off by transposing the centre and the right rod. Then we transpose the new centre (i.e. original right) rod and the left rod; then the (new) centre and the right rod; and so on. The only variability, then, is whether we make the two rods orbit one another clockwise or counter-clockwise during each transposition. From a practical point of view these two ways of stirring would seem entirely equivalent, and a mechanism that accomplishes one stirring protocol can be adapted to accomplish the other. In terms of energetics or forces on the stirrers the choice of sense of rotation makes no difference whatsoever. However, it is absolutely crucial from a topological point of view and, as we shall see, is a determining factor in the quality of mixing achieved.

Three lines of dye were introduced into the apparatus as shown in figure 2(*a*) left or right. Photographs of these dye lines were taken through the flat bottom of the cylindrical container. Within the container the fluid motion is assumed to be essentially two-dimensional, except near the top and bottom boundaries. In order to capture the motion in a plane perpendicular to the rods, we used a sheet of light about halfway between the top and bottom of the fluid to illuminate the dye and take pictures. The true fluid motion is, of course, not exactly two-dimensional, but the assumption of plane motion within most of the container seems adequate and reasonable.

A sequence of snapshots of the dye line configurations for two different stirring protocols are shown in the two columns of figure 2.† In the left column the transpositions of stirrers are always done clockwise. In the right column the transpositions are alternately clockwise and counter-clockwise, i.e. in the first step the centre and right rod orbit clockwise; in the second step the centre and left rod orbit counter-clockwise; in the third step the centre and right rod orbit clockwise; and so on. The experiment is simple in concept, and not too difficult in practice. The photographs provide

† There appear to be some 'double lines' in figure 2. These are due to optical reflections in the photograph.

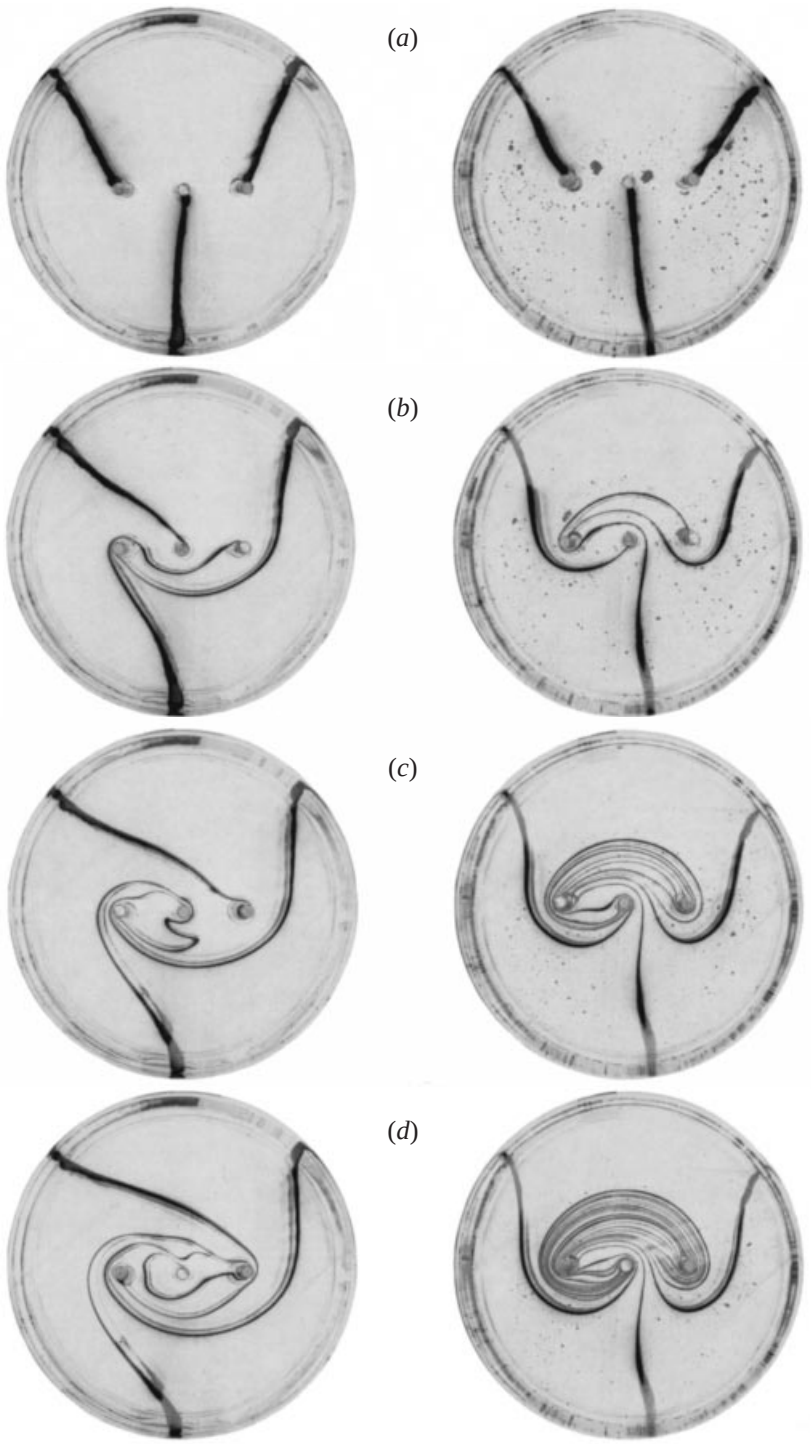

FIGURE 2 (*a–d*). For caption see facing page.

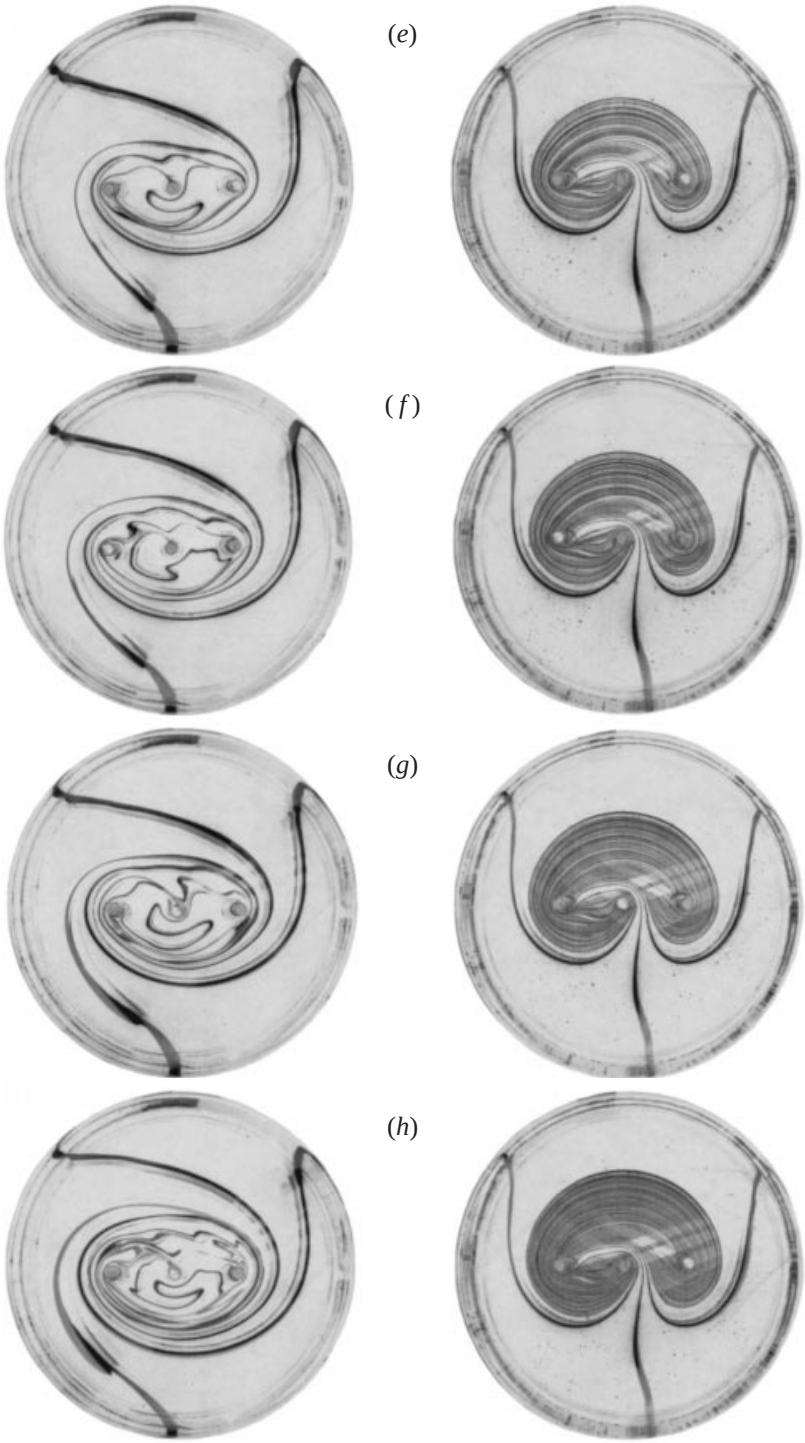

FIGURE 2. Experimental pictures of stirring by two different 'stirring protocols'. Panels on the left are for stirring by a 'finite-order' protocol (cf. §3); on the right by a pseudo-Anosov protocol. (*a*) Initial condition. Configuration after (*b*) 1; (*c*) 2; (*d*) 3; (*e*) 4; (*f*) 5; (*g*) 6; and (*h*) 9 iterations.

convincing evidence that dramatically different types of fluid stirring and mixing are taking place depending on the choice of stirring protocol. The amount of stretching and close layering of material lines is clearly much greater in the case shown in the right column of figure 2 than in the left column. Both cases display chaotic advection, but in the right column some more efficient mechanism has, somehow, been 'built in'. Furthermore, in this case a very interesting self-similar structure – shaped somewhat like an inverted heart – emerges. It appears to have a cusp at the bottom where fluid is entering it. Remarkably, this structure appears devoid of large 'islands'. After just a few steps the individual streaks of dye are so thin and so close together that a substantial amount of diffusion and blurring of individual lines must be taking place. In this sense the process seen in the right column is not reversible after even three or four steps, whereas the case in the left column maintains clearly distinct material lines even up to nine steps of the stirring process.

The reader is challenged to explain the qualitative differences between the picture sequences in the two columns of figure 2 on the basis of solutions to equation (1.1) using any reasonable approximation scheme for the velocity components. To accentuate the issues note that the topological considerations show the geometry of the stirrer orbits, the shape of the external boundary, or the cross-sections of individual stirrers to be immaterial – the same qualitative differences will appear for two similar stirring protocols applied with, say, stirrers of square cross-section in a square flow domain, and with the stirrers orbiting one another along a polygonal path. In § 2 we discuss the notion of isotopy, which makes precise the notion of invariance under continuous deformation.

In § 3 we state the classification theorem of Thurston–Nielsen theory. The most interesting and important consequence of Thurston–Nielsen theory is that for certain stirring protocols, in particular the one used in the right column of figure 2, the fluid displacement is isotopic to a pseudo-Anosov map. Pseudo-Anosov maps are generalizations of linear Anosov maps, such as the 'cat map' of the periodic square onto itself, well known from an often reproduced figure in the book by Arnol'd & Avez (1968). In § 3 we review some of the properties of linear Anosov maps and explain how the pseudo-Anosov maps arising in Thurston–Nielsen theory are related to them.

In § 4 we introduce a new point of view. We imagine the stirrers and the stirred fluid domain as the two-dimensional 'space' part of a three-dimensional space–time. In this 2+1 dimensional space–time the trajectories of the three stirrers as they are moved about in the fluid trace out a braid with three strands. For three stirrers, the case pursued in detail in § 5, each segment of the braid, corresponding to a step in the stirring protocol, can be represented by a 2×2 matrix. Longer pieces of the braid are then represented by matrices that arise by multiplying the elemental matrices. Thurston–Nielsen theory implies that the matrix representation retains information on certain features of the stirred motion, in particular its stretching rate in the pseudo-Anosov case. The theory shows that within a subdomain of the fluid the stirring is everywhere hyperbolic, i.e. that it is a stretching by a certain factor λ along one direction, and a contraction by λ^{-1} along another. The theory says nothing about the extent of this domain, but judging from figure 2 it is on the scale of the distance between stirrers. The stretching factor λ can be calculated from the 2×2 matrix just mentioned. By incompressibility this matrix has determinant 1, and λ and λ^{-1} arise as its eigenvalues. For the protocol used in the right column of figure 2 we find $\lambda = \frac{1}{2}(3 + \sqrt{5})$.

Our concluding § 6 discusses the connection of our approach and results to horseshoe maps, which have been proposed as the 'engines' of chaotic advection, and

the Melnikov method for detecting the onset of chaos when an integrable system is perturbed, which has been used extensively in chaotic advection studies. We also comment on applications to the stirring by interacting, discrete vortices, and to the design of inserts in static mixers.

The topological approach adopted here looks at a very general level of structure, eliminates many specifics of the problem, but retains enough information to draw conclusions that have wide applicability, and that therefore classify different flows into 'universality classes'. This general insight has already led workers in fluid mechanics to study topological approaches to the subject in many other contexts (see, for example, the proceedings edited by Moffatt & Tsinober 1990, or the general-audience article by Ricca & Berger 1996). The term 'topological fluid mechanics' has been used for this emerging disciplinary area.

An overview of this work was presented at the 49th Annual Meeting of the American Physical Society, Division of Fluid Dynamics in Syracuse, NY (Boyland, Aref & Stremler 1996; Aref, Boyland & Stremler 1996).

2. Isotopy

Consider a domain in the plane filled with fluid. By some stirring action this fluid domain is mapped onto itself. We wish to explore the consequences of the most basic assumptions regarding this mapping, which in the fluid mechanics literature are usually made without much comment. In keeping with a continuum fluid description, we assume the mapping instant by instant to be differentiable, one-to-one, and that its inverse is differentiable, i.e. it is a *diffeomorphism*. We shall also assume that area is preserved, corresponding to an incompressible fluid, although most of our results hold without that assumption. For a viscous fluid we insist that every point at the boundary of the fluid region moves with the (solid) boundary that delimits it. To consider an inviscid fluid, we would allow slip along the boundary.

Consider the prototypical case of fluid inside a circular disk, D, with a number, k, of identical, movable, cylindrical stirrers inserted as indicated in figure 1. The choice of disk shape or the shape of the cross-section of any stirrer is irrelevant, as we shall see, but for ease of visualization we consider all boundaries to be circular. We assume that the configuration is always of the kind indicated in figure 1 – we are not interested in singularities associated with the collision of stirrers or of stirrers hitting the boundary. The region occupied by fluid in the initial configuration is designated R_k, where the subscript reminds us of how many stirrers ('holes') there are in the fluid region. The general stirring action consists of a motion of one or more stirrers along specified trajectories within the disk such that at the end of a 'stirring cycle' the stirrers are at the same positions as when the cycle began (although generally permuted, and possibly rotated about their axes). We are also going to allow the outer boundary to rotate during the stirring cycle. The prescribed motion of stirrers and external boundary is called the *stirring protocol*. For simplicity let the fluid be very viscous, so that we may assume it only moves when the stirrers move, and that it comes to rest immediately when the stirrers stop. It will be clear from the nature of our arguments that they can be extended to other physical situations as well, but this case is the simplest to consider.

During the motion of the stirrers each fluid particle moves from an initial position to a final position. In general, instant by instant the region occupied by fluid changes due to the motion of the stirrers, i.e. the diffeomorphism is not from R_k to R_k (although the domain and range of the diffeomorphism are topologically equivalent – each a

disk with k holes). At the end of the cycle the fluid again occupies the original region R_k. We call the diffeomorphism of R_k onto R_k produced by one stirring cycle the *stirred motion.*

There is a particularly simple class of diffeomorphisms corresponding to those cases in which the stirrers simply rotate about their axes while remaining in fixed positions and in which we also allow the outer boundary to rotate. For these diffeomorphisms the fluid region is R_k for all time, not just at the end of each stirring cycle. As mentioned above, prior studies of chaotic advection in Stokes flow have employed diffeomorphisms of precisely this type (cf. Aref & Balachandar 1986; Chaiken *et al.* 1986; Jana *et al.* 1994). We call these diffeomorphisms *fixed stirrer motions.* A fixed stirrer motion, then, is a one-parameter family of diffeomorphisms, with time t as the parameter. We shall designate such a family by a Greek lowercase letter with time as a subscript, e.g. ψ_t. For $t = 0$, i.e. before any stirrer or the disk boundary has moved, we clearly have $\psi_0 = \mathrm{id}$, the identity.

In the topological theory the fixed stirrer motions are augmented very substantially by considering the class of all diffeomorphisms that map R_k onto itself while keeping the stirrers at fixed positions (and only allowing them and the outer boundary to rotate). In the terminology of mechanics we may think of this expanded class of diffeomorphisms as *virtual motions*, each of which might be produced by an externally imposed, distributed force field acting on the fluid while the stirring takes place. The augmented class of diffeomorphisms is a very large collection of mappings indeed. We shall call all of them the *fixed stirrer diffeomorphisms.* (It is even possible to relax the constraint of incompressibility and enlarge the class of fixed stirrer diffeomorphisms further to include compressible flows, but this will not be necessary in our discussion.) Every fixed stirrer motion is, of course, a fixed stirrer diffeomorphism, but the latter class contains many more mappings. It will become apparent shortly why it is important to work with this larger class of mappings even though it must contain many elements with a very artificial fluid-mechanical interpretation.

2.1. *Isotopy to the identity*

Let $h\colon R_k \to R_k$ be a stirred motion corresponding to some stirring protocol involving the k stirrers. One says that h is *isotopic to the identity* if there exists a parametrized set of fixed stirrer diffeomorphisims, ψ_τ, $0 \leqslant \tau \leqslant 1$, such that $\psi_0 = \mathrm{id}$ and $\psi_1 = h$. Thus, if h is isotopic to the identity, the same net result could have been obtained by a fixed stirrer diffeomorphism. From a fluid mechanics standpoint we might have been content to write the definition of isotopy in terms of fixed stirrer motions. The definition given is clearly much more demanding: the class of mappings in the definition is huge compared to the actual possible motions accessible by rotating the stirrers about fixed positions and rotating the outer boundary. The main point will be the topological result that *for $k \geqslant 3$ there exist stirred motions that are not isotopic to the identity*. In other words, for $k \geqslant 3$ there are stirred motions that are not equivalent to motions with a fixed geometry (up to in-place rotations) even if all possible force distributions for moving particles around are allowed!

It is obvious that for $k = 0$ (no internal stirrers) any stirred motion h is isotopic to the identity. For if h corresponds to a rotation of the disk boundary by some angle α, the family of fixed stirrer diffeomorphisms can simply be chosen such that ψ_τ corresponds to rotation by an angle $\tau\alpha$.

Already the case of a single stirrer, $k = 1$, is more challenging. Is h isotopic to the identity in this case? The answer is affirmative. Let us consider a particular version of the problem of stirring by a single stirrer in which the stirrer moves on a circular

trajectory, concentric with the disk. In this case fixed stirrer diffeomorphisms can be found in the following way: Consider the flow from a frame of reference that moves with the stirrer maintaining its axes parallel to the axes in the fixed 'laboratory' frame with which the moving frame coincides at $t = 0$. In this moving frame the geometry of disk and stirrer is fixed, except for possible rotations in place of either or both boundaries. Hence, the fixed stirrer diffeomorphisms produced by rotating the stirrer and disk boundaries in accord with the circular orbit of the stirrer augmented by the overall rotation of the stirrer will produce the same stirring in the 'laboratory' frame that the moving stirrer does. In particular, this family of fixed stirrer diffeomorphisms shows that h is isotopic to the identity.

It is from stirred motions that are not isotopic to the identity that one expects to see the effect that we call 'topological chaos'. Intuitively speaking, these stirred motions 'tangle up' the fluid so badly that the process cannot be replicated (or, equivalently, undone) within a very broad class of mappings if the configuration of stirrers is held fixed. We use the term 'topological chaos' to refer to complexity that cannot be removed by continuous deformations of the shape of the fluid region, or by continuous modification in the actual trajectories of stirrers, i.e. modifications that do not change their starting and ending points and the way in which they loop around one another. In many applications of chaotic advection with fixed stirrers one starts from an integrable situation with a homoclinic or heteroclinic point, a saddle point of the steady streamline pattern. Then one perturbs this situation by modulating the motion of the boundaries in time. Standard dynamical systems ideas of the breakdown of a smooth saddle connection lead to chaos in the advection problem due to the modulated flow. The underlying situation, however, is always one where chaos can be introduced or eliminated by continuous change of a parameter. The 'topological' chaotic advection under discussion here is of an intrinsically different origin. If certain topological features of the motion of the stirrers are established, we are assured that chaos of a predetermined type is present in the advection problem. There is no continuously variable parameter. The appearance and type of chaos is stable to changes in the shape and size of the container, to stirrer outlines and dimensions, to the exact geometry of stirrer trajectories, and so on. We use the phrase 'topological chaos' since this kind of chaotic advection is 'built in' by topological properties of the stirrer motion, in effect, by the overall design of the stirring device and not by the setting of parameters such as speed or frequency of agitation.

Basic topological theorems state that for $k = 0$ or 1 any stirred motion is isotopic to the identity. For $k = 2$ either the stirred motion itself or its second iterate is isotopic to the identity (see Birman 1975; Seifert & Threlfall 1980). This leads us to study the case of three stirrers. Recalling the literature on chaotic advection, we note that in this sense all the configurations investigated to date have been 'too simple'. It has been possible to produce chaos, of course, as has been documented many times, but the richer topological chaos that can be achieved with three (or more) stirrers has not been pursued.

2.2. *Isotopic diffeomorphisms*

We extend the concept of isotopy to isotopy between two general mappings and show that it too has a natural physical interpretation. Two diffeomorphisms are said to be isotopic if one can be deformed continuously into the other. Formally, two diffeomorphisms f and g from R_k to R_k are said to be *isotopic* if gf^{-1} is isotopic to the identity (§ 2.1), i.e. there is another diffeomorphism, h, that is isotopic to the identity, such that $g = hf$. Thinking in terms of fluid stirring in R_k, two stirred motions f and

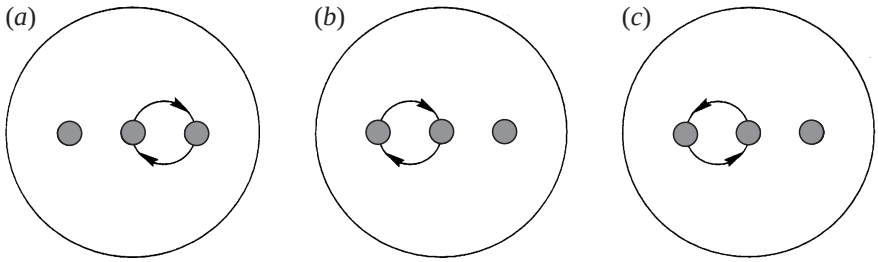

FIGURE 3. Illustration of stirred motions: (*a*) R_+; (*b*) L_+; (*c*) L_-.

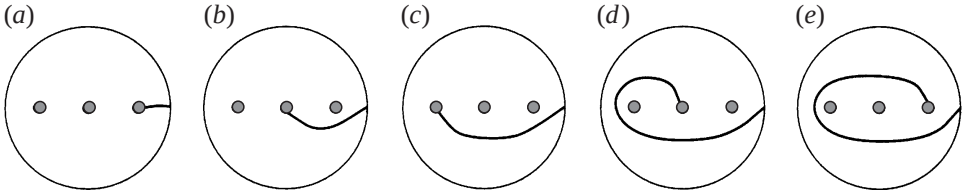

FIGURE 4. Illustration of finite-order stirred motion: (*a*) Initial position of Γ; (*b*) $R_+(\Gamma)$; (*c*) $f(\Gamma)$; (*d*) $f^2(\Gamma)$; (*e*) $f^3(\Gamma)$.

g are isotopic if we can accomplish the same stirred motion as g by first performing the stirred motion dictated by f and then follow it by a fixed stirrer diffeomorphism h. Given a stirred motion (or, in general, any diffeomorphism), f, all diffeomorphisms that are isotopic to f are called its *isotopy class*. The fixed stirrer diffeomorphisms introduced above are then simply the class of diffeomorphisms isotopic to the identity.

The notion of *isotopic curves* is closely related. By a *simple loop* in R_k we mean a closed curve that has no self-intersections. More precisely, a simple loop, Γ, can be produced by a continuous, one-to-one map from the circle, S^1, into R_k. Two simple loops, Γ_1 and Γ_2, are said to be isotopic if one can be continuously deformed into the other through a family of simple loops. Isotopy of simple arcs with their ends on the boundary of the fluid region is defined similarly. Note that isotopy allows the ends of the arc to move along the boundary corresponding either to inviscid boundary conditions or to rotation of the boundary for viscous flow.

It follows from the isotopy extension theorem (Hirsch 1994) that Γ_1 and Γ_2 are isotopic if and only if there exists a diffeomorphism, h, isotopic to the identity, such that $h(\Gamma_1) = \Gamma_2$. In particular, if a given diffeomorphism, h, is isotopic to the identity, then for any simple loop, Γ, the image of Γ under h must be isotopic to Γ.

Using the ideas just introduced we now consider two stirred motions of R_3 onto itself, one of which has its third iterate isotopic to the identity, whereas the other does not. Referring to figure 3(*a*) let R_+ denote the stirred motion produced by rotating the middle and right stirrer about the midpoint of the line joining them through 180° in a clockwise sense. Similarly, let L_+ denote the stirred motion produced by rotating the middle and left stirrer about the midpoint of the line joining them through 180° in a clockwise sense (figure 3*b*). If, instead, we rotate by 180° in the counterclockwise sense, we designate the stirred motion L_- (figure 3*c*). We now consider the two stirred motions that arise by following R_+ by L_+ and L_-, respectively, i.e. $f = L_+R_+$, and $g = L_-R_+$.

Consider the material line, Γ, a simple arc, initially connecting the right stirrer to the boundary, as shown in figures 4(*a*) and 5(*a*). Using only the continuity of the fluid

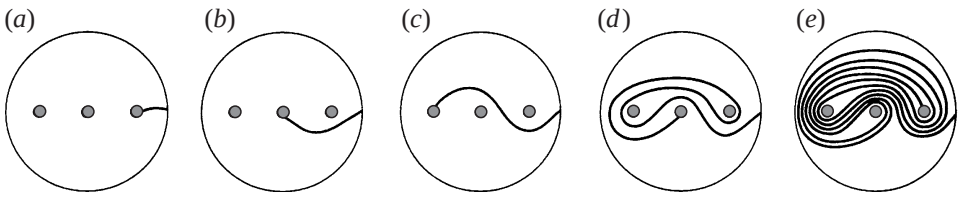

FIGURE 5. Illustration of pseudo-Anosov stirred motion: (*a*) Initial position of Γ; (*b*) $R_+(\Gamma)$; (*c*) $g(\Gamma)$; (*d*) $g^2(\Gamma)$; (*e*) $g^3(\Gamma)$.

motion we see that under R_+ the line Γ must be deformed as indicated in figures 4(*b*) and 5(*b*). We do not know, of course, exactly where the line $R_+(\Gamma)$ will be. This depends on details of the dynamics governing the fluid motion. However, this line must clearly still connect the boundary and the stirrer, and because of the motion being clockwise, it must somehow pass 'below' the right position – more precisely, it must be isotopic to what is shown in the figure. Application of the stirred motions L_+ and L_-, respectively, to produce $f(\Gamma)$ and $g(\Gamma)$, now yields the results shown in figures 4(*c*) and 5(*c*). The difference is readily apparent: no continuous fluid motion holding the stirrers fixed (i.e. no fixed stirrer motion) can deform the material line in figure 4(*c*) to that in figure 5(*c*). Thus, f is not isotopic to g. If we continue the iterations, the difference becomes ever more dramatic. Figures 4(*e*) and 5(*e*) show the results (up to isotopy) of three iterates. At this point each stirrer has returned to its initial position. By rotating the outer boundary the arc in figure 4(*e*) can be deformed back to its initial position, figure 4(*a*) up to isotopy, and f^3 is therefore isotopic to the identity. Figure 5(*e*) indicates that this clearly is not the case for g^3.

3. Thurston–Nielsen theory

It turns out that, in fact, most stirred motions are not isotopic to the identity. Thurston–Nielsen theory (see references given in § 1) describes and categorizes diffeomorphisms in terms of isotopy classes. The classification theorem, stated in precise mathematical terms in § 3.1 below, says that within each isotopy class there exists a special diffeomorphism, called the *Thurston–Nielsen* (*TN*) *representative*, which is in a precise topological and dynamical sense the simplest in the isotopy class. Once the properties of the TN representative are understood, we know what dynamical and topological complexity must be present in every diffeomorphism of the isotopy class in question.

3.1. *The Thurston–Nielsen representative*

We give first a precise statement of the

THURSTON–NIELSEN CLASSIFICATION THEOREM: *If f is a homeomorphism of a compact surface, S, then f is isotopic to a homeomorphism, φ, of one of the following types*:

(i) *Finite order*: $\varphi^n = \mathrm{id}$ *for some integer* $n > 0$;

(ii) *Pseudo-Anosov*: φ *preserves a pair of transverse, measured foliations,* $\mathscr{F}_u$ *and* $\mathscr{F}_s$, *and there is a* $\lambda > 1$ *such that* φ *stretches* $\mathscr{F}_u$ *by a factor* λ *and contracts* $\mathscr{F}_s$ *by* λ^{-1};

(iii) *Reducible*: φ *fixes a family of reducing curves, and on the complementary surfaces* φ *satisfies* (i) *or* (ii).

We pause to explain the various technical terms in the theorem. First, the theorem applies to the more general homeomorphisms, i.e. invertible, continuous, but

not necessarily differentiable mappings. For fluid mechanics applications we usually assume differentiability and so we have focused on diffeomorphisms above. The theorem deals with compact surfaces, i.e. finite regions of stirred fluid. It applies only to two-dimensional flows.

The theorem says that the TN representative, φ, is of one of three basic types: *finite order*, *pseudo-Anosov* or *reducible*. The finite-order homeomorphisms are the simplest. They have the property that φ composed with itself a certain finite number of times is the identity. For the fixed stirrer diffeomorphisms of R_k considered previously one can show that the only finite-order homeomorphism is the identity itself. Hence, the TN representative for the isotopy class containing the identity is the identity.

The second type of TN representative, the pseudo-Anosov (henceforth abbreviated pA) homeomorphisms, are dynamically very complicated. We discuss their properties further in § 3.2. When φ is of this type we expect particularly efficient stirring of the fluid. The 'cat map' on the torus is an example of a (pseudo-)Anosov map. The 'measured foliations' in the theorem statement are the collections of all the stable and unstable manifolds of all points in S. The last phrase in (ii) reflects that there is stretching and contraction everywhere by the factors λ and λ^{-1}, respectively (see § 3.2). The presence of a pA homeomorphism in an isotopy class has strong implications for the dynamics (in our case the stirring of the fluid) produced by all mappings in the class. This is brought out by another important result known as

HANDEL'S ISOTOPY STABILITY THEOREM: *If φ is pseudo-Anosov and f is isotopic to φ, then there is a compact, f-invariant set, Y, and a continuous, onto mapping $\alpha : Y \to S$, so that $\alpha f = \varphi\alpha$.*

This theorem makes precise the sense in which the dynamics of the pA map φ are present in the dynamics of any isotopic map f. The set Y contains the 'memory' of the pA dynamics, because for any $x \in S$ there is $y \in Y$ with $\alpha(y) = x$. Thus, $\varphi(x) = \varphi\alpha(y) = \alpha f(y)$. Further, from $\alpha f = \varphi\alpha$, $\alpha(f^n(y)) = \varphi^n(\alpha(y)) = \varphi^n(x)$ for all n. Thus, α sends the orbit of y under f to that of x under φ, and so every orbit of the pA map φ has a counterpart in some orbit of f contained in Y. This clarifies the meaning of the term 'topological chaos': the complex dynamics of the pA map remains after continuous deformation, i.e. under isotopy. In Handel's theorem the set Y is perhaps not all of S, and the map α may be many-to-one. This allows the dynamics of f to be more complicated than the dynamics of φ, but the dynamics of f can never be less complicated. The theorem does not, however, say anything about the size of Y relative to the size of the full domain S.

The third and last type of TN representative is called *reducible*, since in this case there is a collection of disjoint, simple loops (the 'reducing curves' in the classification theorem) with the property that they are permuted by the TN representative. Cutting along these loops, we obtain a collection of smaller domains, and on each of these the TN representative is either finite order or pA.

A given isotopy class will only contain a TN representative of one type. Thus, we can call the isotopy class itself finite order, pseudo-Anosov or reducible depending on what kind of TN representative it contains. An important question, then, is how does one tell to what type of isotopy class a given stirred motion belongs? And, after this is ascertained, what implications does the nature of the TN representative have for the dynamics and mixing properties of the stirred motion itself? There is an algorithm due to Bestvina & Handel (1995; see also Franks & Misiurewicz 1993; Los 1993) that answers both of these questions. There is a computer implementation of

this algorithm†, but the algorithm itself is quite complicated and we are content here to discuss cases that are of relevance to stirred motions with three stirrers where all calculations can be displayed explicitly.

3.2. *Anosov and pseudo-Anosov maps*

Pseudo-Anosov maps are a generalization of the more familiar Anosov mappings. We first briefly review the theory of linear Anosov diffeomorphisms on the two-dimensional torus T^2. This theory provides a convenient way to understand the properties of the pseudo-Anosov maps that are of importance here. In addition, the linear Anosov maps on T^2 are found to be closely connected with the pseudo-Anosov maps on R_3. (The appropriate piece of analysis has been relegated to the Appendix). For more detailed information on linear Anosov maps, Markov partitions and invariant foliations, see Devaney (1989), Robinson (1995) or Katok & Hasselblatt (1995).

To obtain a linear Anosov map on T^2, one starts with a positive integer matrix,

$$\boldsymbol{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{3.1}$$

with unit determinant, $ad - bc = 1$, and $a + d > 2$. The matrix $\boldsymbol{M}$ defines a mapping, f_M, on the plane given by $f_M(x, y) = (ax + by, cx + dy)$. Since this map preserves the integer lattice, it can be used to induce a map, ϕ_M, on the two-torus, T^2, thought of as a square with opposite sides identified. Perhaps the simplest such matrix is

$$\boldsymbol{M} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}. \tag{3.2}$$

In this case ϕ_M is often called *Thom's toral automorphism* or the *cat map*.

The conditions on the matrix $\boldsymbol{M}$ ensure that it has two distinct eigenvalues $\lambda > 1$ and $1/\lambda$. The eigendirection corresponding to λ is called the *unstable direction*; the eigendirection corresponding to $1/\lambda$ is called the *stable direction*. Since the Jacobian matrix of ϕ_M at every point is $\boldsymbol{M}$, the map uniformly stretches by λ in the unstable direction and contracts by $1/\lambda$ in the stable direction. All the unstable directions (i.e. all the unstable manifolds of individual points) fit together into what is called the *unstable foliation*. This structure is invariant under the action of ϕ_M. The conditions on $\boldsymbol{M}$ imply that its eigenvectors have irrational slope. Thus, the unstable foliation is a wrapping of the torus by lines with irrational slope. The *stable foliation* is defined analogously.

An Anosov diffeomorphism always gives rise to a *Markov partition* that allows one to code the orbits using symbolic dynamics (see the references given above). For a linear Anosov mapping, ϕ_M, the Markov partition can be chosen so that its *transition matrix* is $\boldsymbol{M}$. The trace of $\boldsymbol{M}^n$ counts the number of fixed points of $(\phi_M)^n$. Thus, the number of periodic orbits of order n of ϕ_M grows as λ^n.

Perhaps less well known are the topological aspects of the linear Anosov diffeomorphisms on the two-torus. Simple loops on T^2 can be represented by pairs of relatively prime integers. The pair (p, q) can be thought of as the arc in the plane that connects the origin to the point (p, q), or if this arc is 'pushed down' to the torus, a loop that wraps p times around one direction of the torus and q times around the other. The

† There is a version of this algorithm due to T. Hall available for download on the World Wide Web. Interested readers should contact the first author at boyland@math.ufl.edu for further information.

image of the loop under ϕ_M is just the loop represented by $\boldsymbol{M}$ acting on (p, q). Its nth iterate is $\boldsymbol{M}^n\binom{p}{q}$, and so lengths of loops grow like powers of the largest eigenvalue of $\boldsymbol{M}$, namely λ^n. Using a similar argument, one sees that the asymptotic direction of iterated loops is controlled by the unstable eigendirection. Thus, under iteration, all loops converge to the unstable foliation of ϕ_M. Similar remarks hold for the behaviour of sets, i.e. in our application 'patches' of advected fluid, under iteration.

Even less known is the fact that these basic properties of Anosov maps persist under isotopy (see Franks 1970). More precisely, if g is isotopic to ϕ_M, then the lengths of loops grow under iteration of g at least at the rate λ^n. Further, the periodic points of ϕ_M are present (in the appropriate sense) in g, so the number of fixed points of g^n grows at least at the rate of λ^n. Finally, under iteration by g, loops and patches eventually converge (in a precise topological sense) to the unstable foliation of ϕ_M.

Pseudo-Anosov maps are quite similar to linear Anosov maps with one essential difference. The unstable foliation of a linear Anosov map can be seen as defining a non-vanishing vector field on the two-torus. Such vector fields cannot exist on other surfaces (except the annulus). Thus, for example, no diffeomorphism of the fluid region R_k can have an unstable foliation that defines a vector field that is non-vanishing everywhere. However, the region R_k can have a homeomorphism of pA type. In pA maps there is still uniform stretching and contraction at each point by a stretch factor λ. The unstable and stable directions still fit together into invariant unstable and stable foliations, but these foliations must have a finite number of points (called singularities or prongs) where there are three or more stable directions and the same number of unstable directions. Additional singular behaviour is also allowed on the boundary.

Despite the presence of these singularities, a pseudo-Anosov map, ϕ, shares most of the basic properties of Anosov maps. There is again a Markov partition with transition matrix $\boldsymbol{M}$ that allows one to code the dynamics of ϕ, and the largest eigenvalue of $\boldsymbol{M}$ is the stretching factor $\lambda > 1$. The number of fixed points of ϕ^n grows like λ^n as does the length of non-trivial loops under iteration. In addition, the iterated loops converge to the unstable foliations. Finally, according to Handel's theorem (§ 3.1), these properties are shared by any diffeomorphism that is isotopic to ϕ.

From the perspective of fluid stirring pA diffeomorphisms are especially attractive. They give area-preserving maps of the region with uniform stretching and contraction at each point. Thus, there are no elliptic islands in which material becomes trapped, impeding mixing. In fact, this property, when stated a bit more precisely, uniquely characterizes the pA map in an isotopy class (up to a change of coordinates). Unfortunately, we do not at this time know how to realize an actual pA map over a given fluid domain by a motion of stirrers using realistic fluid dynamical equations of motion. However, we can identify many pA isotopy classes and, as noted above, the complexity of the pA map in the class then provides a lower bound for the complexity of any element in its isotopy class. Hence, using three (or more) stirrers we may realize over a subset of the fluid domain the very complicated dynamical and topological motion associated with pA maps. This is one of the main conclusions of this paper. With fixed stirrers there is no obvious advantage of increasing the number of stirrers from two to three (or more). However, once we start moving the stirrers about, there are significant advantages in terms of stirring quality to be gained by going from two to three stirrers if we move them appropriately. The underlying reason for these advantages is that with moving stirrers we can 'build in' topological chaos to the stirring process, as we have seen.

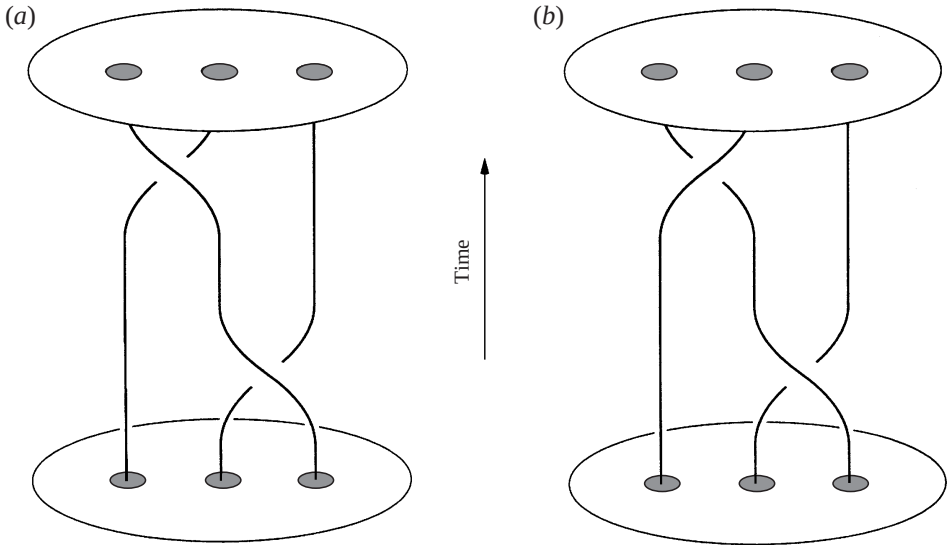

FIGURE 6. Stirred motions as braids in (2+1)-dimensional space–time, here for three strands. (*a*) Finite-order stirring protocol; (*b*) pseudo-Anosov stirring protocol.

4. Braids and isotopy classes

We now introduce a new point of view which, in § 5, will be used as a means for generating a matrix representation of the stirring protocol, and a calculational procedure for stretching rates in the pA case. We augment the two-dimensional flow region with a time axis, thus creating a (2+1)-dimensional 'space–time' in which to discuss our stirring problem. We imagine the motion of stirrers and fluid particles represented in this 'space–time' by trajectories of the form $(x(t), y(t), t)$, where t is time, and $x(t)$, $y(t)$ are the coordinates of a given physical point at time t. The movement of a stirrer, or the induced movement of a fluid particle, produces in this way a continuous curve, a 'world line', in our three-dimensional space–time continuum.

We are particularly interested in the k 'world lines' of our stirrers. Because they interchange positions and move around one another, their space–time trajectories will define a *physical braid on k strands*. A formal definition follows: Pick k points, $P_1, P_2, \ldots, P_k$ in the disk D. Place two copies of D in (2+1)-dimensional space–time, one directly above the other. The first, 'earlier' copy of the disk, D_0, is on the plane $\tau = 0$ (where τ is a non-dimensional time); the second, 'later' copy, D_1, on the plane $\tau = 1$. The physical braid, b, is a collection k non-intersecting arcs or 'strands' that connect the points P_i on D_0 to the points P_i on D_1 (see figure 6). Each distinguished point in the first disk is connected to exactly one distinguished point in the second disk. Since time increases, the strands always move upward. The braid keeps track of how the stirrers are permuted during one cycle of the stirred motion.

The key is that *braids* specify isotopy classes on a region R_k. (For general information on braids, see Birman 1975.) An algebraic description of braids allows us to 'label' stirred motions and their isotopy classes. *Artin's braid group*, B_k, gives an algebraic description of braids on k strands – indeed, braids can be thought of as forming the algebraic structure known as a group. The *generators* of the group are the simple braids σ_i, $i = 1, \ldots, k$, that interchange the ith and $(i+1)$st strands, and their inverses σ_i^{-1}. We choose to let σ_i denote an interchange of stirrers i and $i+1$ such that these 'orbit' one another in a clockwise sense. Thus, if the three stirrers considered in

figures 3–5, or in figure 6, are labelled 1, 2, 3 from left to right, the stirred motion R_+ would be represented by the braid σ_2 (the first interchange of two strands in figure 6(*a*) or (*b*)). The stirred motions L_+ and L_- would be represented by σ_1 and σ_1^{-1}, respectively (see the second interchange of strands in figure 6(*a*), (*b*), respectively). The composition of two generators, $\sigma_i\sigma_j$, represents the braid obtained by putting the braid σ_i 'after' (i.e. in the sense of figure 6 'on top of') σ_j. Thus, in our earlier example in § 2, the braid corresponding to f is $\sigma_1\sigma_2$, whereas the braid corresponding to g is $\sigma_1^{-1}\sigma_2$.

Since one wants physical braids that can be continuously deformed into one another to have the same algebraic description, one needs to add certain relations to the definition of the group, namely $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$, and for $|i-j| \geqslant 2$, $\sigma_i\sigma_j = \sigma_j\sigma_i$ (see Birman 1975, figure p. 8).

An element of the braid group will be called a *mathematical braid* to distinguish it from a physical braid. Each physical braid can be assigned a mathematical braid by projecting it onto a given plane perpendicular to the flow plane and keeping track of over- and under-crossings. Changing the projection plane can clearly change the resulting braid. It is not too difficult to see that the resulting change is always conjugation in the group B_k, i.e. if the original braid is β, the braid in the new projection will be of the form $\alpha\beta\alpha^{-1}$, where the braid α reflects the change of coordinates from the old projection plane to the new one. It is also the case that if one physical braid can be deformed into another (without breaking the strands), then the two physical braids are assigned the same mathematical braid.

The braid group has many uses, but the application of importance here is its role as a device for labelling stirred motions of a region R_k and their isotopy classes. From the intuitive picture given of the correspondence between stirring protocols and physical braids (cf. figure 6) it is clear that each stirring protocol corresponds to a braid, and to each braid there corresponds a stirring protocol.

We saw in § 2 that a stirring protocol with k stirrers gives rise to a diffeomorphism of R_k. We have just seen that such a stirring protocol also yields a physical braid and thus a mathematical braid. Now we need to connect the isotopy classes of the stirred motions to mathematical braids. There is almost a unique correspondence, but since we have allowed rotation along the boundaries of R_k in defining isotopy, we must be a bit careful in stating this correspondence. For example, the stirred motion on three strands represented by $(\sigma_1\sigma_2)^3$ is isotopic to the identity because it can be 'undone' by rotating the outer boundary once. The correct statement is that two stirring protocols with k stirrers yield isotopic diffeomorphisms if (and only if) their mathematical braids are equal up to multiplication by some number of full twists of the outer boundary of R_k. In algebraic language the precise statement is that the elements of the braid group are equal in B_k modulo the subgroup generated by $(\sigma_1\sigma_2 \ldots \sigma_{k-1})^k$.

Intuitively, this last result is seen to be true as follows: consider two stirring protocols and the physical braids they generate. If these can be deformed into one another, it implies, on one hand, that they have the same mathematical braid and, on the other hand, that the three-dimensional deformation that takes one physical braid to the other can be used to generate a family of two-dimensional diffeomorphisms that carry out the isotopy of one stirred motion to the other. In more mathematical terms, the proof proceeds by showing that the set of isotopy classes is a group with multiplication as the rule of composition. This group is then shown to be isomorphic to the quotient of B_k discussed above by showing that the isotopy class group is generated by classes that switch two stirrers, i.e. by the classes that correspond to the elements σ_i. Further details may be found in Boyland, Stremler & Aref (1999).

FIGURE 7. Transformation of the two lines, I and II, connecting the stirrers under R_+. (*a*) Initial position of I and II; (*b*) position of transformed segments I′ and II′ up to deformation.

5. The case $k = 3$

The case $k = 3$ is the first for which Thurston–Nielsen theory allows the existence of a stirred motion isotopic to a pA diffeomorphism. We expect such motions to stir the fluid particularly well. Fortunately, for this case virtually all aspects of the Thurston–Nielsen theory can be computed explicitly using a connection with torus maps. We have already indicated how stirring protocols for the region R_3 lead to physical braids, and how these braids label isotopy classes. It turns out that on the two-torus each isotopy class of diffeomorphisms contains exactly one linear Anosov map, ϕ_M, of the kind discussed in § 4.2 (isotopy of diffeomorphisms on the torus is defined analogously to that on the regions R_k). In Thurston–Nielsen theory, this linear Anosov map is the TN representative in its class. This particular diffeomorphism, then, is represented by a certain matrix with integer entries, and its properties can be calculated from this matrix. The correspondence between torus maps and diffeomorphisms on R_3 preserves the matrix. Hence, each stirred motion on R_3 is represented by a certain 2×2 matrix with integer entries, which yields properties of the TN representative for the isotopy class to which the stirred motion belongs. There are a number of technical details to be considered, in particular how to map the torus onto R_3, that we have collected in the Appendix. Here we proceed directly and heuristically to a representation of the braid group on three strands by 2×2 matrices with integer entries.

5.1. *Matrix representations of braids*

Consider the set of three stirrer positions and the two material lines connecting them as shown in figure 7(*a*). Under the stirred motion corresponding to σ_2 the stirrers and material lines are transformed into figure 7(*b*) (modulo deformations that can be accomplished with fixed stirrers). Hence, since the line I is mapped to I′, which 'is like' I and II of the original configuration, and since II is mapped to II′, i.e. in essence to itself, we can represent σ_2 by the matrix

$$\boldsymbol{M}_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{5.1}$$

We may view the assignment of 0 and 1 to the entries in this matrix in the spirit of an 'incidence matrix' as used in circuit theory: the rows correspond to the original material lines I and II, and the columns to their transformed versions I′ and II′. Since I′ 'consists of' I and II, we have placed a 1 in both entries of the first column; similarly, since II′ is just II, in the second column we have placed a 0 in the first row and a 1 in the second row. Another interpretation comes from treating I and II as boxes in a Markov partition.

Analogously, the matrix representing σ_1^{-1} is taken as (see figure 8)

$$\boldsymbol{M}_1^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{5.2}$$

We now construct the matrices corresponding to other simple braids by invoking

FIGURE 8. Transformation of the two lines, I and II, connecting the stirrers under L_-. (*a*) Initial position of I and II; (*b*) position of transformed segments I′ and II′ up to deformation.

the group properties of the braid group to which this representation should be faithful. In particular, the matrix corresponding to σ_1 is

$$\boldsymbol{M}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}. \tag{5.3}$$

Thus, if, as in §§ 3 and 4, we consider the stirred motion f corresponding to the braid $\sigma_1\sigma_2$, we find that its matrix representation is

$$\boldsymbol{M}_f = \boldsymbol{M}_1\boldsymbol{M}_2 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}. \tag{5.4}$$

This matrix has the property $\boldsymbol{M}_f^3 = -\mathbf{1}$, where $\mathbf{1}$ is the 2×2 unit matrix. Hence, $\boldsymbol{M}_f^6$ is the identity, in accord with the finite-order nature of f.

On the other hand, the matrix representation of the braid $\sigma_1^{-1}\sigma_2$, which corresponds to the stirred motion g from §§ 3 and 4, is

$$\boldsymbol{M}_g = \boldsymbol{M}_1^{-1}\boldsymbol{M}_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \tag{5.5}$$

i.e. precisely the matrix that appears in the cat map. This matrix has eigenvalues $\frac{1}{2}(3 \pm \sqrt{5})$.

TN theory shows that the stirred motion 'inherits' a number of features from the underlying linear map represented by the matrix $\boldsymbol{M}_g$ (see the Appendix). Thus, the isotopy class containing the stirred motion g stretches some material lines by a factor $\frac{1}{2}(3 + \sqrt{5})$ on each iteration, and the number of fixed points of g is bounded below by $\mathrm{Tr}(\boldsymbol{M}_g^n)$. Since,

$$\boldsymbol{M}_g^n = \begin{pmatrix} F_{2n} & F_{2n-1} \\ F_{2n-1} & F_{2n-2} \end{pmatrix}, \quad n \geqslant 1, \tag{5.6}$$

with F_n the Fibonacci numbers $F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, \ldots$, we find that $\mathrm{Tr}(\boldsymbol{M}_g^n) = F_{2n} + F_{2n-2}$. We conclude that g has at least three fixed points; that at least seven points are fixed under g^2 (the three fixed points are included here, so there are two different period-two *orbits*); at least 18 points are fixed under g^3 (so there are five different period-three *orbits*); at least $F_{20} + F_{18} = 15\,127$ points are fixed under g^{10}, and so on. For large n, $\mathrm{Tr}(\boldsymbol{M}_g^n)$ grows approximately as $\frac{1}{2}(3 + \sqrt{5})^n \approx 2.62^n$.

In the context of stirring with three stirrers we note that repeatedly performing the stirred motion g corresponding to $\sigma_1^{-1}\sigma_2$ is, in the sense of line stretching, the most efficient stirring protocol. More precisely, of all 'words' consisting of $2n$ letters, each letter being a generator of the braid group B_3, the 'word' $(\sigma_1^{-1}\sigma_2)^n$ yields an isotopy class for which the stretching factor is greatest (Mezic, D'Allessandro & Dahleh 1999). In more physical terms, if we are allowed to perform $2n$ switches, each being the switch of a pair of adjacent stirrers, we get the largest stretching by using a 'pigtail' braid or, equivalently, an 'eggbeater' motion.

6. Discussion and conclusions

The present paper provides an application of a major result of modern topology to the problem of stirring of a fluid. We have shown in outline how Thurston–Nielsen theory applies to a prototypical problem of the type indicated in figure 1. An illustration of the kind of improvement possible in a physical experiment has been displayed in figure 2. There are a number of remarks that should be made when relating this work to the earlier body of literature on chaotic advection. In this final section we turn to some of these issues.

6.1. *Generalized horseshoe maps*

It has long been recognized that stretching and folding is central to the creation of chaos. The simplest example of this behaviour is provided by what is known as *Smale's horseshoe map*. In the context of chaotic advection studies it has been suggested that horseshoes are the 'engines' of mixing. To discuss connections of horseshoes with TN theory we need to establish some terminology and review some mathematical background (for additional information see Franks 1982; Guckenheimer & Holmes 1983; Devaney 1989; Katok & Hasselblatt 1995).

By Smale's horseshoe map, H, we mean a specific diffeomorphism of the disk that is constructed by taking a rectangle, stretching it, folding it into a horseshoe shape, and then placing the folded structure over the original rectangle. The resulting diffeomorphism contains an invariant set, X, that is uniformly hyperbolic, i.e. H uniformly stretches in one direction and contracts in the other, and these stable and unstable directions line up under iteration. The set X contains all the periodic points (except a single attracting fixed point), and, indeed, all the so-called non-wandering points. The rectangles that are used for the construction can be used to code the dynamics as all possible sequences of 0 and 1. In other words, they form a Markov partition with transition matrix

$$\boldsymbol{N} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{6.1}$$

The eigenvalues 2 and $\frac{1}{2}$ of the matrix $\boldsymbol{N}$ give Lyapunov exponents $\pm\log 2$, while the number of periodic points fixed by H^n is given by $\mathrm{Tr}(\boldsymbol{N}^n) = 2^n$.†

A *generalized horseshoe map* (abbreviated GHS) refers to a construction that is similar to Smale's horseshoe, but one allows any number of rectangles that stretch and fold over themselves in a perhaps topologically complex manner (see figure 10 in the Appendix for an example). As with Smale's horseshoe, there is a compact, invariant set, X, on which the dynamics can be coded using the boxes as a Markov partition, but in this case we have as many symbols as boxes and some transitions may not be allowed. For simplicity we restrict the box-pulling so that X is indecomposable, i.e. has a dense orbit. This will happen if the transition matrix has an iterate that is strictly positive. Once again, the eigenvalues of the transition matrix give the Lyapunov exponents and the growth rate of the number of periodic points. Further, the structure of the invariant manifold template, i.e. the stable and unstable foliations, can be read off using the symbolic dynamics and the box-pulling picture.

In certain contexts the invariant set X is called the generalized horseshoe, and in others the map itself is given that name. In Smale's original terminology (Smale 1967) the set X is an example of a basic set, and the GHS is an *Axiom A diffeomorphism*. Whatever the terminology these maps or sets are the most common models for

† $\mathrm{Tr}(\boldsymbol{N}) = 2$, and since $\boldsymbol{N}^2 = 2\boldsymbol{N}$, it follows that $\boldsymbol{N}^n = 2^{n-1}\boldsymbol{N}$, and so $\mathrm{Tr}(\boldsymbol{N}^n) = 2^{n-1}\mathrm{Tr}(\boldsymbol{N}) = 2^n$.

chaos – the dynamics are very complicated, yet may be understood using the symbolic description.

The GHS provides a somewhat different way to view TN theory. Start with a given diffeomorphism and begin performing continuous deformations of the map. For a stirred motion we may think of changing the parameters of the stirring protocol (but not its braid!) and applying an external, distributed force field. The goal of all these deformations is to reduce or simplify the overall dynamics of the diffeomorphism. For example, we may squeeze down elliptic islands and move their eigenvalues so that they become flip saddles (i.e. saddles with negative eigenvalues); we may eliminate any sink–saddle pairs via saddle-node bifurcations; and we may reverse period-doubling (or, in general, period-multiplying) bifurcations. Since we are performing continuous deformations, all the resulting maps are in the same isotopy class, but what is left if and when the process terminates? And, if we perform the simplification process in different ways, do we get essentially different maps at the end? From the point of view of dynamics, the heart of TN theory is the answer to these questions. The simplest final map is the TN representative, and because it arises as the essentially unique termination of the simplification process, its dynamics must be present in every diffeomorphism in its isotopy class.

In simple domains, e.g. when there are zero, one or two stirrers in the disk, the dynamics can always be reduced to something simple and non-chaotic. But in other isotopy classes, the simplest map is chaotic, and so we attempt to make it hyperbolic so that its dynamics can be understood. The usual procedure in TN theory is to make the simplest map pA, in which case it is hyperbolic everywhere except for a finite number of singularities. However, we could just as well make the simplest map a GHS. The GHS has the advantage of providing a clearer visualization: the invariant manifold template is obvious, and the symbolic coding is unique (in pA maps a small set of points is multiply coded). It has the disadvantage of sometimes having a few extra periodic points. Let us call the simplest GHS in an isotopy class its *Axiom A representative*.† Figure 10 shows the Axiom A representative in the isotopy class of our pA stirring protocol.

Now, in general, each isotopy class contains many GHSs. In fact, the Birkhoff–Smale theorem says that a GHS is always present when there is a transverse heteroclinic or homoclinic intersection. GHSs are thus a ubiquitous signature of chaos. What is special about the Axiom A representative is that the structure of its GHS is intrinsic to the isotopy class. The way its Markov boxes are pulled exactly characterizes the topology that distinguishes its isotopy class from any other. One way to view the persistence of the Axiom A representative's dynamics is that *every* map in its isotopy class has 'boxes' that are pulled in a toplogically identical way. This is why the periodic orbits and, indeed, all the dynamics of the pA map, are present in every isotopic map. And, to repeat a point made earlier, this is why we use the term 'topological chaos' to describe maps isotopic to pA maps.

In contrast, Smale's horseshoe is isotopic to the identity, and thus its dynamics can be removed via isotopy. Similar remarks hold for all the GHSs that one gets from heteroclinic and homoclinic intersections. These maps are undeniably chaotic, but the chaos has a relation to the ambient topology that is fundamentally different from that of the Axiom A representative in a pA class.

This distinction is made especially clear by examining the behaviour of simple loops under iteration. As discussed in § 3, these converge to the unstable foliation

† One can also choose a map in the class with a hyperbolic strange attractor, see Appendix.

of the pA map. Topologically this is identical to their converging to the unstable manifold template of the Axiom A representative. This is expected behaviour when iterating the Axiom A representative, but the convergence holds for any isotopic diffeomorphism, and this illustrates another facet of the preservation of the invariant manifold template.

Finally, it is important to note that the theory does not say that every map in the isotopy class has the same GHS as the Axiom A representative (see Handel's theorem in § 3.1). Rather, it could have something bigger (but not smaller), and there is no information obtained from the theory about the hyperbolicity of dynamics of the stirred motion.

6.2. *Connection to Melnikov's method*

The general class of techniques that goes under the name of Melnikov methods is perhaps the most common method for showing that a system is chaotic. Both the Melnikov method and TN theory yield the existence of GHS in the dynamics, but the details of their application and the conclusions obtained are quite different.

The simplest case of application of the Melnikov method is to a periodically modulated, two-dimensional ODE, such as advection by two-dimensional flow. One usually studies the time-T diffeomorphism induced on the plane, where T is the period of a time-dependent stirring. A parameter controls the amplitude of the time-dependent component. When there is no time dependence, i.e. for stirring by the steady flow, we require that there is a homoclinic (or heteroclinic) loop. When time dependence is turned on, the Melnikov function measures (to first order) the crossing of stable and unstable manifolds, which were coincident for steady flow. If the Melnikov function crosses zero transversally, there is a transverse, homoclinic intersection for the time-T map. Thus, the Birkhoff–Smale theorem tells us that a GHS is embedded in the dynamics.

The method is very powerful because it allows one to do analytic calculations on explicit systems, the result of which can give very interesting information about the dynamics. There are many variants and refinements of the method (Wiggins 1990) and it has been extended to give additional information, such as lobe size, that is important in transport problems (Wiggins 1992; Beigie, Leonard & Wiggins 1995). One of the main shortcomings of the method is that it is perturbative. One must begin with a system in which there is an analytically known homoclinic (or heteroclinic) loop, and the computations of the method only work rigorously for small perturbations. In addition, the GHS that one obtains will always be confined to a band around the unperturbed homoclinic or heteroclinic loop. The GHS are never topologically intrinsic; they can always be deformed away. In fact, the method explicitly shows how this can happen, namely, by turning off the time dependence and reverting to advection by steady flow. Thus, the Melnikov method never yields topological chaos in the sense of this paper.

In contrast, TN theory gives dynamics that are topologically intrinsic, and holds for deformations of any size (within the isotopy class). The GHS from the Melnikov method are hyperbolic, whereas TN representative is only known to be semiconjugate to a GHS. On the other hand, TN theory gives a great deal more information about the dynamics of the GHS, since it yields the Axiom A representative. Thus, one gets such information as the number of periodic points, stretching rates, invariant manifold templates, and so on.

An obvious shortcoming of TN theory is that it only yields information if the map in question is in a pA isotopy class, which is often not the case. In fact, the domain in which the diffeomorphism is defined must be sufficiently complicated for a pA

class to even exist. Thus, most chaotic advection studies to date have used domains that are too simple for a pA class to exist, at least based on the location of the stirrers. Nevertheless, plenty of chaotic advection has been explored and reported! There is a well developed theory in which one removes known periodic orbits from the domain, thereby creating the necessary topological complexity (see Boyland 1994; Hall 1993), and so it is possible that pA isotopy classes have been produced in stirring configurations with fixed stirrers, but the experiments have not yet been analysed in these terms. A major advantage of the approach advocated in this paper is that computing the type of an isotopy class is very robust, and needs only a small amount of combinatorial data about the map, e.g. its braid.

A crucial restriction of TN theory is that it only works for diffeomorphisms in two dimensions, whereas Melnikov's method works in any dimension (although the theory and the computations become more difficult in higher dimensions). MacKay (1990) has speculated on the importance of Thurston–Nielsen theory to steady three-dimensional flows. Both TN theory and Melnikov's method only give a lower bound on the dynamical complexity of the system being analysed. They give no control, for example, over the total area occupied by elliptic islands in an area-preserving diffeomorphism. Such information is typically very important to the mixing efficiency.

6.3. *Other applications*

In this paper we have considered the application of TN theory to a relatively simple case of 'batch mixing' of a viscous fluid and we have studied two protocols with three stirrers in detail. TN theory applies to more stirrers as well, and there is much to be explored with reference to such 'metric issues' as optimizing the mixed region by using specially shaped containers, varying the stirrer diameter and its trajectory, and so on.

The stirring problem discussed here was not the first to which we applied the theory. Much as the original introduction of the concept of chaotic advection (Aref 1984) was motivated by consideration of advection by three point vortices on the infinite plane (the 'restricted' four-vortex problem) by Aref & Pomphrey (1980, 1982), we were first led to apply TN theory to the problem of advection by three point vortices of total strength zero. The vortices play the role of stirrers – with the hydrodynamics being that of ideal flow rather than Stokes flow – and one can consider braids and isotopy classes of the advecting motion produced by the three dynamically interacting point vortices. On the infinite plane the vortex dynamics problem was solved in detail some years ago (Rott 1989; Aref 1989). There are four distinct regimes of motion. We have checked that in all cases the map for the advection of a passive particle by the three vortices is in a finite-order isotopy class, or a reducible class with all finite-order components.

Recently, we have generalized the method of solution to three point vortices in a periodic strip (Aref & Stremler 1996) and to three vortices in a periodic parallelogram (Stremler & Aref 1999). In the former case one has to impose the condition that the vortices have total strength zero; in the latter this follows from the periodicity of the flow. With either type of periodic boundary condition we find some regimes of motion that lead to advection isotopic to a pA map, and for such regimes we can apply the apparatus of TN theory. We report separately on this application (Boyland *et al.* 1999).

It is also possible to use the results obtained here to design efficient 'static' mixers for steady flow in tubes. If we imagine inserting a set of three strands inside a straight tube, the mapping of inlet to outlet will play the role of our two-dimensional diffeomorphism, and the braid formed by the strands can be chosen to make this diffeomorphism isotopic to a pA map, which will imply more efficient transverse mixing. Since it is the topology and not the size of the tubes that matters, it may

be possible to enhance mixing by keeping the inserted strands relatively thin, thus avoiding the large pressure drops that often arise in mixers of this type.

The mechanism of producing topological chaos introduced here involves motions on the scale of the overall motion of the system of stirrers. Thus, one may hope that the chaotically stirred region will be of a scale set by the configuration of the stirrers and their motions, i.e. a scale that is largely under the control of the designer of the apparatus. This is certainly the case in the simple experimental illustration given above, where the chaotic region fills a region within the container of about the size one would expect based on the paths of the stirrers. In prior illustrations of chaotic advection the scale of the chaos always arises through the dynamic response of the fluid to the in-place motion of inner and outer boundaries, and the scale on which advective chaos is seen depends on the 'tuning' of some control parameter, and is difficult to anticipate or predict.

We thank V. V. Meleshko for discussions and S. T. Thoroddsen for access to laboratory facilities and for help in constructing the experimental apparatus. This work was supported in part by NSF grant CTS-9311545. M.A.S. acknowledges the support of an ONR graduate fellowship.

Appendix. Isotopy classes on the torus and on R_3.

In the case of three stirrers we are led to study the isotopy classes of diffeomorphisms of R_3 (in the notation of figure 1). Virtually all aspects of the Thurston–Nielsen theory can be computed explicitly for this case. This is due to the close connection between these isotopy classes and linear automorphisms of the two-torus, which is fairly well-known mathematical 'folklore'. The connection was used to study isotopy classes in Birman (1975) and for dynamics by Katok (1979), Birman & Williams (1983), Boyland & Franks (1989), and others. The pA map corresponding to the cat map (and represented by the braid $\sigma_1^{-1}\sigma_2$) is a familiar example. Indeed, W. Thurston and D. Sullivan painted a picture of its 'advection pattern' across from the elevators on the 7th floor of the Berkeley Mathematics Department in 1971!

As described in § 3, we obtain a map ϕ_M on the torus by projecting the linear map on the plane derived from an integer matrix

$$\boldsymbol{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{A 1}$$

In our discussion here we still require $ad - bc = 1$, but we put no restriction on the sign of the integers or on the value of the trace.

The first result is that an isotopy class of diffeomorphisms on the two-torus contains exactly one linear automorphism ϕ_M (isotopy of diffeomorphisms on the torus is defined analogously to isotopy on the regions R_k). In Thurston–Nielsen theory, this linear automorphism is the TN representative in its class. Further, the trace of the matrix $\boldsymbol{M}$ determines the type of the TN representative. If $|\mathrm{Tr}(\boldsymbol{M})| > 2$, we are in the Anosov case discussed in § 3.2 (an Anosov map is also considered to be pA). If $|\mathrm{Tr}(\boldsymbol{M})| < 2$, one can check that for $n = 3, 4$ or 6, $\boldsymbol{M}^n = \mathbf{1}$, the unit matrix, and so $(\phi_M)^n$ is the identity map.† These are all finite-order cases. Finally, if $\mathrm{Tr}(\boldsymbol{M}) = \pm 2$, $\boldsymbol{M}$ is either $\mathbf{1}$, $-\mathbf{1}$ (two more finite-order cases), or is similar to a matrix of the form

† The point is that $ad - bc = 1$ implies $\boldsymbol{M}^2 = \mathrm{Tr}(\boldsymbol{M})\boldsymbol{M} - \mathbf{1}$. Thus, $\mathrm{Tr}(\boldsymbol{M}) = 0$ implies $\boldsymbol{M}^2 = -\mathbf{1}$, and $\boldsymbol{M}^4 = \mathbf{1}$. For $\mathrm{Tr}(\boldsymbol{M}) = 1$, $\boldsymbol{M}^2 = \boldsymbol{M} - \mathbf{1}$, $\boldsymbol{M}^3 = \boldsymbol{M}^2 - \boldsymbol{M} = -\mathbf{1}$, and $\boldsymbol{M}^6 = \mathbf{1}$. Finally, for $\mathrm{Tr}(\boldsymbol{M}) = -1$, $\boldsymbol{M}^2 = -\boldsymbol{M} - \mathbf{1}$, $\boldsymbol{M}^3 = \mathbf{1}$.

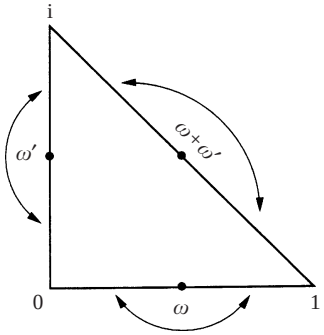

FIGURE 9. Obtaining the sphere via identification of a pair of subarcs in each side of a triangle.

$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$. This is the reducible case, since ϕ_M leaves invariant the loop in the torus corresponding to the vector (1, 0).

The next step is to connect these torus maps to homeomorphisms on R_3. We begin by recalling the construction of the toral map ϕ_M (see § 3.2). Let f_M the map of the plane induced by the matrix $\boldsymbol{M}$. Now,

$$f_M(x+n, y+m) = f_M(x, y) + (n', m') \tag{A 2}$$

for some integers n' and m'. Thus, by reducing modulo 1 in both coordinates, f_M induces a map ϕ_M on the unit square with opposite sides identified. This identification yields a topological torus, which concludes the topological description of the construction. To express it in formulae, start by letting $e(x, y) = (\exp(2\pi i x), \exp(2\pi i y))$ and consider e to be a map from the plane to the torus. The inverse of e, denoted e^{-1}, is multi-valued but the definition $\phi_M = e f_M e^{-1}$ makes sense because of (A 2), which says that the choice of the value of e^{-1} does not matter.

Next we see how similar considerations allow us to use f_M to induce a map of the sphere S^2. Equation (A 2) can be viewed as expressing a symmetry of the map f_M. Note that f_M has another symmetry, namely,

$$f_M(-x, -y) = f_M(x, y). \tag{A 3}$$

This means that we can project f_M to a map Φ_M defined on the space where we reduce points in the plane modulo integers *and* we treat a point and its opposite as the same. Equivalently, f_M induces a map on the space that we get by identifying opposing edges of the unit square as well as pairs of points that correspond under S, where S is the rotation of the unit square by π about its centre. This space is obtained from the triangle with vertices $(0, 0)$ $(1, 0)$ and $(0, 1)$ with the side identifications as indicated in figure 9. This identification yields the sphere, so we have given the topological description of the construction of the homeomorphisms of the sphere. To express this construction in formulae, the role of the projection e used in the torus case may be played by a Weierstrass $\wp$-function. Recall that $\wp$ is doubly periodic – we fix the periods to be $2\omega = 1$ and $2\omega' = \mathrm{i}$, so that $\wp$'s (double) poles are all on the integer lattice. We use the notation $\wp$ for the function obtained with these particular choices. Note that $\wp$ may be viewed as a map from the plane to the Riemann sphere S^2, or alternatively, from the torus to S^2. Now we define a homeomorphism of the sphere by $\Phi_M = \wp f_M \wp^{-1}$. Once again, $\wp^{-1}$ is multi-valued, but the double periodicity of $\wp$ and the symmetry, $\wp(-z) = \wp(z)$, coupled with (A 2) and (A 3), show that the choice of the value of $\wp^{-1}$ does not matter.

We have now obtained from f_M a diffeomorphism on the sphere S^2 and we want to use it to obtain a diffeomorphism on the region R_3. For this, the first necessary observation is that the point $(0,0)$ remains fixed under any map f_M, and so the corresponding point on the sphere, $\mathscr{P}(0) = \infty$, is a fixed point of Φ_M. Thus we may remove ∞ from S^2 and the map Φ_M still makes sense because $\Phi_M(\infty) = \infty$. A disk may be obtained by replacing the point removed with a circle. We can still think of Φ_M as defined on this disk if we take care with the definition on the boundary circle.

One final observation is necessary to obtain a diffeomorphism of R_3. The three points $e_1 = \mathscr{P}(\omega) = \mathscr{P}(\frac{1}{2}), e_2 = \mathscr{P}(\omega + \omega') = \mathscr{P}((1+\mathrm{i})/2)$, and $e_3 = \mathscr{P}(\omega') = \mathscr{P}(\mathrm{i}/2)$ are permuted by Φ_M, since for any matrix $\boldsymbol{M}$ of the type considered here

$$\boldsymbol{M}\begin{pmatrix} \dfrac{m}{2} \\ \dfrac{n}{2} \end{pmatrix} = \begin{pmatrix} \dfrac{am+bn}{2} \\ \dfrac{cm+dn}{2} \end{pmatrix}, \tag{A 4}$$

and the condition $ad - bc = 1$ implies that the numerators cannot both be even unless m and n are both even. So, we may also remove the three additional points e_1, e_2, and e_3 and replace them by circles that are permuted under Φ_M. Thus we obtain a homeomorphism on R_3 that we shall denote Ψ_M.

Recalling that we may also think of $\mathscr{P}$ as a map from the torus to S^2, we see that we could also define $\Phi_M = \mathscr{P}\phi_M\mathscr{P}^{-1}$. Thus, Ψ_M can be completely understood from the corresponding toral automorphism ϕ_M. In particular, each Ψ_M will be the TN representative in its isotopy class and its TN type will be the same as the corresponding toral automorphism ϕ_M. Furthermore, each isotopy class contains a diffeomorphism Ψ_M, but note that as a consequence of (A 3) $\boldsymbol{M}$ and $-\boldsymbol{M}$ will yield the same map on the sphere.

The main result of these considerations is that to determine the TN type of an isotopy class on R_3, we just need to determine the Ψ_M in that class. The key to finding the appropriate $\boldsymbol{M}$ lies in the connection between the braid group coordinates for isotopy classes and the algebraic structure of the collection of matrices $\boldsymbol{M}$. The algebraic fact that we need is that any matrix $\boldsymbol{M}$ of the type we have been considering here can be written as a product

$$\boldsymbol{M} = \boldsymbol{L}^{n_1}\boldsymbol{R}^{n_2}\ldots\boldsymbol{L}^{n_{k-1}}\boldsymbol{R}^{n_k}, \tag{A 5a}$$

where $\boldsymbol{R}$ and $\boldsymbol{L}$ are the matrices $\boldsymbol{M}_2$ and $\boldsymbol{M}_1$, respectively, from § 6,

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{L} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \tag{A 5b}$$

and the n_i are positive or negative integers (we allow $n_1 = 0$ or $n_k = 0$; Coxeter & Moser 1972, p. 85). The induced maps, Ψ_R and Ψ_L, respectively, provide the key to connecting the matrix $\boldsymbol{M}$ to the isotopy class of Ψ_M and its braid description.

In order to understand the braids of the maps Ψ_R and Ψ_L we need to see how they move the points e_1, e_2 and e_3. These points are the constants designated by the same symbols in the theory of elliptic functions. With our choice of periods it turns out that $e_2 = 0$, and e_1 and e_3 are both real with $0 < e_1 = -e_3 (\approx -1.72)$. Now, viewing the plane from the negative real axis, one checks that Ψ_L yields an isotopy class with braid description σ_1 while Ψ_R yields σ_2. The correspondence of matrices to braids respects multiplication, i.e. the matrix corresponding to $\boldsymbol{M}$ in (A 5a) is the braid

$$\beta_M = \sigma_1^{n_1}\sigma_2^{n_2}\ldots\sigma_1^{n_{k-1}}\sigma_2^{n_k}. \tag{A 5c}$$

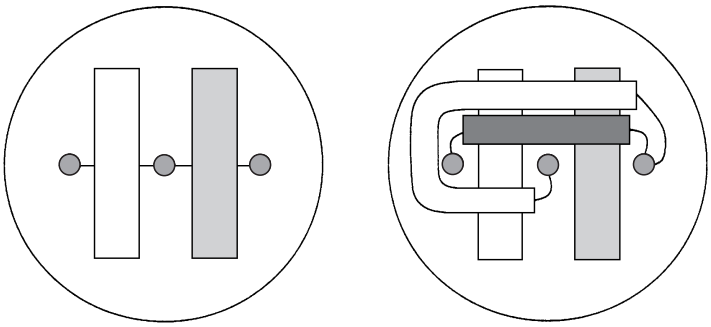

FIGURE 10. The motion of the Markov boxes in the Axiom A representative in the class of the pseudo-Anosov stirring protocol.

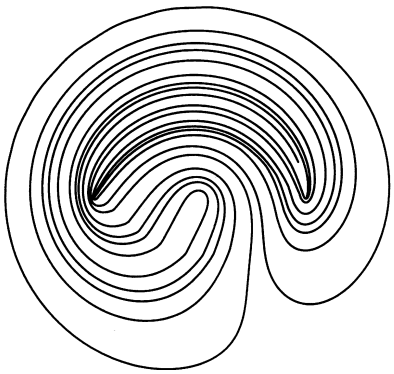

FIGURE 11. An unstable manifold of the pseudo-Anosov map in the class of the pseudo-Anosov stirring protocol (compare figure 2, right column).

We would like to invert this process, namely for a braid β written in terms of the generators we replace each occurrence of σ_1 by $\boldsymbol{L}$ and each occurrence of σ_2 by $\boldsymbol{R}$ (and similarly for inverses) and so obtain a matrix $\boldsymbol{M}$ such that Ψ_M has isotopy class 'coordinates' β_M. For this to make sense we need to check that braids that are equal (for example $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$) yield the same matrix. This is a straightforward computation. Recall also that our braid coordinates were only defined up to full twists around the outer boundary, i.e. up to multiplication by a power of $(\sigma_1\sigma_2)^3$. The matrix corresponding to $(\sigma_1\sigma_2)^3$ is minus the identity matrix, so its presence will not influence the absolute value of the trace of $\boldsymbol{M}$. Thus, for example, the braid $\sigma_1^{-1}\sigma_2$ corresponds to the matrix $\boldsymbol{L}^{-1}\boldsymbol{R} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, and $\sigma_1\sigma_2$ corresponds to $\boldsymbol{LR} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$ (cf. § 5.1).

In summary then, given a stirring protocol with three stirrers, we generate a braid as described in § 5. After writing this braid in terms of the generators σ_1 and σ_2, we obtain the matrix $\boldsymbol{M}$. The map Ψ_M is then isotopic to the stirring diffeomorphism generated by the protocol. Further, the trace of $\boldsymbol{M}$ determines the TN type of Ψ_M and thus of the isotopy class. In particular, if $|\mathrm{Tr}(\boldsymbol{M})| > 2$, we are in the pseudo-Anosov case and we know that our stirred motion must have complicated topology and dynamics. The eigenvalues of $\boldsymbol{M}$ give the stretching factor. To understand the topology of the invariant manifold template it is usually easiest to study the Axiom A representative (as described in § 6.1). We may construct this map using the arcs

shown in figure 7(*a*). These arcs are 'fattened' into Markov boxes and then pulled as required by the topology. Figure 10 shows the Axiom A representative in the class of our pA stirring protocol, i.e. for the braid $\sigma_1^{-1}\sigma_2$. The map in this isotopy class with a strange attractor is the Plykin map (cf. Guckenheimer & Holmes 1983, p. 264).

Figure 11 shows a portion of the unstable manifold of a point for the pA map represented by $\sigma_1^{-1}\sigma_2$. The picture was produced using the Weierstrass $\mathscr{P}$-function to project an unstable manifold of the cat map (i.e. the unstable eigendirection) onto the plane. Since $\mathscr{P}$ has poles, some rescaling was necessary. As noted in § 3.2, under iteration by a pA map arcs, loops and sets converge to the unstable foliation. This property is shared by any map isotopic to the pA map. Thus, this structure should emerge under repetition of the pA stirring protocol in a physical fluid. A comparison of figure 11 and the right column of figure 2 clearly shows this to be the case.

REFERENCES

Aref, H. 1984 Stirring by chaotic advection. *J Fluid Mech.* **143**, 1–21.

Aref, H. 1989 Three-vortex motion with zero total circulation: Addendum. *Z. Angew. Math. Phys.* **40**, 495–500.

Aref, H. 1990 Chaotic advection of fluid particles. *Phil. Trans. R. Soc. Lond.* A **333**, 273–289.

Aref, H. 1991 Stochastic particle motion in laminar flows. *Phys. Fluids* A **3**, 1009–1016.

Aref, H. (Ed.) 1994 *Chaos Applied to Fluid Mixing*. Special issue of *Chaos, Solitons Fractals* **4**, 380 pp.

Aref, H. & Balachandar, S. 1986 Chaotic advection in a Stokes flow. *Phys. Fluids* **29**, 3515–3521.

Aref, H., Boyland, P. L. & Stremler, M. A. 1996 Topological fluid mechanics of stirring: Applications. *Bull. Am. Phys. Soc.* **41**, 1683.

Aref, H. & Pomphrey, N. 1980 Integrable and chaotic motions of four vortices. *Phys. Lett.* A **78**, 297–300.

Aref, H. & Pomphrey, N. 1982 Integrable and chaotic motions of four vortices I: The case of identical vortices. *Proc. R. Soc. Lond.* A **380**, 359–387.

Aref, H. & Stremler, M. A. 1996 On the motion of three point vortices in a periodic strip. *J. Fluid Mech.* **314**, 1–25.

Arnol'd, V. I. & Avez, A. 1968 *Ergodic Problems of Classical Mechanics*. Benjamin.

Beigie, D., Leonard, A. & Wiggins, S. 1995 Invariant manifold templates for chaotic advection. *Chaos, Solitons Fractals* **4**, 5–124.

Bestvina, M. & Handel, M. 1995 Train-tracks for surface homeomorphisms. *Topology* **34**, 109–140.

Birman, J. 1975 *Braids, Links and Mapping Class Groups*. Annals of Mathematics Studies, Princeton University Press.

Birman, J. & Williams, R. 1983 Knotted periodic orbits in dynamical systems II: Knot holders for fibered knots. *Contemp. Maths* **20**, 1–60.

Bouasse, H. 1931 *Tourbillons*, Vol. 1. Librairie Delagrave, Paris.

Boyland, P. 1994 Topological methods in surface dynamics. *Topology Appl.* **58**, 223–298.

Boyland, P. L., Aref, H. & Stremler, M. A. 1996 Topological fluid mechanics of stirring: Theory. *Bull. Am. Phys. Soc.* **41**, 1683.

Boyland, P. & Franks, J. 1989 *Notes on Dynamics of Surface Homeomorphisms: Lectures by P. Boyland and J. Franks, Notes by C. Carroll, J. Guaschi and T. Hall*, pp. 1–48. University of Warwick Preprint.

Boyland, P. L., Stremler, M. A. & Aref, H. 1999 Topological fluid dynamics of point vortex motions. *Physica* D (submitted).

Casson, A. & Bleiler, S. 1988 *Automorphisms of Surfaces after Nielsen and Thurston*. Cambridge University Press.

Chaiken, J., Chevray, R., Tabor, M. & Tan, Q. M. 1986 Experimental study of Lagrangian turbulence in a Stokes flow. *Proc. R. Soc. Lond.* A **408**, 165–174.

Coxeter, H. & Moser, W. 1972 *Generators and Relations for Discrete Groups*, 3rd Edn. Springer.

Devaney, R. 1989 *An Introduction to Chaotic Dynamical Systems*. Addison–Wesley.

Fathi, A., Laudenbach, F. & Poenaru, V. 1979 Travaux de Thurston sur les surfaces. *Asterique*, 66–67.

Franks, J. 1970 *Anosov diffeomorphisms*. AMS Proceedings of Symposia in Pure Mathematics, vol. XIV, pp. 61–93.

Franks, J. 1982 *Homology and Dynamical Systems*. CBMS Lectures, vol. 49, American Mathematical Society.

Franks, J. & Misiurewicz, M. 1993 Cycles for disk homeomorphisms and thick trees. *Contemp. Math.* **152**, 69–139.

Guckenheimer, J. & Holmes, P. 1983 *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer.

Hall, T. 1993 Weak universality in two-dimensional transitions to chaos. *Phys. Rev. Lett*. **71**, 58–61.

Handel, M. 1985 Global shadowing of pseudo-Anosov homeomorphisms. *Ergod. Theor. Dyn. Syst*. **5**, 373–377.

Hirsch, M. W. 1994 *Differential Topology*. Springer.

Jana, S. C., Metcalfe, G. & Ottino, J. M. 1994 Experimental and computational studies of mixing in complex Stokes flows: the vortex mixing flow and multicellular cavity flows. *J. Fluid Mech.* **269**, 199–246.

Katok, A. 1979 Bernoulli diffeomorphisms on surfaces. *Ann. Math.* **110**, 529–547.

Katok, A. & Hasselblatt, B. 1995 *Introduction to the Modern Theory of Dynamical Systems*. Encyclopedia of Mathematics and its Applications, vol. 54. Cambridge University Press.

Los, J. 1993 Pseudo-Anosov maps and invariant train tracks in the disc: a finite algorithm. *Proc. Lond. Math. Soc*. **66**, 400–430.

Lundsgaard-Hansen, V. 1993 Jakob Nielsen (1890–1959). *Math. Intelligencer* **15**(4), 44–53.

MacKay, R. S. 1990 Postscript: Knots for three-dimensional vector fields. In *Topological Fluid Mechanics* (ed. H. K. Moffatt & A. Tsinober), p. 787. Cambridge University Press.

McRobie, F. A. & Thompson, J. M. T. 1993 Braids and knots in driven oscillators. *Intl J. Bifurcation Chaos* **3**, 1343–1361.

Meleshko, V. V. & Aref, H. 1996 A blinking rotlet model for chaotic advection. *Phys. Fluids* **8**, 2393–2399.

Mezic, I., D'Allessandro, D. & Dahleh, M. 1999 Control of mixing in fluid flows: a maximum entropy approach. *IEEE Trans. Automatic Control* **44**, 1852–1863.

Moffatt, H. K. & Tsinober, A. (Eds.) 1990 *Topological Fluid Mechanics*. Cambridge University Press.

Ottino, J. M. 1989 *The Kinematics of Mixing: Stretching, Chaos and Transport*. Cambridge University Press.

Ottino, J. M. 1990 Mixing, chaotic advection and turbulence. *Ann. Rev. Fluid Mech.* **22**, 207–253.

Ricca, R. L. & Berger, M. A. 1996 Topological ideas and fluid mechanics. *Physics Today* **49** (12), 28–34.

Robinson, C. 1995 *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos*. CRC Press.

Rott, N. 1989 Three-vortex motion with zero total circulation. *Z. Angew. Math. Phys.* **40**, 473–494.

Seifert, H. & Threlfall, W. 1980 *A Textbook of Topology*. Academic.

Smale, S. 1967 Differentiable dynamical systems. *Bull. Am. Math. Soc*. **73**, 747–817.

Stremler, M. A. & Aref, H. 1999 Motion of three point vortices in a periodic parallelogram. *J. Fluid Mech.* **392**, 101–128.

Thurston, W. 1988 On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Am. Math. Soc*. **19**, 417–431.

Wiggins, S. 1990 *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer.

Wiggins, S. 1992 *Chaotic Transport in Dynamical Systems*. Springer.