

STATISTICAL PROPERTIES OF CHAOTIC SYSTEMS

D. S. ORNSTEIN AND B. WEISS

ABSTRACT. The theory of smooth dynamical systems and the theory of abstract dynamical systems (ergodic theory), although having the same roots, have for many years developed quite independently of one another. These theories have now matured to the point where they can be combined to shed light on the nature of chaotic behavior.

Introduction

1. Overview

- 1.1 Smooth dynamical systems
- 1.2 Abstract dynamical systems
 - 1.2.1 Abstract systems
 - 1.2.2 Concrete systems
- 1.3 Stationary processes
- 1.4 The picture from the point of view of abstract dynamical systems
- 1.5 Smooth systems and hyperbolic structure
- 1.6 Isomorphism of smooth systems
- 1.7 α -congruence
- 1.8 α -congruence and smooth systems
- 1.9 Historical overview of isomorphism theory of chaotic systems
- 1.10 More recent results in abstract isomorphism theory

2. α -congruence

- 2.1 Strong stochastic stability
- 2.2 Stability under random perturbations
- 2.3 Scaling time
- 2.4 Bernoulli flows and Markov processes
- 2.5 Long-term versus short-term behavior and simulation
- 2.6 Instability, or when α -congruence fails

Received by the editors June 5, 1988 and, in revised form, July 28, 1989.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 28Dxx,
28Fxx, 58Fxx, 70D05.

This work was supported in part by a grant from the National Science Foundation, DMS-09876.

- 2.7 Further directions for α -congruence
- 3. A survey of some abstract ergodic theory
 - 3.1 The \bar{d} -distance
 - 3.2 Entropy and α -entropy
 - 3.3 Extremal processes
 - 3.4 Very weak Bernoulli (VWB)
 - 3.5 Finitely determined processes (FD)
 - 3.6 Isomorphism theorems
 - 3.7 Factors, \bar{d} -limits and the relativized theory
- 4. A survey of some smooth chaotic systems
 - 4.1 Introduction
 - 4.2 Geodesic flows
 - 4.3 Anosov flows
 - 4.4 Axiom A flows
 - 4.5 Partially hyperbolic systems and billiards
 - 4.6 Proving that smooth systems are Bernoulli
 - 4.7 Smooth models for abstract systems
- 5. Proofs of the new results
 - 5.1 Proofs for strong stochastic stability (§2.1)
 - 5.2 Proofs for §2.4: Bernoulli flows and Markov processes
 - 5.3 Proofs for §2.2: Random perturbations
 - 5.4 Proofs for §2.5: Long-term versus short-term behavior
 - 5.5 Proofs §2.3: Scaling time
 - 5.6 Proofs for instability (§2.6)
 - 5.7 Some proof for §3

Appendix by David Fried

Bibliography

INTRODUCTION

Ergodic theory arose from an attempt to study the long-term¹ statistical behavior of dynamical systems. The early successes of ergodic theory were Poincaré's recurrence theorem and Birkhoff's ergodic theorem. These results were proved without using the topological or differential structure that comes with a dynamical system. This led to abstract ergodic theory which studies measure-preserving transformations (in discrete) time and measure-preserving flows (in continuous time) on abstract measure spaces or, equivalently, abstract dynamical systems where all but the probability structure is abstracted out.

¹We watch our system forever.

From another direction measure-preserving flows (or transformations) were introduced in order to model the stationary processes of probability theory. The level of abstraction of abstract ergodic theory is what allowed one to put dynamical systems and random processes into the same mathematical framework.

One of the problems concerning measure-preserving transformations that most interested ergodic theorists was the classification of the transformations that arise from independent processes—the Bernoulli shifts. The problem was considered important partly because of the role that independent processes have in probability theory and partly because of the simplicity of Bernoulli shifts. The first step in this direction was taken in 1958 when Kolmogorov and Sinai showed that not all Bernoulli shifts were isomorphic. This was done by transplanting the Shannon entropy from information theory to ergodic theory.

The problem was solved in 1970 when it was shown that Bernoulli shifts were completely classified by their entropy. The method used to prove this also gave a number of other results with the following common theme: There is a unique measure-preserving flow that is the most random flow possible (this is called the Bernoulli flow).

Independently of the development of abstract ergodic theory, Hopf, Anosov, Sinai, Pesin, etc. made great progress in elucidating the structure of specific systems. They showed that certain systems had *hyperbolic structure*, a sharp form of “sensitivity to initial conditions,” which could be used to prove things like ergodicity.

The proof of the isomorphism theorem for Bernoulli shifts did not use the independence property of Bernoulli shifts. Instead it rested on a property (FD, or finitely determined) derived from independence. It was observed that hyperbolic structure could be used to check FD. In this way many concrete systems were shown to be isomorphic to the Bernoulli flow. These are the only cases so far of chaotic (positive entropy) systems that have been determined up to isomorphism.

The main focus of this article will be the recent results that show that the isomorphisms produced by the abstract theory in some cases still preserve much of the concrete geometric structure that was abstracted away by abstract ergodic theory.

The first result along these lines was motivated by the purely topological results of the Anosov–Smale school about structural

stability. Here isomorphism theory provides a fairly exact statistical analog of these results.

Further motivation came from the growing interest in chaos. A central issue here is understanding the phenomenon of systems that obey deterministic laws such as Newton's laws, but whose behavior seems to be random. Here isomorphism theory goes beyond sensitivity to initial conditions, which is a kind of randomness, by showing that in many cases the same system can be produced by Newton's laws or by a random mechanism based on coin flipping.

A much discussed question about chaos is whether randomness comes from the effects of small random perturbations (like thermal agitation) whose cumulative effect is large or from inherent deterministic laws. Here the isomorphism theory can show that, in certain cases, the process produced by small random perturbations (where effects in principle are cumulative) can also be produced by small random perturbations in the device through which we are observing the system. These latter effects are clearly not cumulative.

Chaos is sometimes studied by computer simulation. A result in this direction is the following: Bernoulli systems and (essentially) only Bernoulli systems have the property that long-term behavior can be modeled by a finite state machine (e.g., a computer) equipped with a roulette wheel.

Another question is whether long-term behavior is observable. It turns out that Bernoulli systems also have the property that their long-term behavior can be reconstructed from an observation of the system that lasts a (sufficiently long) finite time.

The plan of this paper is the following.

§1 is a detailed introduction. §2 contains new results that relate to chaos. These results are based on results in abstract ergodic theory and results about specific smooth systems. §3 gives the abstract background on which §2 is based, while §4 gives the smooth dynamical background. §5 gives the proofs for our new results. These are based on fairly long proofs which we do not reproduce but which can be found in the literature. The appendix by David Fried is about the relationship between structural stability and stochastic stability.

1. OVERVIEW

1.1 Smooth dynamical systems. For chaotic systems there is a very strong sensitivity to initial conditions, in the sense that

arbitrarily small changes in the starting point eventually produce large changes in the trajectory. Individual trajectories are therefore not reproducible and the emphasis is placed, rather, on the overall features of ensembles of trajectories. In this spirit we form the phase space, which is the manifold M of all possible configurations of our system. If, for example, our system is a billiard ball moving with velocity 1 on a table, the phase space would be the three-dimensional manifold of all possible positions and directions of motion of the ball. The laws of physics tell us how a configuration changes and the time evolution of a configuration is represented by a trajectory, or orbit, in the phase space M . We can think of the time evolution, of the system as a whole, as a flow on M (i.e., a one-parameter family of transformations on M or the solution of an ordinary differential equation on M).

There are several approaches to the study of such flows, each of which focuses on certain features of the flow and ignores others. The features that one is concerned with are usually pinned down by the notion of equivalence. For example, the “qualitative theory of ordinary differential equations” initiated by Poincaré and extended by people like Sinai, Anosov, and Smale focuses on the topology of orbits. This is made precise by saying that flows on manifolds are equivalent if there is a homeomorphism between the manifolds that takes orbits to orbits. A stronger equivalence is also studied (in connection with structural stability) in which the manifolds are the same and the homeomorphism moves all points by less than α , thus preserving some of the geometry as well as the topology.

In many cases (e.g., Hamiltonian systems, billiard flows, geodesic flows, Axiom A attractors, etc.) there is an invariant measure on the phase space M that represents the probability that a configuration is in a certain set in the phase space. In this case we are able to study the statistical properties of the system.

Because a homeomorphism may take a set of measure 1 (probability 1) to a set of measure 0, it need not preserve the probability structure. Furthermore, a map that takes orbits to orbits need not preserve the parametrization of the orbits. Therefore, statements involving the evolution of a set E in M can get garbled under equivalence.

Because of the above, when studying the statistical properties of a dynamical system, we replace equivalence by isomorphism.

Definition of isomorphism. Two flows, f_t acting on M , preserving a measure μ , and \bar{f}_t acting on \bar{M} , preserving $\bar{\mu}$, are isomor-

phic if there is an invertible measure-preserving map ϕ between (M, μ) and $(\overline{M}, \overline{\mu})$ that takes orbits of f_t to orbits of \overline{f}_t in a time-preserving manner, except possibly for a collection of orbits of measure 0 (i.e., $\phi f_t(x) = \overline{f}_t \phi(x)$ for all x except for an invariant set of measure 0).

1.2. Abstract dynamical systems.

1.2.1. *Abstract systems.* We can think of an equivalence class under isomorphism as an abstract dynamical system. More concretely we can define an abstract dynamical to be a one-parameter family f_t of invertible measure-preserving transformations of an abstract (probability) measure space X where the map from $X \times \mathbf{R}$ to X defined by $(x, t) \mapsto f_t(x)$ is measurable (i.e., the transformations are put together in a measurable way), and $f_{t_1}(f_{t_2}) = f_{t_1+t_2}$.

1.2.2. *Concrete systems.* If we ignore sets of probability zero, then we can think of a concrete dynamical system as an abstract system together with a function P , where P is what we really see. If f_t is a flow on a manifold M , then P would be the identity function from M as an abstract measure space to M as a manifold. For an experimenter the function P might be just a finite-valued function defined on M that describes his possible observations.

1.3. **Stationary processes.** We have been describing measure-preserving flows in the context of dynamical systems. They can also arise from stationary processes² like Markov processes.

The mathematical model for a continuous time stationary process taking values in a metric space M is a flow f_t on an abstract measure space X and a function P from X to a metric space M . (P can be thought of as the result of an observation on X .)

We justify the model as follows: A stationary process (like a diffusion on M) is a measure on all possible paths.

If we start with (f_t, X, P) then for each x , the function from t to M , $P(f_t(x))$, can be thought of as a possible path and X puts a measure on these paths.

²See a formal description in the footnote at the beginning of §3.

On the other hand, given a stationary process the paths, functions from t to M , form the measure space X , P tells us where the path is at time 0, and f_t shifts the path by t .

Note that (f_t, X, P) can be thought of either as a stationary process or as a concrete system.

1.4. The picture from the point of view of abstract dynamical systems.

I. A central consequence of the isomorphism theory is the existence of a unique abstract system that is the most random abstract system possible. We call this system the Bernoulli flow or B_t . (The statement that B_t is the most random possible is, of course, not well defined and would not make sense if we did not identify isomorphic flows. However, we will now present theorems about B_t that make this clear.³)

In order to understand the nature of B_t , let us focus for a moment on the case of discrete time. It seems clear that the most random kind of stochastic process is the process of independent random variables, in its simplest version the tossing of a fair coin. The corresponding transformation, the *Bernoulli shift* $(\frac{1}{2}, \frac{1}{2})$, is defined as the shift⁴ on $A^{\mathbb{Z}}$ with product measure where A has 2 points each of measure $\frac{1}{2}$. (The general Bernoulli shift $B(\dots p_i \dots)$, $p_i > 0$, $\sum p_i = 1$ is defined similarly.) The simplest form of the isomorphism theorem says:

Theorem 1.4.1. *$B(\dots p_i \dots p_k)$ is isomorphic to $B(\dots q_j \dots q_l)$ if and only if $\sum p_i \log p_i = \sum q_j \log q_j$.*

The class of independent processes is thus fairly simple from the point of view of isomorphism. A stronger theorem is the following.

Theorem 1.4.2. *There is a flow B_t such that B_1 is isomorphic to the Bernoulli shift $(\frac{1}{2}, \frac{1}{2})$, and for any t_0 , B_{t_0} is also a Bernoulli shift; in fact we get all Bernoulli shifts by varying t_0 .⁵ B_t is unique up to a constant scaling of the time parameter (i.e., if \tilde{B}_t is another*

³The results of this section may be found in [O 1].

⁴Points in $A^{\mathbb{Z}}$ are doubly infinite sequences $\{a_i\}_{-\infty}^{\infty}$. The shifted sequence is the sequence $\{\bar{a}_i\}$, where $\bar{a}_i = a_{i-1}$, $A = \{H, T\}$, $a_j = H$ or $a_j = T$. The general Bernoulli shift is defined formally in §3.6.

⁵More precisely, finite entropy shifts. For infinite entropy, see below. Also, note that B_t refers to a one-parameter family of transformations indexed by t while B_1 and B_{t_0} refer to a single transformation in the family.

flow such that for some t_0 , \tilde{B}_{t_0} is a Bernoulli shift, then there is a constant c such that B_{ct} is isomorphic to \tilde{B}_t .

II. To further justify our claim that B_t is the most random abstract flow possible, and to see that it occupies a unique place among chaotic flows, we will compare B_t with all the flows that could be considered random or chaotic. We will describe the class of random flows by eliminating the ones that are generally considered nonrandom. These are called *completely predictable* and are characterized by the property that all observations on the system are predictable in the sense that if we make the observation at regular intervals of time (i.e., every hour on the hour), then the past determines the future. (By an observation we simply mean a measurable function P on M —we can think of this as a stationary process.) It is not hard to prove that these are exactly the flows of *zero entropy*.

Theorem 1.4.3. *Any flow f_t that is not completely predictable (i.e., f_t has positive entropy) has B_{ct} as a factor. (The c 's are those for which the entropy of B_{ct} is not greater than the entropy of f_t .) The only factors of B_t are B_{ct} , $0 < c \leq 1$.*

The flow (g_t, Y) is a factor of (f_t, X) if there is a $\phi: X \rightarrow Y$ such that the inverse image of any measurable set is a set of the same measure and $\phi(f_t(x)) = g_t(\phi(x))$ for all x except for an invariant set of measure zero.⁶ (If we insist that ϕ is one-to-one and invertible, then f_t and g_t are isomorphic.) We could also think of a factor of f_t as f_t acting on an invariant sub- σ -algebra (these arise naturally from observations that do not generate).

Theorem 1.4.3 implies that if a system has any observation that is not predictable, then the set of processes arising from observations on the system includes all the processes that can arise from observations on B_t . Furthermore, it can also be shown that if f_t is not isomorphic to B_t , then there is some observation on f_t that is more predictable than any observation on B_t in the sense that it is not VWB and not extremal. (VWB and extremal characterize the processes arising from B_t in terms of predictability and will be discussed in §3.) Furthermore, this observation cannot be approximated by observations on B_t in the sense of α -disparate (§2.6).

⁶An eigenfunction in our terminology is a rotation factor.

III. One concrete realization of B_t is given by the following stationary process: Pick $\alpha > 0$ and $\beta > 0$, with α/β irrational. Flip a coin and wait α units of time if the outcome is heads (β if the outcome is tails) before flipping again. The isomorphism theory implies that when we vary α , β , and the bias of the coin, we get flows that are isomorphic to each other and to B_t (except for changing the unit of time, i.e., $B_t \rightarrow B_{ct}$).

If α/β is rational we get the direct product of B_t and the flow of rigid rotations of the circle.

IV. So far, we have ignored the case of infinite entropy. (Note that smooth flows on compact manifolds must have finite entropy.) There are analogous results for infinite entropy that complete the picture.

If in our discussion of Bernoulli shifts we want to include all independent processes (i.e., define Bernoulli shifts as those transformations that arise from independent processes) then we must consider the cases when A is countably infinite or has a continuous part. In the countably infinite case we still have isomorphism if and only if $\sum p_i \log p_i = \sum q_i \log q_i$. If $\sum p_i \log p_i = \infty$ or if A has a continuous part, then we get the unique Bernoulli shift of infinite entropy B^∞ .

There is also a unique Bernoulli flow of infinite entropy B_t^∞ . $B_{t_0}^\infty$ is the Bernoulli shift of infinite entropy for any t_0 . If f_{t_0} is the Bernoulli shift of infinite entropy for some t_0 then f_t is isomorphic to B_t^∞ .

The only factors of B_t^∞ are B_t^∞ or B_{ct} . The only factors of B^∞ are B^∞ or $B(\dots p_i \dots)$. B_t^∞ is a factor of any flow of infinite entropy. B^∞ is a factor of any transformation of infinite entropy. A concrete realization of B_t^∞ is given by any Poisson process on the line.

V. The picture for flows not isomorphic to B_t is very complicated and poorly understood.

At one extreme we have the completely predictable (zero entropy) flows. These include billiards on a rectangular table with no obstacle and the horocycle flow on a manifold of negative curvature, and the rigid rotation of the circle.

The remaining flows, those that have some measurement that is not predictable (positive entropy), are generally considered chaotic

and are characterized by positive Lyapunov exponents.⁷

Among the flows having some nonpredictable measurement are those flows where all measurements are nonpredictable. These are called K flows or Kolmogorov flows (i.e., f_t is K if it has no zero entropy factors; it is easy to see that B_t is K).

It was once hoped that the picture would be fairly simple, at least in discrete time. The hope was that all K transformations were Bernoulli shifts (Bernoulli shifts are easily seen to be K) and that every transformation was either K , or zero entropy or the direct product of K and zero entropy (the Pinsker conjecture). Both of these conjectures turned out to be false, both in continuous and in discrete time.

Theorem 1.4.4. *There is a flow (transformation) that is not K , or zero entropy or the direct product of K and zero entropy.*

Theorem 1.4.5. *There are uncountably many nonisomorphic K flows (K transformations) of the same entropy. (For flows this means that they cannot be made isomorphic by a constant rescaling of time.)*

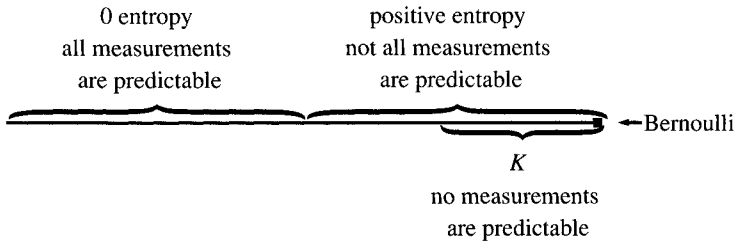


FIGURE 1

1.5. Smooth systems and hyperbolic structure. Certain chaotic systems have what is called hyperbolic structure. For a diffeomorphism D of a two-dimensional manifold, this would mean roughly, two invariant one-dimensional foliations, one of which expands exponentially under D while the other contracts exponentially. Hopf, Sinai, and Anosov used the hyperbolic structure to show that geodesic flow on manifolds of negative curvature was ergodic and in fact K . These results depended on delicate properties of the foliation and successive weakening of these conditions—Anosov flows, Axiom A flows, partially hyperbolic flows—were

⁷In the popular terminology of chaos, positive entropy corresponds to sensitivity to initial conditions. This makes sense even if there is no invariant measure.

studied. In addition, Sinai, Pesin, and many others found hyperbolic structure in a variety of specific systems. For example, Sinai [Si 4] found a partially hyperbolic structure in billiards with a convex obstacle and used this to prove ergodicity and the K property.

Hyperbolic structure was also used by Anosov [Ano 1] and Smale [Sm] to show that for certain systems, a small perturbation in the defining vector field did not change the topology of the orbits and did not change the geometry of orbits very much. More specifically there is a homeomorphism from the space on which the flow is defined, to itself that moves points by $< \epsilon$ and takes orbits of the original flow to orbits of the perturbed flow in an order-preserving way.

1.6. Isomorphism of smooth systems. Combining hyperbolic structure and a variant of the isomorphic theorem of §1.4, we proved that the *geodesic flow on a surface of negative curvature* was isomorphic to B_t . This was the first nontrivial flow whose measure-theoretic structure was determined exactly. Furthermore, the proof gave a general method for checking if a concrete flow was B_t . This has led, over the years, to showing that a very large number of very different looking concrete flows are B_t .^{8,9} (In fact, the only nontrivial case where something is shown to be isomorphic to a Bernoulli shift or flow depends on the isomorphism theorem.)

Here is a small sampling of the flows that have been shown to be isomorphic to B_t or \overline{B}_t , the direct product of B_t and the flow that rotates the circle. Geodesic flow on a manifold of negative curvature of any dimension is B_t . Axiom A attractors are B_t or \overline{B}_t [Rat 1, Bun]. The motion of a billiard ball on a table with a convex obstacle is B_t , [GO] a model for the Lorentz attractor is B_t [Rat 2]. The following theorem of Pesin [Pe] is especially striking:

Theorem 1.6.1. *Let f_t be a smooth flow on a three-dimensional compact manifold that preserves a smooth¹⁰ invariant measure.*

⁸ We can get a better feeling for the result that all geodesic flows on all surfaces of negative curvature are isomorphic (up to rescaling time) by contrasting it with the following result: Horocycle flows on surfaces of negative curvature are isomorphic (up to rescaling of time) only if the surfaces are the same (isometric). This means that the picture does not necessarily simplify when we identify isomorphic flows, and that a measure-preserving flow on an abstract measure space can, in principle, contain smooth information (we can reconstruct the surface) [Fe 0], [Ot].

⁹ Many such examples can be found in [Si 7, Part II].

¹⁰ Absolutely continuous with respect to Lebesgue measure.

After ignoring an invariant set of measure zero, we have that: Any ergodic component on which f_t is not completely predictable has positive measure and on each such component f_t is either B_t or \overline{B}_t .¹¹

B_t can also arise from a variety of stationary processes. A few examples are: A semi-Markov flow (where the particle remains in each of a finite number of states— s_i for a fixed time t_i and then jumps according to the spin of a roulette wheel) is B_t or \overline{B}_t . In particular, if the process has two states, changes states with probability $\frac{1}{2}$, and if t_1/t_2 is irrational, then the process is B_t [O 5]. Poisson processes, continuous time Markov processes on a finite number of states, or Brownian motion with reflecting barriers (so it remains in a finite interval or square) are B_t^∞ . [McCSh, AdShSmo]

Here are a few results for discrete time. In an algebraic context we have the beautiful result of Katznelson, Lind, and Miles and Thomas: [Katz, Lin, MiTn 1, 2, 3]. Any ergodic automorphism of a compact Abelian group is isomorphic to a Bernoulli shift. For one-dimensional maps we have that the natural extension of the continued fraction transformation is Bernoulli; [Smo] the natural extension of the maps $ax(1-x)$, for those a for which there is an absolutely continuous invariant measure, is Bernoulli [Led].

Because of the large variety of systems isomorphic to B_t or \overline{B}_t , we conjecture that most chaos is Bernoulli. To the extent that this is true, the diversity that we see would come from different measurements on the same abstract system. Furthermore, the diversity would be tamed because there are only a countable number of processes that can arise from observations on B_t (if we identify measurements that differ by a small amount except on a small part of the state space). Thus at the outer limit of randomness we get an extra amount of simplicity.

1.7. α -congruence. Because very different systems can be isomorphic, we propose studying a stronger kind of isomorphism which takes into account both the geometry and the statistical properties.

Definition of α -congruence. We will say that two measure-preserving flows f_t and \overline{f}_t on the same compact metric space M are

¹¹Compactness is essential. Any measure-preserving flow can be modeled by a C^∞ flow, preserving a smooth invariant measure, even on a two-dimensional manifold which is, of course, noncompact [ArOW].

α -congruent if they are isomorphic and the map ϕ from M to M that implements the isomorphism moves the points in M by $< \alpha$ except for a set of points in M of measure $< \alpha$.

If we agree that we cannot distinguish points in M that have distance $< \alpha$, and if we are willing to ignore events of probability less than α (experimental error), then α -congruent flows are indistinguishable.

Assuming ergodicity for the flow, the ergodic theorem will imply that if two flows are α -congruent, then, almost surely, orbits of corresponding points in M are within α of each other except for a set of times of density $< \alpha$. Thus α -congruent flows have essentially the same collection of orbits with the same probability distribution.

We will also define α -congruence in a more general context. If f_t and \bar{f}_t are flows on abstract measure spaces X and \bar{X} and P (and \bar{P}) are functions from $X(\bar{X})$ to a metric space then:

Definition. (f_t, X, P) and $(\bar{f}_t, \bar{X}, \bar{P})$ are α -congruent if there is an invertible measure-preserving ϕ , $\phi f_t(x) = \bar{f}_t \phi(x)$ a.e. and $d(P(x), \bar{P}(\phi x)) < \alpha$ except for a set of measure $< \alpha$ (d denotes distance in the metric space).

An alternative wording of the above definition is the following: (f_t, X, P) and $(\bar{f}_t, \bar{X}, \bar{P})$ can be jointly modeled by a third system $(f'_t, X', P' \cup \bar{P}')$ where (a) P' (and \bar{P}') generate the entire σ -algebra under f'_t , (b) (f'_t, X', P') (and (f'_t, X', \bar{P}')) give the same concrete systems as (f_t, X, P) (and $(\bar{f}_t, \bar{X}, \bar{P})$), and (c) \bar{P}' and P' differ by less than α except on a set of measure $\leq \alpha$. This says that our systems differ by a small change in the function that tells us what we see.

P and \bar{P} can be thought of as the functions that make the abstract system (f_t, X) and (\bar{f}_t, \bar{X}) into concrete systems.

1.8. α -congruence and smooth systems. One kind of α -congruence result is a statistical version of structural stability. Consider, for example, a rectangular billiard table with a convex obstacle. Pick $\alpha > 0$. Now perturb the obstacle so that the shape and the curvature change by a small amount (depending on α). The original and the perturbed billiard flow will then be α -congruent. (We may have to rescale time in one of the flows, by a small amount, i.e., change f_t to f_{ct} , $|1 - c| < \alpha$, c is a constant.) We also

get an α -congruence result for essentially the same systems and perturbations for which structural stability holds. In addition, we get α -congruence results for some systems that are not structurally stable.

The main difference between structural stability and the above α -congruence result is that the homeomorphisms in structural stability preserve the topology while the map in α -congruence preserves probabilities. Moreover, in α -congruence the parametrization of orbits is preserved (or rescaled by a constant) and therefore corresponding sets evolve in the same way. On the other hand we lose all information about events of probability zero.

Another stability result says that in certain cases adding noise to a system will have the same effect as leaving the system alone but adding noise to the device through which we view the system. Suppose we add noise to an axiom A diffeomorphism, D , by after applying D to a point x , jumping with uniform distribution in a sphere of small radius around $D(x)$. Our result says that there is a viewer whose state changes randomly but independently of D , and where the point that we see when looking through the viewer (which depends on the point we are looking at and the state of the viewer) is with high probability close to the point we are looking at. If we look at the unperturbed diffeomorphism through this viewer, we see our perturbed system exactly (all joint probabilities are the same). The interest in this result lies in the fact that the effects of the first perturbations are clearly cumulative, while the effects of the viewer are not since the viewer does not interfere with D , and only misreads a point slightly.

Another α -congruence result says that the randomness of systems isomorphic to B_t manifests itself in a more concrete way. Consider billiards with a convex obstacle. Pick $\alpha > 0$. We can define a process on the same table, whose mechanism is governed by tossing a (biased) coin, that is α -congruent to billiards. The process can be described as follows: Our ball will always be at one of a finite number of points on the table. It will stay at each point p for time $t(p)$ and will then jump to one of a pair of points according to the flip of our biased coin. (The pair of points depends on p .)

If we think of the function that implements the above α -congruence as a device or viewer which looks at the billiard ball and according to its position and velocity sees one of the special points on the table, then our result could be phrased as follows: There is a viewer that is deterministic and does not distort by $> \alpha$ for a

collection of configurations of probability $> 1 - \alpha$ and when we look at the billiards through this viewer we see a ball that moves according to the flips of the coin.

An additional α -congruence result says that the long-term behavior of a Bernoulli flow on a manifold is finitely accessible: There is an algorithm that takes an observation of our system for a sufficiently long, but finite, interval of time and produces a stationary process that could have been produced by watching our system (forever) through a viewer whose state changes randomly but with high probability, distorts very little.

1.9. Historical overview of isomorphism theory of chaotic systems. Isomorphism theory for chaotic systems began in 1958 when Kolmogorov and Sinai introduced the concept of entropy (§2.4) into ergodic theory and used it to solve a long-standing problem by showing that not all Bernoulli shifts were isomorphic. They showed that the entropy of $B(p_1, p_2, p_3)$ was $p_1 \log p_1 + p_2 \log p_2 + p_3 \log p_3$, etc. and that shifts of different entropy could not be isomorphic [Ko 1, Ko 2, Si 1].

In 1962 Sinai [Si 2] showed, in discrete time, that a Bernoulli shift was a factor of everything of equal or larger entropy. In 1967 Adler and Weiss [AdW 1,2] proved an isomorphism theorem for automorphisms of the 2-torus.

Entropy also gave the break up into completely predictable and not completely predictable and completely unpredictable [Ko 1, Ko 2, Si 1].

On the concrete side Sinai and Anosov showed that a large class of systems (including billiards with obstacles) had a “good hyperbolic structure” and used this to show that they were completely unpredictable [Si 3, Ano 1, Si 4].

Anosov also used *hyperbolic structure* to show that these systems were stable in a topological (rather than statistical) sense (structural stability) [Ano 1].

In 1970 Ornstein showed that Bernoulli shifts of the same entropy were isomorphic [O 2]. The method was different from the Adler–Weiss proof or the Sinai proof.

By a different set of ideas Ornstein showed that the completely unpredictable class contained more than the Bernoulli shifts and that not every transformation was the direct product of a completely unpredictable and a completely predictable transformation (the Pinsker conjecture) [O 7, O 8, O 9, OSh].

By the method used to prove the isomorphism theorem for Bernoulli shifts, Ornstein showed that Bernoulli shifts can be embedded in a flow and there is a unique Bernoulli flow B_t which strings together all of the Bernoulli shifts [O 6], that B_t was a factor (modulo scaling) of any flow that was not completely predictable (the continuous time extension of Sinai's 1962 result), and that the only factor of B_t is B_{ct} , $0 < c \leq 1$ [O 1].

The connection with concrete systems was made when Ornstein and Weiss showed that the geodesic flow on a surface of negative curvature is isomorphic to B_t [OW 3]. Since then it was shown that a large class of specific flows was isomorphic to the Bernoulli flow. This was done by using the hyperbolic structure (elucidated by Sinai–Anosov–Pesin and others) to check a criterion, VWB (see §2.6), which makes the isomorphism theorem work [GO, Pe, Rat, Bun].

In many cases the isomorphisms produced by the theory do not move points very much and these are the results about α -congruence and stability that we will focus on in the next section.

1.10. More recent results in abstract isomorphism theory. We have so far described the history of the isomorphism results that we have been concerned with in this paper. The subject, however, does not end here and some of the deepest results in isomorphism theory are outside the scope of this paper. We will now try to give the reader a hint of these results.

Thouvenot [Tho 1, Tho 2] introduced the idea of relativizing with respect to a factor, and showed that the isomorphism arguments could be relativized. This gave criteria for when a factor had an independent Bernoulli complement. Our results involving viewers whose state changes use this theory.

A sample result of the Thouvenot theory is that a factor of the direct product of a zero entropy and a Bernoulli is again of that form.

Thouvenot's theory can also be obtained by substituting a different metric for the \bar{d} metric of §3 [Rud 2]. Feldman introduced yet another metric and initiated the study of measure-theoretic equivalence where one weakens the definition of isomorphism by allowing orbits to map to orbits with a time change that varies from point to point along the orbit [Fe 1]. Ornstein, Rudolph, and Weiss modified the isomorphism machinery to accommodate Feldman's

metric [ORudW]. This direction culminates in Rudolph's equivalence theory [Rud 2].

Rudolph has a very deep theorem about the Bernoulliness of compact extensions of a Bernoulli shift [Rud 3, Rud 4]. The Bernoulliness of many concrete examples relies on this result. Rudolph's compact extension result together with an isomorphism theorem for actions of a skew product of the integers and a compact group allows him to classify¹² the factors of a Bernoulli shift with finite fibers [Rud 5]. There are results that indicate that the classification of all factors of a Bernoulli shift has much of the complexity of the general classification problem for transformations [O 10].

Ornstein and Weiss [OW 2] developed an isomorphism theory for actions of a general amenable, unimodular group. This builds on Rudolph's isomorphism theory for Z skew compact group, the isomorphism theorems for Z^n and R^n , Feldman's r -entropy [Fe 2], and new proofs of the Shannon–McMillan theorem.

The above results build on the original proof of the isomorphism theorem for Bernoulli shifts but in many cases there is a new level of intricacy and ingenuity.

In a different direction, Keane and Smorodinsky [KeSmo] have proved a version of the isomorphism theorem for Bernoulli shifts where the codes are finitary. The Adler–Weiss isomorphism theorem for toral automorphisms was also finitary and there is now a large and active area concerned with various kinds of finitary codes.

Lastly we should mention that there is a rich and beautiful collection of counterexamples. In particular we mention Rudolph's theory of "minimal self joinings," a systematic method of constructing certain kinds of examples [Rud 6].

2. α -CONGRUENCE

2.1. Strong stochastic stability. The individual orbits of chaotic systems are highly unstable but in certain cases the system as a whole will be stable. To discuss the stability of a certain system, we must specify the allowed perturbations and give a way of measuring their size. We must also specify the sense in which we want the perturbed system to be close to the unperturbed system. In all of our results "close" will mean α -congruent, for small α . Since

¹²We want an isomorphism of the whole system that takes corresponding factors to each other.

we have results for many kinds of perturbations, we will make a definition that focuses only on the sense of closeness.

Definition. f_t is stochastically stable under a certain kind of perturbation if given f_t and $\alpha > 0$ and if \bar{f}_t is a small enough perturbation (of the right kind) of f_t there is a constant c , $|1-c| < \alpha$ and f_t and \bar{f}_{ct} are α -congruent.

(f_t and \bar{f}_t cannot be isomorphic unless the entropies are the same. c is the unique scaling of time that makes the entropies of f_t and \bar{f}_t the same.)

We will postpone for the moment a discussion of the kinds of systems and the kinds of perturbations for which we have stochastic stability and focus on what stochastic stability tells us, by comparing it with a celebrated kind of topology stability called structural stability (or Ω stability). These topological stability results say that if f_t is a flow of a certain kind on a manifold M , then given f_t and $\alpha > 0$, if \bar{f}_t is a small enough perturbation of the right kind, there is a homeomorphism of M to M that takes orbits of f_t onto orbits of \bar{f}_t and moves points by $< \alpha$.

The conclusion of stochastic stability differs from that of structural stability in three ways: The homeomorphism is replaced by an invertible measure-preserving map. The map takes orbits rigidly onto orbits (i.e., after rescaling time, the map is an isomorphism). Instead of moving all the points by $< \alpha$, there is a set of points of measure $> 1 - \alpha$ that are moved by $< \alpha$. (We will discuss the necessity of this weakening in Remark 3 below.)

For example, if the perturbation only changes the speed of the flow and does this in a nonconstant way, then the structural stability homeomorphism could be the identity while the stochastic stability map must permute orbits.

Before giving a discussion of a general class of systems and perturbations for which we have stochastic stability, we will give some concrete examples:

Theorem 2.1.1. *Let f_t be the geodesic flow on a manifold L of negative curvature. Fix $\alpha > 0$. If \bar{f}_t is the geodesic flow on L resulting from a small (given α) change in the Riemannian structure on L ,¹³ then f_t and \bar{f}_{ct} are α -congruent where $|c-1| < \alpha$.*

¹³The corresponding metric tensors should be C^1 -close.

(Note that a geodesic flow on L is really a flow on M , the unit tangent bundle to L , and preserves the Riemannian measure on M).

Theorem 2.1.2. *Let f_t be the billiard flow on a square table with a convex obstacle. Fix $\alpha > 0$. If \bar{f}_t is the flow resulting from a small (given α) perturbation in the obstacle and its curvature, then f_t will be α -congruent to \bar{f}_{ct} where $|1 - c| < \alpha$.*

(Note that f_t (and \bar{f}_t) is a flow on a three-dimensional manifold M , consisting of positions on the table and directions; f_t (and \bar{f}_t) preserve Lebesgue measure.)

*A general result is that we have stochastic stability for a class of situations that include essentially the same systems and perturbations for which we have structural stability.*¹⁴ We will devote the rest of this section to a more careful statement of the above claim [Ma 1, Ma 2].

A result (due mainly to Smale and Mané) characterizes the flows that are structurally stable under C^1 perturbations of the defining vector field as axiom A systems (with some extra technical conditions). Axiom A systems are, except for a set of measure zero, a finite union of axiom A attractors, and we will therefore focus on the stochastic stability axiom A attractors. These are discussed in §4.4 but we will summarize some points needed for this discussion.

Axiom A attractors are smooth flows defined on a manifold M or a subset Ω of M . These flows are a generalization of geodesic flows on a manifold of negative curvature where Anosov first proved structural stability.

C^2 axiom A attractors do have a canonical invariant measure, the Sinai–Bowen–Ruelle measure (or SBR measure). This measure is believed to be physically relevant. When we consider an axiom A attractor we will always endow it with its SBR measure.

To make sense out of stochastic stability, both the flow and its perturbation must have an invariant measure. A C^2 flow that is sufficiently close C^1 to an axiom A attractor is itself an axiom A attractor and therefore has an SBR measure.

Definition of strong stochastic stability (SSS). We will say that a C^2 axiom A attractor, f_t on M , is strongly stochastically stable

¹⁴It is hoped that stochastic stability is a much more general phenomenon than structural stability, see §2.7.

or SSS if given $\alpha > 0$: if \bar{f}_t is a C^2 flow on M that is sufficiently close C^1 to f_t , then f_t and \bar{f}_{ct} are α -congruent for $|c - 1| < \alpha$.

We turn to the question: Which C^2 axiom A attractors are SSS?

We first note that C^2 axiom A attractors come in two flavors, that is, we have the following dichotomy.

Type 1. f_t is a suspension. This means that there is a nontrivial eigenfunction (i.e., a function g from M to the circle such that $g(f_t(x)) = R_{\gamma t}(g(x))$ where $R_{\gamma t}$ is rotation by γt). If f_t has a measurable eigenfunction, then it has a continuous one and is not topologically mixing. f_t is isomorphic to the direct product of the Bernoulli flow and the flow that rotates the circle.

Type 2. f_t has no measurable eigenfunction. f_t is topologically mixing. f_t is isomorphic to the Bernoulli flow.

The relevance of the above discussion is that a flow of type 1 is not isomorphic to a flow of type 2. It is also easy to see that if f_t is type 1, then we can make an arbitrarily small perturbation, \bar{f}_t , such that \bar{f}_t is type 2.

This means that a type 1 flow cannot be SSS and a type 2 flow that is the C^1 limit of type 1 flows cannot be SSS.

Our theorem is that the above are only obstructions to SSS.

Theorem 2.1.3. *A C^2 axiom A attractor is SSS if and only if it is not the C^1 limit of C^2 suspensions.*

The question now arises: Can an axiom A attractor be a C_1 limit of suspensions but not a suspension itself? David Fried clarified this situation for us and the complete answer is given in the Appendix. Here is part of the story: It can happen, but it cannot happen if f_t is Anosov of dimension ≤ 4 . It cannot happen if the expanding and contracting foliations are not jointly integrable. It cannot happen if $\dim H^1(U, R) = 1$ (U being a small open neighborhood of the attractor).

Theorem 2.1.4. *If f_t is an axiom A attractor that is not a suspension, even if it is a C_1 limit of suspensions, then one still has a lot of stability: either the perturbation \bar{f}_t is α -congruent to f_{ct} or to $R_t \times f_{ct}$, $|c - 1| < \alpha$.*

$R_t \times f_{ct}$ acts on the product of the circle and the manifold. Here R_t is the flow that rotates the circle with small period and

the isomorphism ϕ takes a point (γ, x) , where γ is on the circle and x is in M , to a point x' in M . x and x' are closer than α with probability $> 1 - \alpha$ (i.e., we are using the second definition of α -congruence).

The α -congruence to $R_t \times f_{ct}$ could be pictured as follows: We could think of R_t as a viewer whose state or condition changes cyclicly. When we look at a point x in M through the viewer, we see a point x' ($x' = \phi(\gamma, x)$) in M . x' depends only on x and the state γ of the viewer. When we look at f_{ct} through the viewer, we see exactly \bar{f}_t . We could recover the point we are looking at and the state of the viewer from the orbit that we see, i.e., there is an inverse viewer (whose state does not change).

We could, of course, also think of the α -congruence between \bar{f}_t and f_{ct} as being implemented by a viewer whereby we look at a point x in M and see a point $\phi(x)$ in M . Thus if we watch f_{ct} through this viewer, we will see \bar{f}_t (and looking at \bar{f}_t backwards through this viewer we see f_{ct}).

The main point of statistical stability is that the effect of a C^1 perturbation is in principle cumulative, whereas the distortions produced by a viewer are clearly not cumulative.

Remark 1. The difficulties associated with types 1 and 2 do not occur in the theory of structural stability because we do not insist in this theory that orbits map to orbits rigidly. If we gave up on this requirement in the statistical case, the difficulty would similarly disappear.

Proposition. *Let f_t be the flow on an axiom A attractor M . Then given $\alpha > 0$, if \bar{f}_t is a C^2 flow, sufficiently close C^1 to f_t , there is an invertible measurable¹⁵ map from M to M that takes orbits to orbits and moves all points by $< \alpha$ except for a set of measure $< \alpha$.*

Remark 2. The reason we do not strengthen SSS by requiring that ϕ move all points by $< \epsilon$ is that for axiom A flows there is only one orbit of the perturbed system that is uniformly close to a given orbit of the original system. If, for example, the perturbation only changes the speed of the flow, then each orbit must map onto itself. It is easy to see that this cannot be done rigidly and measurably unless the speed change is very special. If we strengthened the

¹⁵If ϕ is not an isomorphism, it cannot take an invariant measure to an invariant measure.

conclusion of Theorem 1 when L has dimension 2, then L and its deformation would have to be isometric. This follows from a theorem of Otal [Ot].

Remark 3. In Theorem 2.1.2 it is important not to change the curvature very much. If we approximate the boundary of our obstacles arbitrarily well by polygons, then (*), after ignoring a collection of orbits of total probability zero, any orbit of the original flow and any orbit of the perturbed flow will be D apart on the average, where D is the average distance between points in the state space. (This is because the perturbations have entropy 0.) If we take sufficiently good smooth approximations to the polygon, (*) will still hold but our perturbation will now have positive entropy and will even be Bernoulli.

Here are a few general comments on the meaning of SSS.

Two flows that are close in the C^1 -topology behave in a similar way for some finite amount of time. The closer they are, the longer this finite amount becomes, but it remains finite. Thus in a sense, knowing $\{f_t\}$ up to a small error in C^1 determines only the short-term behavior of the system. The SSS of f_t means that the approximate short-term behavior actually determines the long-term behavior arbitrarily far out in time (even though the individual orbits are sensitive to initial conditions).

The number of different f_t on M that are SSS is severely limited: For fixed α , there are only a countable number of different f_t up to α -congruence.

Stability properties are important from the point of view of experimentation or simulation. Because we can never duplicate an experiment exactly and in general we do not even know the theoretical model exactly, we need some stability properties in order to observe, simulate, or experiment with a system.

Remark 4. We also have an α -congruence theorem for axiom A diffeomorphisms, but because we cannot scale time the result is not as clean.

Theorem 2.1.5. *Let f be a connected axiom A attractor on M and pick $\alpha > 0$. If \bar{f} is a C^2 diffeomorphism on M that is sufficiently close in C^1 to f , then either f is α -congruent to the direct product of \bar{f} and a Bernoulli shift, or \bar{f} is α -congruent to the product of f and a Bernoulli shift.*

We can interpret this result as saying that one of the diffeomorphisms can be reproduced exactly (ignoring probability 0) by

looking at the other through a viewer whose state changes randomly but, with probability $> 1 - \alpha$, distorts by $< \alpha$. The viewer will not lose information in the sense that the orbit of f_t that we are looking at and the states of the viewer can be reconstructed from the orbit of the perturbed process that we see.

2.2. Stability under random perturbations. If a system is sensitive to initial conditions, then one might expect that small random effects such as thermal agitation would be magnified and that the resulting system would evolve in a very different way from the unperturbed system. On the other hand, the effects of small random perturbations in the device through which we view a system (leaving the system alone) are not cumulative (we leave an orbit alone and only misread it slightly). In spite of the above we will show that in certain cases the process that results from small perturbations whose effects are cumulative (intrinsic perturbations) could be reproduced exactly by looking at our unperturbed system through a viewer that distorts randomly but not very much.

The intrinsic perturbations that we have in mind will have infinite entropy and therefore cannot be α -congruent or even isomorphic to the unperturbed system. Some randomness has been added but our point is that the same process could have been produced by adding the randomness in an innocuous way (through a random viewer).

The general setup is the following: We will start with a flow f_t (or diffeomorphism f) on a manifold M , together with a collection of intrinsic perturbations and a way of measuring the size of these perturbations.

Definition. We will say that f_t (or f) is stable under these intrinsic perturbations if given $\alpha > 0$: if the perturbation is small enough (relative to α), we can construct a “viewer” such that

- (i) the state of the viewer changes randomly but independently of f_t ,
- (ii) the viewer distorts by $< \alpha$ with probability $> 1 - \alpha$ (the point in M that we see depends on the point in M that we are looking at and the state of the viewer),
- (iii) the process that results from looking at f_t through the viewer is identical (all joint probabilities are the same) to the process that results from the small intrinsic perturbation,

- (iv) the viewer does not involve extra information in the sense that the orbit of f_t that we are looking at and the state of the viewer can be reconstructed from the orbit of the perturbed process that we see.

Note. (iv) also means that f_t and the viewer produce our process in a nonredundant way. As well it means that f_t together with the viewer is, measure theoretically, the minimal mechanism capable of producing our process (the intrinsic perturbation).

Our viewer can be modeled as follows: The states of the viewer will form a measure space V . A measure-preserving flow on V , denoted by v_t , will govern the way that the state of the viewer changes with time. The dynamical system that results from f_t and the viewer will be the direct product of v_t and $f_t: (v_t \times f_t, V \times M)$. (The measure on $V \times M$ is a product measure.)

The way the viewer distorts will be described by a function Q from $V \times M$ to M . If we look at x in M through the viewer and the state of the viewer is w in V , then the point in M that we see will be $Q(w, x)$. The reliability of the viewer will be the expected distance between x and $Q(w, x)$ (using product measure on $V \times M$). In addition Q will generate under $v_t \times f_t$ so that Q does not lose information.

An intrinsic perturbation can be thought of as a stationary process with values in M or a measure on \overline{M} , the collection of all paths in M . We model this by $(s_t, \overline{M}, \overline{P})$ where s_t shifts each path in \overline{M} by t and \overline{P} is the function from \overline{M} to M that tells us where the path is at time 0.

Restating our previous definition in the language of α -congruence, we get:

Definition. f_t is stable under intrinsic random perturbations of a certain kind if, given $\alpha > 0$, any sufficiently small intrinsic perturbation (denote the resulting process by (s_t, X, \overline{P}) where X is the space of paths and \overline{P} tells us where the path is at time 0) is α -congruent to the direct product of f_t and some flow v_t (call it $(v_t \times f_t, V \times M)$) relative to the functions \overline{P} and \overline{Q} where $\overline{Q}(w, x) = x$, w in V and x in M .

This agrees with our previous definition if we let our viewer be $(v_t \times f_t, V \times M, Q)$ where Q is the image of \overline{P} under the α -congruence.

An alternative wording of our definition is the following: In $(v_t \times f_t, V \times M, \overline{Q})$ we can make a small change in \overline{Q} (obtaining Q), and $(v_t \times f_t, V \times M, Q)$ models our intrinsic perturbation exactly and Q generates under $v_t \times f_t$.

(A) Our first example will concern a diffeomorphism (f, M) because it is easiest to describe. The perturbation will be a Markov process on M obtained as follows: Apply f to x and then jump with uniform distribution in a ball of radius r around $f(x)$. Then apply f again to the resulting point, etc. This will give a unique probability measure on the space X of paths through M which is invariant under the shift S . Our model for this Markov process is (S, X, \overline{P}) where \overline{P} tells us where a path is at time 0. The size of the perturbation is measured by r .

Theorem 2.2.1. *An axiom A attractor that is Bernoulli is stable under the above kind of intrinsic random perturbation.*

(B) Our next example concerns perturbations that are not arbitrarily small but occur very rarely. All Bernoulli flows are stable under such perturbations (think of a billiard table with a convex obstacle that is bumped once in a great while).

Start with a stationary process that alternates between long random periods of quiet (all at least of size L) and short active periods of fixed size Δ . During the quiet periods we flow using $f_t(x)$. If we are at x at the start of an active period, then we diffuse for a period of time equal to Δ to a new point x_Δ which has an absolutely continuous distribution $\mu(x)$ which we assume dominates a fixed constant times the volume element in a ball of fixed radius around x . The nature of the diffusion is not relevant. From x_Δ one continues to flow with f_t until the end of the next quiet period. Fixing the $\mu(x)$'s and measuring the size of the perturbations by $1/L$ we have:

Theorem 2.2.2. *Every Bernoulli flow on a compact manifold preserving a smooth invariant measure is stable under the above intrinsic perturbations.*

(C) There is a continuous time analog to Theorem 2.2.1 but it is technically harder to describe because the perturbation takes place infinitesimally (i.e., our intrinsic perturbation results from a diffusion).

One formalism goes as follows. To the flow f_t we associate its differential equation

$$\frac{d(f_t x)}{dt} = V(f_t x),$$

where V is a C^2 -vector field on M . The random perturbation will be given by $w_t = (w_t^1, \dots, w_t^d)$, the standard d -dimensional Brownian motion, and C^2 -vector fields on M , X_1, \dots, X_d . The diffusion process x_t can be described by a stochastic differential equation

$$dx_t = V(x_t) dt + \sum_{i=1}^d X_i(x_t) dw_t^i$$

or by an infinitesimal generator

$$(*) \quad L = V + \frac{1}{2} \sum_{i=1}^d X_i X_i.$$

Corresponding to the transition probability measure $\mu(x)$ above, we now have $P(t, x, dy)$ which describes the distribution of a point at time t that was at x at time 0. If L is a nondegenerate elliptic operator, then these transition probabilities $P(t, x, \cdot)$ have a unique invariant measure on the manifold, say ν , and then a shift invariant measure P_ν is defined on the space of all continuous maps from \mathbf{R} to M , $X = C(\mathbf{R}, M)$, so that

$$P_\nu(\{x_t \in A\}) = \nu(A) \quad \text{for all } t$$

and

$$\begin{aligned} &P_\nu(\{x_{t_0} \in A_0, x_{t_1} \in A_1, \dots, x_{t_k} \in A_k\}) \\ &= \int \cdots \int P(t_1 - t_0, u_0, du_1) P(t_2 - t_1, u_1, du_2) \\ &\quad \cdots P(t_k - t_{k-1}, u_{k-1}, du_k) d\nu(u_0) \end{aligned}$$

for $t_0 < t_1 < \cdots < t_k$ and sets $A_0, \dots, A_n \subset M$, where the integration is over all $u_i \in A_i$, $0 \leq i \leq R$.

An invariant measure μ for the flow f_t also gives a measure on X , say dm , which is described in a similar way by saying that $m(\{x_{t_0} \in A_0, \dots, x_{t_k} \in A_k\})$ equals the μ measure of the set of those points u_0 in A_0 such that $f_{t_i - t_0}(u_0) \in A_i$. In a sense this is unnecessary for μ , since X with dm can be identified, up to a null set, with M itself. However, for the Markov process this trajectory space is essential and it is there that we will describe the stability. We metrize X by giving it the topology of uniform

convergence on compact sets. For the perturbation results we now introduce a parameter in (*) and write

$$L^\epsilon = V + \frac{\epsilon^2}{2} \sum_{i=1}^d X_i X_i,$$

μ^ϵ for the invariant measure on M , and x_t^ϵ for the corresponding Markov process. The appropriate measure on X will now be P_{μ^ϵ} .

We shall assume that the basic operator L is a nondegenerate elliptic operator in the relevant domain. In case we are interested in a proper attractor Ω_0 , f_t , with shrinking neighborhood U , then it is natural to suppose that the perturbation term is gradually killed in a smooth way so that at the boundary of U it vanishes. This guarantees that the trajectories of the Markov process never leave U .

The basic analysis of such perturbations was carried out by Y. Kifer in [Ki 3], and his results are crucial to our proof of the stability result.

Theorem 2.2.3. *A C^2 hyperbolic attractor (f_t, μ) with its SBR measure that is a Bernoulli flow is stable under the intrinsic perturbation described above.*

2.3. Scaling time. If we increase the speed of a flow, the flow will become more random in the sense that its entropy will increase. In the case of a Bernoulli flow on a compact manifold M this increase in randomness manifests itself as follows:

Theorem 2.3.0. *(Informal) $(B_{(1+\gamma)t}, M)$, $\gamma > 0$ can be reproduced exactly by watching (B_t, M) through a random viewer. If γ is small enough relative to α , then the viewer will be α -reliable. For all $\gamma > 0$ the viewer will not involve extra information in the sense that the path (of B_t) that we are looking at and the states of the viewer can be reconstructed from the path (of $B_{(1+\gamma)t}$) that we see.*

In the language of α -congruence the above theorem reads as

Theorem 2.3.1. *Let (B_t, Y, P) be the Bernoulli flow where P identifies Y with the manifold M . Given α , if $\gamma > 0$ is small enough then $(B_{(1+\gamma)t}, Y, P)$ is α -congruent to $(B_t \times B_{\gamma t}, Y \times Y, \bar{P})$ where $\bar{P}(r_1, r_2) = P(r_1)$.*

We interpret the above theorem in terms of the viewers as follows: $(B_t \times B_{\gamma t}, Y \times Y)$ is the direct product of (B_t, Y) and

$(B_{\gamma t}, Y)$. P allows us to identify (B_t, Y) and (B_t, M) . Our viewer is $(B_{\gamma t}, Y)$. If Q is the image of P under the α -congruence, then $Q(r_1, r_2)$ tells us what we see when we look at r_1 in M and the viewer is in state $r_2 (r_2 \in Y)$. α -congruence implies that Q is close to \bar{P} and thus our viewer is reliable.

We can also slow B_t by looking at it through a deterministic viewer but we cannot reconstruct the orbit we are looking at from the orbit that we see.

The above theorem can fail dramatically in the non-Bernoulli case. For example, if f_t rotates the unit circle by t , then any orbit of f_t and any orbit of f_{ct} will differ by $\frac{1}{2}$ on the average, no matter how close c is to 1, i.e.,

$$\frac{1}{T} \int_0^T |t\theta - ct\bar{\theta}| dt \rightarrow \frac{1}{2}.$$

There is also a K flow (f_t, X, P) such that the orbits produced by (f_t, X, P) and the one produced by (f_{ct}, X, P) , $c \neq 1$, differ in the same quantitative sense: There is an $\alpha > 0$ (that does not depend on c) and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T d(P(f_t(x)), P(f_{ct}(\bar{x}))) dt \geq \alpha$$

for all x, \bar{x} in X and all $c \neq 1$ and d denotes the distance between points in the range of P .

2.4. Bernoulli flows and Markov processes. Our next α -congruence results describe, in a more concrete way, the randomness of systems isomorphic to B_t . We will do this by comparing our flow f_t on a manifold M , to a process that is generally thought of as random. We will define a semi-Markov process¹⁶ on M to be a process that stays at one of a finite number of points p_i in M for time t_i and then uses a roulette wheel to decide which of the p_i to jump to. For the purposes of our comparison we could focus on a special semi-Markov process where we take 2^N points in M and label each of these by a sequence of 0's and 1's of length N . Fixing a (biased) coin (we get head with probability p), we get the label of the point to which we jump by erasing the first digit and flipping our coin to decide whether to add a 0 or a 1 to the end (e.g., $01111 \Rightarrow \begin{smallmatrix} 1111 & 0 \\ 1111 & 1 \end{smallmatrix}$). The holding times t_0, t_1 will depend in

¹⁶The theory of Markov partitions which has been developed for many of these flows is a topological notion and does not give rise to measure-theoretic Markov processes like the ones we have in mind here.

some arbitrary (but fixed) way on the label, e.g., we could take t_β with $\beta =$ middle digit of the label.

Theorem 2.4.1. *If f_t is a Bernoulli flow or \overline{B}_t on a metric space M then, given α , there is a (special) semi-Markov process on M that is α -congruent to f_t .*

We will now discuss the meaning of our theorem in the case of a billiard ball moving on a square table with a convex obstacle.

To better visualize the result we will think of our isomorphism φ (of the α -congruence) as being implemented by a measuring device or viewer. Thus, when our ball has a certain position p and velocity q , we will see, when looking through the viewer, one of a finite number of points on the billiard table, $\varphi(p, q)$. ($\varphi(p, q)$ is a path of the semi-Markov process but we will only consider the position at time zero.) We will say that our viewer is ϵ -reliable if for all but ϵ of the p, q of $\varphi(p, q)$ is within ϵ of p . Our viewer is supposed to model a measuring device of finite accuracy or resolution. Note that this viewer, unlike the viewer in §2.2, is completely deterministic.

Our theorem says the following: Given $\alpha > 0$, there exists an α -reliable viewer such that the system seen through this viewer is a semi-Markov process. Furthermore the orbit of the billiard ball can (with probability one) be reconstructed from the orbit as seen through the viewer, i.e., the appearance of randomness did not come from loss of information.

The billiard system and the billiard system seen through the viewer differ in that the billiard system as seen through the viewer produces only a finite amount of information in a finite time. This distinction, in terms of observations of finite accuracy, amounts to the following: We can reconstruct, from what we see through the viewer, the state at time 0 to any degree of accuracy by taking observations (of fixed accuracy) far into the past and future. To reconstruct the state at time 0 of the billiard seen directly we have to make observations of finer and finer accuracy, but these can be made at time 0.

Our theorem also tells us that certain semi-Markov systems could be thought of as being produced by Newton's laws (billiards seen through a deterministic viewer) or by coin flipping. This may mean that there is no philosophical distinction between processes governed by roulette wheels and processes governed by Newton's

laws¹⁷. In this connection we should note that our model for a stationary process (§1.2) means that random processes have a deterministic model. This model, however, is abstract, and there is no reason to believe that it can be endowed with any special additional structure. Our point is that we are comparing, in a strong sense, Newton's laws and coin flipping.

Variants of Theorem 2.4.1 can be found in §5.2.

2.5. Long-term versus short-term behavior and simulation. Since we cannot observe a system forever, we might ask about the connection between long-term behavior and behavior that can actually be observed. If our system is Bernoulli, then there is a good connection:

If we have no knowledge of the system except that it is isomorphic to B_t , and we are allowed to observe a single orbit then, with probability one, if we see a long enough piece of the orbit we can construct a model that is α -congruent to the direct product of the flow we are trying to model and another Bernoulli flow. Thus, even though we have watched for only a finite time, our simulation is identical to a process produced by watching our system forever through a random viewer that distorts by $< \alpha$ with probability $> 1 - \alpha$. In addition, the viewer will not involve extra information in the sense that the orbit we are watching and the state of the viewer can be reconstructed from the path as seen through the viewer.

Stating this more explicitly, we get:

Theorem 2.5.1. *Given f_t, X, P , where (f_t, X) is Bernoulli and $\alpha > 0$, then for a.e. $x \in X$, and T large enough, the algorithm below will produce a process $({}_T\bar{f}_t, {}_T\bar{X}, {}_T\bar{P})$ that will be α -congruent to the direct product of (f_t, X, P) and another flow $({}_Tg_t, Y)$ relative to the function ${}_T\bar{P}$ and \tilde{P} where $\tilde{P}(y, x) = P(x)$, $x \in X, y \in Y$. If P identifies X with a manifold M , then we can think of $({}_Tg_t, Y)$ as a viewer and if ${}_T\hat{P}$ is the image of ${}_T\bar{P}$ under the α -congruence, then we see ${}_T\hat{P}(y, x)$ when the viewer is in state y and we are looking at x in M .*

We now describe the algorithm that produces our simulation of f_t, X, P .

The input to the algorithm will be an initial segment of a

¹⁷The popular literature emphasizes the distinction between “deterministic chaos” and “real randomness.”

specific orbit of the original stationary process. The output will be a stationary process.

The initial segment of the orbit can be viewed as a function from the interval $[0, T]$ into a metric space. The algorithm takes any function $Q: [0, T] \rightarrow M$, (i.e., that maps the interval $[0, T]$ into a metric space M), and constructs a stationary process ${}_T\bar{f}_t, {}_T\bar{X}, {}_T\bar{P}$. (The result is that as $T \rightarrow \infty$ the new process is a good approximation of the original one.)

The algorithm uses a function $L(T)$ that tends to ∞ and satisfies

$$\frac{\log T}{L(T)} \rightarrow \infty \quad \text{as } T \rightarrow \infty.$$

To be specific one can take $L(T) = \log T / \log \log T$.

The algorithm deals with a continuous input, the flow $f_t(x)$ where t is real. Naturally, in any digital realization the time variable must be discretized. However, the input $\{P(f_t(x))\}_{t=0}^T$ ((f_t, X, P) is the process we are trying to simulate) might be in some analog form, as a physical tape, so that we can deal with such an input without worrying about discretizing time. In the same way we will speak of choosing a random point t_0 , uniformly distributed in an interval $[a, b]$ with the meaning being—either do so with some analog device or, if time is discretized, do so discretely. The theorem requires a minor modification of the following algorithm which is given later.

Algorithm A.

A1. Input: For $t \in [0, T]$ set $Q(t) = P(f_t(x))$ with $x \in X$ the starting point of the observations.

A2. For $N = 1, 2, 3, \dots$ choose a random point t_N , uniformly distributed on $[0, T - L(T)]$ independently of the previous choices, and set

$$G((N - 1)L(T) + s) = Q(t_N + s) \quad \text{for } 0 \leq s \leq L(T).$$

Comment. We are concatenating the block $\{Q(s)\}_{s \in [t_N, t_N + L(T)]}$ to the right of the orbit that we have already built up.

A3. Output: The path $G(s), 0 \leq s < \infty$.

This algorithm, for large enough T (and a.e. x), will yield a good approximation to the original process. Note that the algorithm does not require knowledge of the dynamics of f_t . The function $L(T)$ is universal¹⁸ and will work for all Bernoulli flows

¹⁸This is the main technical point beyond the isomorphism theory and requires a new Shannon–McMillan–Breiman type theorem [OW 4].

in spite of the lack of uniformity of the ergodic theorem. The only problem might come from a bad starting point, but we assume implicitly that the data with which we are presented reflect the physically significant invariant measure. Unfortunately, the algorithm as such fails to give α -congruence since the resulting process may not have sufficient entropy. We can correct for this by artificially adding some fixed entropy. To do this we add intervals of length one between the successive blocks of lengths $L(T)$ that the algorithm is concatenating, and label these intervals independently with a collection of labels of size $2^{L(T)}$. This will ensure that the resulting process, for T large enough, will have entropy greater than the entropy of the given process (f_t, P) .

Remark. The result in §2.4 (and the result above) implies that the Bernoulli flows can be modeled by a finite state machine (a computer) equipped with a roulette wheel. We may ask if there is an algorithm that allows us to produce a good simulation if we know the equations of motion.

The usual approach is the following:

We discretize time and phase space and approximate the vector field to determine how one of our discrete points in phase space will move to another discrete point in one unit of time. If we discretize finely enough and approximate the vector field well enough, then we can iterate this procedure and approximate a solution to our equations for a long period of time. However, if we keep iterating, our simulation will eventually be periodic (a computer is a finite state machine) and no periodic orbit can be close most of the time to a generic orbit of B_t or any mixing flow.

If, however, our computer is equipped with a random device that acts like a roulette wheel, then the situation can be remedied in the following seemingly paradoxical way: Iterate our equation L times and then jump in a random way to another point, flip a coin whether to decide to wait 0 or β units of time (β/L is irrational, $0 < \beta < 1$), iterate L times again, etc. If our original system is B_t , and the invariant measure has bounded density with respect to Lebesgue measure, then we can prove that if L is large enough and we have discretized finely enough and approximated well enough (depending only on L), then the computer simulation will be α -congruent to the direct product of the original flow and the Bernoulli flow. This means that the approximation is good for all time and even that the simulated process could have been

produced by watching the system through a random viewer that distorts by $< \alpha$ with probability $> 1 - \alpha$. (In addition, the viewer will not lose information in the sense that the orbit we are watching and the state of the viewer can be reconstructed from the path as seen through the viewer.)

Our random jumps could be taken to be uniform over the whole discretized phase space or uniform over a discretized ball of radius r around our position when we jump. (Our computer simulation can be thought of as a flow if we stay at each discrete point for our discrete unit of time.)

2.6. Instability, or when α -congruence fails.

I. All of our results about α -congruence that we have presented depend on Bernoulliness in an essential way and are for the most part false, in the non-Bernoulli case. Not only are the relevant processes not α -congruent; they even fail to satisfy part of the defining properties of α -congruence in a quantitative way. The reason for this is that any non-Bernoulli process is quantitatively different from all Bernoulli processes. To make this precise we introduce the notion of processes being α -disparate.

Definition of α -disparate. Processes (f_t, X, P) , $(\bar{f}_t, \bar{X}, \bar{P})$ are α -disparate if for almost every $x \in X$, and almost every $\bar{x} \in \bar{X}$, the set of t 's for which

$$d(P(f_t(x)), \bar{P}(\bar{f}_t, (\bar{x}))) > \alpha$$

has density $> \alpha$. (d measures the distance between points in the common range of P and \bar{P}).

The appropriate sets of x and \bar{x} can be described independently of this definition, as the generic points of the flows f_t, \bar{f}_t , respectively. These are the typical points for which time averages converge to the spatial averages for all continuous functions of the process.

For two fixed processes, the supremum of the α 's for which they are α -disparate defines a metric on processes, called the \bar{d} -metric. It plays an important role in the isomorphism theory of Bernoulli flows and will be discussed in detail in the next chapter. Here is the basic result:

Theorem 2.6.1. *If (f_t, X, P) is not Bernoulli and P is a generator, then there is a positive $\alpha > 0$, such that (f_t, X, P) and any Bernoulli $(\bar{f}_t, \bar{X}, \bar{P})$ are α -disparate.*

Let's consider the results of §2.2 in this light. There the intrinsic perturbations give rise to Bernoulli flows regardless of the nature of the original flow. (In B we must assume that the process that produces the interference is Bernoulli.) *Thus for a non-Bernoulli flow f_t , there is an $\alpha > 0$ such that all those intrinsic perturbations, no matter how small, are α -disparate from f_t .* In this way microscopic perturbations that are too small to see will have a cumulative macroscopic effect. This effect will be observable in a strong way as we shall see a little later on.

II. In Theorem 1 we see the extent of our being able to model by Bernoulli processes. Suppose, however, that we are interested in modeling with semi-Markov flows or semi-multi-step Markov flows (or in discrete time, Markov or multi-step Markov processes). If a semi-Markov flow or process has no periodic factor then it is Bernoulli and in general it is the direct product of a Bernoulli and a rotation. A system will be α -disparate from this larger class if it is not the direct product of a Bernoulli flow and the inverse limit of rotations (which can be thought of as a clock that gives seconds, minutes, hours, days, ... and so on with an infinite number of nested periods).¹⁹ This makes good sense both in discrete time and continuous time.

The class of products of Bernoulli and inverse limits of rotations is stable under factors and \bar{d} -limits. This follows from Thouvenot's relative isomorphism theory and Fieldsteel's extension to continuous time.

The next theorem is due to D. Rudolph and G. Schwarz [Rud-Sch] in discrete time and K. Park [Pa] in continuous time.

Theorem 2.6.2. *(f_t, X, P) , P generating, is the \bar{d} -limit of semi-multi-step Markov flows²⁰ if and only if f_t is the direct product of a Bernoulli flow and an inverse limit of rotations.*

Thus, if f_t is not the direct product of a Bernoulli flow and an inverse limit of rotations then, given a generating partition P , there is an $\alpha > 0$ such that (f_t, P) is α -disparate from any semi-multi-step Markov flow.

The output of a finite state machine equipped with a roulette wheel can be modeled by a semi-multi-step Markov flow (or, in

¹⁹A more formal description is given in §3.7.

²⁰In a semi- n -step Markov flow we remain at each of a finite number of states s_i , a fixed length of time t_i and then jump with a probability distribution that depends on s_i and the previous n states.

discrete time, a multi-step Markov process) and Theorem 2 tells us exactly which processes can be modeled by the outputs of such machines. *In particular the existence of a non-Bernoulli K automorphism means that there is a discrete time process with no deterministic part whose randomness cannot be explained in terms of roulette wheels.*

Since most of our results involve modeling or approximating by Bernoulli or semi-Markov processes, they fail in a quantitative way when we stray too far from the Bernoulli case.

III. α -disparate processes are quantitatively different and one might ask whether or not this difference can actually be observed. More generally, if we watch a sample output $P(f_t(x))$ of a process (f_t, X, P) and a sample output $Q(g_t(y))$ of (g_t, X, Q) , can we estimate their \bar{d} -distance?

If we are allowed to watch forever, then there is no real problem because the Birkhoff ergodic theorem tells us that, with probability one, the entire process (f_t, X, P) can be reconstructed from a sample infinite orbit $\{P(f_t(x))\}_{t=0}^\infty$ (we assume that P generates).

The following theorem captures our idea of observability in a better way.

Theorem 2.6.3. *There is a universal algorithm with the following properties:*

- (a) *The input is a pair of functions from the interval $(0, T)$ to a metric space, $\{P(f_t(x))\}_{t=0}^{t=T}$, and $\{Q(g_t(x))\}_{t=0}^{t=T}$.*
- (b) *The output is a number D_T (our estimate of the \bar{d} -distance). If $f_t(x)$ is Bernoulli or the direct product of Bernoulli and an inverse limit of rotations, then for almost every x, y chosen according to product measure on $X \times Y$, D_T converges as $T \rightarrow \infty$ to the \bar{d} -distance between (f_t, X, P) and (g_t, Y, Q) .*

Remark 1. The above algorithm is robust in the sense that if we change each input by $< \beta$ in the L_1 sense, then D_T change by $< 2\beta$.

Remark 2. If we make no assumption on (f_t, X, P) or (g_t, Y, Q) , then $\liminf D_t$ will be greater or equal to the \bar{d} -distance between (f_t, X, P) and (g_t, Y, Q) .

Remark 3. One can prove that there is no algorithm that will work for all (f_t, X, P) and (g_t, Y, Q) .

2.7. Further directions for α -congruence. When do we have statistical stability? Structural stability is limited to axiom A and few realistic systems are axiom A . We also ask when is there a canonical measure like the SBR measure, and when do we get stability results for the measure of maximal entropy?

Theorem 2 (see 1.2) shows that we can go beyond axiom A for certain perturbations. Another example due to Eloranta [E] is the following. Perturb the billiard system by taking two copies of a billiard table very close together and cutting a hole where each obstacle is, and connecting the opposite holes with a smooth surface (of negative curvature). The billiard system is then approximated by a geodesic flow and the approximating flow is α -congruent to the billiard system, after a small rescaling of time. ($\alpha \rightarrow 0$ as the tables get closer.) This example is interesting for two reasons. (1) The conclusion of structural stability cannot apply because a.e. orbit will spend a small amount of time circling the hole. (2) The perturbation, although small, gives a qualitatively different system. Thus billiards can be smoothed without changing the long-term dynamics.

If we restrict the kind of perturbation enough, we can get stability results for wide classes of systems. For example, if (f_t, M) is a smooth Bernoulli flow on a compact three-dimensional manifold M preserving a measure absolutely continuous with respect to Riemannian volume, and if we perturb f_t by changing the speed by a small C^2 function (which may vary from point to point along an orbit) then (f_t, M) will be α -congruent either to the perturbed flow \bar{f}_t or to \bar{f}_t times a rotation (this is the analog of Theorem 2.1.4). The question now is when can the alternative be ruled out?

We can also ask for systems other than axiom A that are stable under random perturbations. It seems that this is where the gap between what we can prove and what is true is the greatest. Most reasonable random perturbations are Bernoulli. If our original system is not Bernoulli we have no possibility of stability (see §2.6) so we assume that our original system is Bernoulli. Therefore, what we need to understand is how the invariant measure and the entropy change under the perturbation? For example, we do not know if there is a Bernoulli diffeomorphism of a two-dimensional manifold that is not stable under the perturbation of Theorem 2.2.1 when we jump with uniform distribution in a disk of small radius around the point we land at under the diffeomorphism.

We can also look for different kinds of random perturbations. One example is the following: Take an ergodic automorphism of T , of a compact Abelian group \mathcal{G} . Perturb the T process by taking any stationary process with values in \mathcal{G} , concentrated near 0, and adding this process to the T process. We have stability when the noise process is Bernoulli.

Any time we have a stationary process and another process that is supposed to approximate it, we can ask if the approximation is good for arbitrarily long time intervals. In this spirit, we can ask if the processes are close in the \bar{d} sense or the sense of α -congruence. (For example, Brownian motion confined to a finite interval with reflecting barriers is often approximated by a random walk.) Eloranta has results along these lines that extend the Donsker invariance principle [E].

We can also study α -congruence for other systems like infinite particle systems. Here the group is $\mathbf{Z}^n \times \mathbf{R}$. (\mathbf{Z}^n corresponds to translation of a n -dimensional lattice and \mathbf{R} to the time evolution.) In certain cases Steif [Ste] has shown that the system is α -congruent to a system where the lattice is made up of an infinite number of large disjoint cubes that do not interact with one another. Steif's cases are: (1) The $M < \epsilon$ condition [Lig] where there is only one invariant measure. (2) The two extreme measures for monotone systems—each lattice site wants to imitate its neighbors (see [Lig]). These systems include most of the systems studied. This result is interesting because infinite particle systems are supposed to model large finite systems, and this result says that they do in a very strong sense.

We can also ask for α -congruences that preserve some additional structure. One example is the $(n - 1)$ -frame flow on an n -dimensional manifold of negative curvature, n odd, $n \neq 7$. In this case, we have a group larger than R acting. (We can rotate the frame at a point.) Using the isomorphism theory for this more general group, one can show (using results of Brin and Gromov [BrG]) that if we perturb the metric, the new frame flow will be α -congruent (after a small rescaling of time) to the original one and the conjugating map will take the collection of frames at a point to the collection of frames at another point.

Understanding particular systems has been an ongoing project. Here is a recent example where work still needs to be done: Wojtkowski [Wo 1, Wo 2] has shown that the following system has

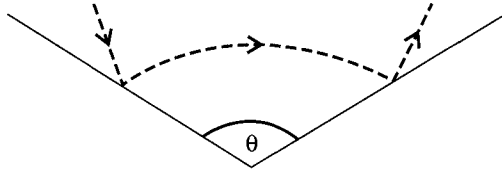


FIGURE 2

positive entropy: a ball of nonzero mass acted on by gravity and bouncing in an elastic way (no energy loss) where the floor is shaped like a wedge and $\pi/2 < \theta < \pi$ (Figure 2).

The Katok–Strelcyn [KatoStr] modification of Pesin theory almost certainly takes care of the singularities and if so, the system is either B_t or \overline{B}_t on each ergodic component. In either case the system would be α -congruent to a semi-Markov process as in §2.4. We could ask if a small change in θ produces an α -congruent system where we can make α arbitrarily small by making the change in θ sufficiently small. (When $\theta = \pi/2$ or $\theta = \pi$ the system has zero entropy and will therefore change discontinuously). The force of gravity could be taken as constant or inversely nonproportional to the distance to the center of the earth. Will the latter small perturbation produce an α -congruent system with small α ? Some of the interest in the above system lies in the fact that it is so close to a ball bouncing on a flat floor—the first system one studies in calculus.

3. A SURVEY OF SOME ABSTRACT ERGODIC THEORY²¹

3.1. **The \overline{d} -distance.** When a stochastic process²² x_n is observed through a noisy medium, the observed process can be represented

²¹The results in this section that are not specifically referenced can be found in [OW 1] or [O 1].

²²A *stochastic process* $\{x_n\}_{-\infty}^{\infty}$ is a collection of random variables defined on some probability space (Ω, Σ, μ) with values in some set A . The measure on $A^{\mathbb{Z}}$ defined by the image of μ under the mapping $\omega \mapsto \{x_n(\omega)\}_{-\infty}^{\infty}$ is called the distribution of the process or the joint distribution of the variables $\{x_n\}$. The process is called *stationary* if for each p , the distribution of $\{x_{n+p}\}_{n=-\infty}^{\infty}$ is the same as that of $\{x_n\}_{-\infty}^{\infty}$. If the process is stationary, then the shift T on $A^{\mathbb{Z}}$ ($x_n(\omega) = x_{n-1}(T\omega)$) preserves μ and we can therefore also model a stationary process by a measure-preserving transformation T acting on a space $X = A^{\mathbb{Z}}$ and a function P on X ($P(\omega) = x_0(\omega)$). Note that $\bigvee_{-\infty}^{\infty} T^i P$ generates the entire σ -algebra of X (up to sets of measure 0). We denote this setup by (T, X, P) , or if we want to emphasize the measure by (T, μ, P) , etc.

by

$$\bar{x}_n = x_n + e_n,$$

where the e_n represent the error process. If the medium is not very noisy, then e_n will be small most of the time and then for most time instants n , x_n will be close to \bar{x}_n . This is the archetypal example of two processes that we wish to think of as being close to one another. The \bar{d} -distance is a formal way of measuring the distance between processes that takes into account not merely the finite distributions but also the long-range behavior in the same way that the x_n and the \bar{x}_n processes above are close. Since discrete time stochastic processes are simpler than continuous time ones, we will begin by explaining the \bar{d} -distance in that case, and further suppose that the processes are finite valued.

When one is presented with two distinct stochastic processes $\{x_n\}$ and $\{y_n\}$, then it makes no sense to ask questions such as what is the probability that $x_0 = y_7$, since these random variables need not be defined on the same probability space. Any way of defining the two processes on the same space is called a *joining* of the two processes. More formally a joining is a stochastic process of pairs $\{\bar{x}_n, \bar{y}_n\}$ such that $\{\bar{x}_n\}$ has the same distribution, or probability laws as $\{x_n\}$ and $\{\bar{y}_n\}$ has the same distribution as $\{y_n\}$. There is always at least one joining of any two given processes—the *independent joining*—obtained by declaring that the x_n variables are all independent of the y_n variables. We shall be dealing with stationary processes and in that case by a joining we shall mean a stationary one, i.e., $\{(\bar{x}_n, \bar{y}_n)\}_{-\infty}^{\infty}$ is also a stationary process. There are cases where the independent joining is the only joining; however, in the situation that we will be studying there are typically many different joinings.

Given two processes $\{x_n\}$, $\{y_n\}$ we seek a joining between them $\{(\bar{x}_n, \bar{y}_n)\}$ so that

$$(*) \quad \Pr\{\bar{x}_0 \neq \bar{y}_0\}$$

shall be as small as possible. The infimum of (*) as we range over all possible joinings is the \bar{d} -distance between $\{x_n\}$ and $\{y_n\}$. Formally

$$\bar{d}(\{x_n\}, \{y_n\}) = \inf_{(\bar{x}_n, \bar{y}_n)} \Pr\{\bar{x}_0 \neq \bar{y}_0\},$$

where $\{(\bar{x}_n, \bar{y}_n)\}$ ranges over all joinings of the given processes. This metric is much more stringent than the metric defined by the finite distributions. Here is a simple example to keep in mind.

Consider the two state Markov chain defined by the transition matrix

$$\begin{pmatrix} \epsilon & 1 - \epsilon \\ 1 - \epsilon & \epsilon \end{pmatrix}$$

and stationary measure $(\frac{1}{2}, \frac{1}{2})$. When $\epsilon = 0$, this is a pure periodic process with precisely two output sequences. As long as $\epsilon > 0$, the \bar{d} -distance between this process and the pure periodic is $\frac{1}{2}$, corresponding to the independent joining (indeed there is no other one). To see the former, observe that by stationarity

$$\Pr\{\bar{x}_0 \neq \bar{y}_0\} = \frac{1}{N} \sum_1^N \Pr\{\bar{x}_i \neq \bar{y}_i\}.$$

Now for a typical \bar{x}_i output when ϵ is very small we will see long strings of alternations between the two states. However these will be out of phase about half the time with respect to either of the purely periodic alternations represented by $\{\bar{y}_i\}$. In finite distributions, of course, this family of Markov chains is continuous all the way down to $\epsilon = 0$. The family is continuous in the \bar{d} -distance for $\epsilon > 0$. While fairly elementary, this fact is nontrivial; see §3.5.

To compute the \bar{d} -distance between two processes in terms of the finite distributions, one proceeds as follows. Define the \bar{d} -distance between $\{x_n\}_{n=1}^N, \{y_n\}_{n=1}^N$ by

$$(**) \quad \bar{d}(\{x_n\}_{n=1}^N, \{y_n\}_{n=1}^N) = \inf_{\{(\bar{x}_n, \bar{y}_n)\}} \frac{1}{N} \sum_{n=1}^N \Pr\{\bar{x}_n \neq \bar{y}_n\},$$

where the infimum is over all joinings (now not required to be stationary) between the x_n 's and the y_n 's. It is a fact that

$$\bar{d}(\{x_n\}_{-\infty}^{\infty}, \{y_n\}_{-\infty}^{\infty}) = \lim_{N \rightarrow \infty} \bar{d}(\{x_n\}_1^N, \{y_n\}_1^N).$$

Our main interest in this metric is the behavior at zero, and so variant definitions which give that same behavior are sometimes more useful. For example, the right-hand side of $(**)$ can be replaced by

$$\inf \left\{ \epsilon : \text{for some joining } (\bar{x}_n, \bar{y}_n), \Pr \left\{ \frac{1}{N} |\{1 \leq n \leq N : \bar{x}_n \neq \bar{y}_n\}| \geq \epsilon \right\} \leq \epsilon \right\}.$$

With this variant it is easy to see what to do if the process is not finite valued but takes values in a fixed compact metric space

(M, d) . We take then the infimum of all $\epsilon > 0$ for which there is some joining such that

$$\Pr \left\{ \frac{1}{N} \sum_{n=1}^N d(\bar{x}_n, \bar{y}_n) \geq \epsilon \right\} < \epsilon.$$

After this discussion of the more standard situation we come to the one we need here. Now we have to begin with a measure-preserving flow f_t on a measure space (X, \mathcal{B}, μ) . For example f_t could be a smooth flow on a manifold X with \mathcal{B} the Borel sets and μ a fixed invariant measure. A process will be defined by such a flow and a mapping

$$P: X \rightarrow M$$

which is a measurable mapping from X to a compact metric space (M, d) . We will sometimes refer to P as a partition of X . For example, if X is a compact manifold, then the identity map with $M = X$ defines a process. To any point $x \in X$, and any $T > 0$, we associate a path in M defined by

$$P^T(x) = \{P(f_t(x)): 0 \leq t \leq T\}$$

and call that the $P-T$ name of x . This map takes the measure μ onto a probability measure on $M^{[0, T]}$, the space of $[0, T]$ -paths in M , which we denote by μ_T . These are the finite distributions of the process P . Had we been considering a finite space M , and a discrete time flow $\{f_n\}$, then these would just be our earlier finite distributions. For two measures μ_T, ν_T on $M^{[0, T]}$ define

$$\bar{d}(\mu_T, \nu_T) = \inf\{\beta\},$$

where β ranges over all positive numbers for which there is some measure λ on $M^{[0, T]} \times M^{[0, T]}$ satisfying

(1) $\pi_1 \circ \lambda = \mu_T, \pi_2 \circ \lambda = \nu_T$; where $\pi_{1,2}$ are the coordinate projections. This is a formal way of saying that λ is a joining of μ_T and ν_T .

$$(2) \lambda \left\{ (u, v): \frac{1}{T} \int_0^T d(u(t), v(t)) \geq \beta \right\} < \beta.$$

Here $u(t)$ and $v(t)$ denote a $[0, T]$ -path or an element in $M[0, T]$ i.e., $u(t) = P(f_t(x))$ for some x .

If ν_T represents the finite distribution of some process (g_t, ν, Q) with values in M then we define

$$\bar{d}((f_t, \mu, P), (g_t, \nu, Q)) = \lim_{T \rightarrow \infty} \bar{d}(\mu_T, \nu_T).$$

For the whole process this guarantees that for any β greater than this \bar{d} -distance there is a joining between them, a measure λ on $(M \times M)^{\mathbf{R}}$, so that

$$\lambda\{(u, v) : d(u(0), v(0)) \geq \beta\} < \beta.$$

The notion of the \bar{d} -distance is implicit in the concept of α -congruence. Indeed two processes that are α -congruent are within α in the \bar{d} -metric, but α -congruence involves much more since the (close) joining of the processes is required to be an *isomorphism* between them.²³

\bar{d} is not very useful for studying deterministic processes like the linear flows on the torus. Here almost any two different flows are far apart in \bar{d} . Indeed, defining

$$f_t(x, y) = (x + at, y + bt), \quad g_t(x, y) = (x + ct, y + dt)$$

for $(x, y) \in \mathbf{T}^2$, if (a, b) and (c, d) are not rationally related then one can show that the only joining possible between processes defined by f_t and g_t are independent joinings. It is precisely for chaotic systems, the ones that have positive entropy, and are very far from such deterministic systems, that the \bar{d} -metric proves to be a very useful tool.

3.2. Entropy and α -entropy. For a continuous mapping T of a compact metric space (M, d) the topological entropy is defined by

$$\lim_{\epsilon \rightarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \{ \text{minimal number of } (N, \epsilon) \text{ balls} \\ \text{that cover the space } M \},$$

where the (N, ϵ) ball means a set B that satisfies

$$d(T^n x, T^n y) < \epsilon, \quad \text{all } 0 \leq n < N, \quad \text{all } x, y \in B.$$

This quantity measures the exponential growth rate of the number of diverging orbits and is denoted by

$$h_{\text{top}}(T).$$

The topological entropy is zero for nice flows like the irrational flow on the torus, and its positivity is an indication of some kind

²³Rudolph has unpublished examples that show that it is possible for processes to be close in the \bar{d} sense and isomorphic without being close in the sense of α -congruence. In these examples there are joinings that are isomorphisms and joinings that produce a good \bar{d} -match but no joining that does both.

of chaotic behavior or sensitive dependence on initial data. In the presence of a measure μ that is T -invariant it is natural to count the minimum number of (N, ϵ) balls required to cover a set in X of large measure rather than all of X . For ergodic measures it turns out that fixing any $c > 0$ and calculating

$$\lim_{\epsilon \rightarrow 0} \limsup \frac{1}{N} \log \{ \text{minimal number of } (N, \epsilon) \text{ balls} \\ \text{needed to fill a set of } \mu\text{-measure } c \text{ in } X \}$$

gives the same number independently of c , and this number is called the *measure-theoretic entropy* of the system (M, T, μ) , denoted by

$$h_\mu(T).$$

For flows f_t , the measure-theoretic entropy was originally defined by means of the discrete time skeleta f_{nd} . It was J. Feldman who showed how to define the entropy of a flow, with respect to a finite partition, directly in terms of the flow [Fe 2]. His definition is close in spirit to the one we just gave and works just as well for compact partitions. For a measurable flow $(X, \mathcal{B}, \mu, f_t)$ and a partition $P: X \rightarrow M$, a compact metric space, define a $P - \alpha - T$ ball to be a set $B \subset X$ that satisfies

$$\frac{1}{T} \int_0^T d(P(f_t x), P(f_t y)) dt < \alpha \quad \text{all } x, y \in B.$$

Now define for $0 < c < 1$

$$h(P, \alpha, c) \\ = \limsup_{T \rightarrow \infty} \frac{1}{T} \log \{ \text{minimal number of } (P, \alpha, T) \text{ balls} \\ \text{needed to fill a set of } \mu \text{ measure } c \text{ in } X \}.$$

For ergodic processes one shows that this limit is independent of $c \in (0, 1)$ and we call that limit the α -entropy of the process, denoted by $h(P, \alpha)$. The entropy of the process (P, f_t) is defined by

$$h(P, f_t) = \lim_{\alpha \rightarrow 0} h(P, \alpha)$$

and finally the entropy of the flow is defined to be

$$h(f_t) = \sup_P h(P, f_t),$$

where the supremum ranges over all compact partitions, or over all finite partitions. This turns out to be the same as the more usual definition which is $h(f_1)$.

We have the formula first proved by Abramov, $h(f_{t_0}) = |t_0|h(f_t)$. Note that $h(f_t)$ refers to the entropy of the flow while $h(f_{t_0})$, where t_0 is thought of as a fixed real number, is the entropy of the transformation f_{t_0} .

It is trivial that under rescaling, $\bar{f}_t = f_{ct}$, we have

$$h(\bar{f}_t, P, \alpha) = |c|h(f_t, P, \alpha),$$

and thus also in the limit as $\alpha \rightarrow 0$

$$h(P, \bar{f}_t) = |c|h(P, f_t).$$

When P is changed to a \bar{P} that is close to P in the sense that for some small $\delta > 0$

$$(*) \quad \mu(x: d(\bar{P}(x), P(x)) < \delta) > 1 - \delta,$$

then for most of the space X , (P, α, T) balls are $(\bar{P}, (\alpha + 2\delta), T)$ balls so that for ergodic processes

$$h(\bar{P}, \alpha + 4\delta) \leq h(P, \alpha) \leq h(\bar{P}, \alpha - 4\delta).$$

In the limit, as $\alpha \rightarrow 0$ this continuity in P is lost and in general it is not true that as $P_n \rightarrow P$ in the metric defined by $(*)$ above,

$$h(f_t, P_n) \rightarrow h(f_t, P).$$

In the totally discrete situation, when the number of elements in the partition is finite and held fixed and the time discretized, the entropy depends continuously on the partition. This continuity allows for some results to be formulated more simply in that situation, and it is the lack of continuity in our setting that is responsible for the cumbersome definition of “finitely determined processes” below.

One can establish a version of the Shannon–McMillan theorem for the α -entropy and it plays a key role in the proofs of the theorems in the rest of this chapter. The remarkable thing is that a real meaning is given thereby not only to the process entropy $h(P, f_t)$ but to the α -entropy for each value of α . Here is one version of the theorem:

Given $\epsilon > 0$, if T is sufficiently large then there is a subset $X_0 \subset X$ satisfying

1. $\mu(X_0) > 1 - \epsilon$;
2. for any $P - \alpha - T$ ball $B \subset X_0$

$$\mu(B) \leq \exp\{-T(h(P, \alpha) - \epsilon)\};$$

3. *there is a covering of X_0 by fewer than*

$$\exp\{T(h(P, \alpha) + \epsilon)\}$$

(P, α, T) balls.

Another way of understanding (3) is to observe that, in the presence of (2), (3) implies that most of the (P, α, T) balls in that covering collection have measures which lie in the

$$\exp\{-T(h(P, \alpha) \pm \epsilon)\}.$$

In the discrete situation one can take $\alpha = 0$ and then there are *canonical* (P, N) balls which are simply the atoms of the partitions $\bigvee_{n=0}^{N-1} f_n^{-1}P$. Thus the classical Shannon–McMillan theorem simply asserts that most atoms of $\bigvee_{n=0}^{N-1} f_n^{-1}P$ have exponential size approximately equal to $\exp\{-Nh(f_1P)\}$. In general the formulation is of necessity more complicated.

The Shannon–McMillan theorem has a simpler statement for very small α . This statement is about centered (P, α, T) balls. We define a (P, α, T) ball centered around $\gamma(t) \in M^{[0, T]}$ to be the set of y such that

$$\frac{1}{T} \int_0^T d(P(f_t y), \gamma(t)) dt < \alpha.$$

If we use centered (P, α, T) balls, the value of α -entropy will change but the above Shannon–McMillan theorem will still hold and we will get the same limit $h(P, f_t)$ as $\alpha \rightarrow 0$. In terms of centered balls we get a Shannon–McMillan theorem for $h(P, f_t)$:

Given $\epsilon > 0$, if α is small enough and T large enough (given α), then there is a subset $X_0 \subset X$, $\mu(X_0) > 1 - \epsilon$ and any (P, α, T) ball with center in X_0 has measure between $\exp\{-T[h(P, f_t) \pm \epsilon]\}$ and there is a covering X_0 by fewer than $\exp\{T[h(P, f_t) + \epsilon]\}$ (P, α, T) balls.

The way in which α -entropy was defined makes it clear that the topological entropy of the flow gives an upper bound for the measure theoretic entropy (for this and the following result see the discussion in Szlenk’s book [Sz]). The fact that there always exists invariant measures which approach the topological entropy lies a little deeper and requires the construction of invariant measures.

With the concept of entropy at our disposal we can return to our earlier discussion of predictable and nonpredictable processes. The results are easier to state in discrete time (for continuous time flows

f_t , these results are best formulated by considering the discrete samplings of the flow $\{f_{t_0 n}\}_{n \in \mathbb{Z}}$. A finite state process $\{x_n\}_{-\infty}^{\infty}$ is said to be predictable if x_0 is a measurable function of the past $\{x_n : n \leq -1\}$. The fundamental connection with entropy is the following: *A finite state process $\{x_n\}_{-\infty}^{\infty}$ is predictable if and only if it has zero entropy.* A flow has zero entropy if and only if the discrete samplings of any measurable function are predictable.

Sometimes deterministic is used instead of predictable. Standard approximation theorems yield that if we are satisfied with predicting x_0 up to a small error then a *finite* number of observations in the past suffice. Examples of such systems include periodic systems such as the simple pendulum or quasi-periodic systems like billiards on a rectangular table with no obstacles. Geometric examples include the horocycle flow on a negatively curved manifold or the geodesic flow on an ellipsoid.

At the other extreme are the systems that are completely nonpredictable, where no nonconstant measurement on the system taken at discrete times (i.e., nt_0) is predictable. These are exactly the systems that have no factors of zero entropy. These systems are also called *K*-systems and for obvious reasons systems with *completely positive entropy*.

By a theorem due to Pinsker, Rohlin and Sinai [RoSi] these systems also have the following characterization: A process $\{x_n\}_{n \in \mathbb{Z}}$ is completely nonpredictable if and only if, for any k , and $\epsilon > 0$, there is an N so that the distribution of $\{x_1, x_2, \dots, x_k\}$ conditioned on $\{x_n : n \leq -N\}$ is within ϵ of the unconditioned distribution but for a set of measure $< \epsilon$ of specific outcomes of $\{x_n : n \leq -N\}$. Informally, any fixed number of observations $\{x_1, \dots, x_k\}$ is approximately independent of the remote past (and remote future). A flow is *K* if and only if any measurement taken at discrete times is completely nonpredictable.

Examples here are all the examples of Bernoulli flows that we have already discussed, but in addition there are non-Bernoulli examples, [O 7]. The first smooth example was given by Katok [Ka 1], and some rather natural smooth ones are discussed in Rudolph [Rud 7].

3.3. Extremal processes. We start by recalling the definition of Bernoulli shifts as those transformations that arise from independent processes. f_t is a Bernoulli flow if f_{t_0} is a Bernoulli shift for some t_0 . There are various characterizations of Bernoulli

processes. The last one to be discovered was the property of extremality due originally to J.-P. Thouvenot and first used extensively by D. Rudolph in his work on extensions of Bernoulli shifts. Since this characterization works for infinite entropy flows as well as finite entropy discrete time actions, we describe it first. To begin with the simplest situation, consider a discrete process $\{x_n\}$ with values in the finite set A . We say that the process is *extremal* if the measure μ_n that it defines on n -names (i.e., A^n) cannot be divided up into a small number of pieces that are far away from μ_n in the \bar{d} -metric. In formal terms we require that for any $\epsilon > 0$, there is some $\delta > 0$ and N_ϵ , and for all $n \geq N_\epsilon$ if μ_n , the measure²⁴ on A^n defined by $\{x_i\}$ is decomposed as a convex combination

$$\mu_n = \sum_{i=1}^{2^{\delta n}} c_i \lambda_i$$

of probability measures λ_i , then most (measured by c_i) of the λ_i satisfy

$$(*) \quad \bar{d}(\mu_n, \lambda_i) < \epsilon.$$

In more precise terms

$$\sum_{i \in I} c_i > 1 - \epsilon$$

where I consists of all those indices that satisfy $(*)$.

The basic fact is that independent processes are extremal in the sense just described. It is easiest to prove this fact using the notion of finitely determined (see §3.5 below). For flows and compact partitions one can use the very same definition for extremality:

Definition. A process (f_t, P) is extremal if given $\epsilon > 0$ there is some $\delta > 0$ and T_0 , and for all $T > T_0$ if μ_T denotes the measure on $P - T$ names and

$$\mu_T = \sum_{i=1}^{2^{\delta T}} c_i \mu_i$$

is a convex decomposition of μ_T into probability measures μ_i , then

$$\sum_{i \in I} c_i > 1 - \epsilon$$

²⁴Here is how the stochastic process $\{x_i\}$ defines the measure

$$\mu_n : \mu_n([u_1 u_2 \dots u_n]) = \Pr\{x_1 = u_1, x_2 = u_2, \dots, x_n = u_n\}.$$

where $i \in I$ if

$$\bar{d}(\mu_i, \mu_T) < \epsilon.$$

Recall that a process P is called a *generator* for the flow $(X, \mathcal{B}, \mu, f_t)$ if the σ -algebra generated by P , $\bigvee_{-\infty}^{\infty} f_t P$ coincides with \mathcal{B} modulo μ -null sets.

Theorem 3.3.1. *If one generator of a flow is extremal so are all processes defined by that flow, and the flow is Bernoulli (isomorphic to B_{ct} for some constant $c > 0$, or to B_t^∞).*

If in our definition of extremal we restricted the convex combinations of measures μ_i to be those that come from partitioning X , we would get an equivalent definition.

Extremality in the latter form is especially easy to interpret in terms of predictability. Suppose f_t is a Bernoulli flow on a metric space M , and that we would like to predict, to within ϵ , the T -name ($M - T$ -name) of a point (in M) that we can see with only finite precision. The finite precision device, through which we view our starting point, partitions M into a finite number of pieces and extremality says that if there are fewer than $2^{\delta T}$ pieces, then with probability $1 - \epsilon$ we can say nothing about T -names if we ignore an average error less than ϵ . Thus exponentially (in T) good precision tells us nothing about the T future if the exponential rate is too slow. This means that if it takes δ^{-1} units of time to get one extra bit of information about the starting point and if δ^{-1} is too large, then no prediction is possible.

Note that the entropy gives the slowest rate for which it is possible to predict arbitrarily well.

3.4. Very weak Bernoulli (VWB). For an independent process the distribution of $\{x_1, \dots, x_n\}$ conditioned on *any* past event A of the form

$$A = \{\omega: x_i(\omega) = \xi_i, i \leq 0\}$$

is the *same* as the unconditioned distribution. You come to the definition of a *very weak Bernoulli process* by relaxing a little bit on this both in the set of A 's and on the relationship of the unconditioned to the conditioned distribution. We are considering here finite-valued discrete times processes P and use the notation μ_N^P for the measure on N -names defined by the distribution of

$$S^{-1}P \vee S^{-2}P \vee \dots \vee S^{-N}P.$$

We have switched to the notation of a measure space (X, \mathcal{B}, μ) , a measure-preserving transformation $S: X \rightarrow X$ and a finite valued partition P of X . The process (S, P) is said to be very weak Bernoulli (VWB) if, given $\epsilon > 0$, there exists N_0 and for all $N \geq N_0$ and all L the set of all atoms A of $\bigvee_{-L}^0 S^{-l}P$ for which

$$\bar{d}(\mu_N^P|A, \mu_N^{\mathcal{B}}) < \epsilon$$

has total measure greater than $1 - \epsilon$. Here we are using $\mu_N^P|A$ to denote the measure on $P - N$ -names obtained by conditioning on the atom A .

The following basic result from the isomorphism theory of Bernoulli shifts provides a way that specific flows can be shown to be Bernoulli since this property is often verifiable [OW 3].

Theorem 3.4.1. (a) *The process (S, P) , P finite, is very weak Bernoulli (VWB) if and only if*

$$\left(X, \bigvee_{-\infty}^{\infty} S^{-n}P, \mu, S \right)$$

is a Bernoulli transformation.

(b) *If for some t_0 , and finite partitions P_i, P_{i+1} refines P_i , and together they generate under f_{t_0} (i.e., $\bigcup_i (\bigvee_{-\infty}^{\infty} f_{t_0}^{-n}P_i)$ is the whole σ -algebra) and (f_{t_0}, P_i) is VWB then f_{t_0} is isomorphic to $B_{c,t}$ for some $c > 0$, or B^∞ . Conversely if f_{t_0} is isomorphic to $B_{c,t}$, $c > 0$, or B^∞ then for any fixed finite partition P , and any $t_0 \neq 0$, (f_{t_0}, P) is VWB.*

Here is a variant of this property where both L and μ are given infinite values. For C in a set of atoms of $\bigvee_{-\infty}^0 T^{-n}P$ that has full measure

$$(*) \quad \bar{d} \left(\bigvee_1^{\infty} T^{-n}P|C, \bigvee_{-1}^{\infty} T^{-n}P \right) = 0.$$

Note that since the first distribution is *not* stationary, this doesn't imply that the processes are the same. The \bar{d} -distance in (*) should be interpreted as: for some joining between the conditional variables $\{x_n^C\}_1^\infty$ and $\{x_n\}_1^\infty$ we have

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_1^N \Pr\{x_n^C \neq x_n\} = 0.$$

This property may be compared with one of the characterizations of K processes that we gave in §3.2. Instead of saying there that $\{x_1, \dots, x_n\}$ becomes independent of the remote past $\{x_n: n \leq -N\}$, we could equally well say that a fixed time range $\{x_n: N \leq n \leq N+k\}$ in the distant future becomes independent of the past $\{x_n: n \leq -1\}$. On the other hand, for VWB processes the past gives vanishingly little information (on average) of the long-term future.

In terms of how knowledge of the past allows us to predict the future we get the following table (observations with finite precision at discrete time intervals).

Class of flow	Knowledge of past for a generating observation tells us
0 entropy	Everything about the future
Positive entropy	Not everything about the future
K	Arbitrarily little about fixed time in distant future
Bernoulli	Vanishingly little about long-term future

3.5. Finitely determined processes (FD). One of the key concepts that emerged from the proof of the isomorphism theorem for Bernoulli processes is that of processes whose long-range behavior is fixed or determined in the \bar{d} -metric by their finite distribution and entropy. Thus, after fixing the short-range behavior (i.e., the distribution of some finite number of variables), the long-range behavior (i.e., the joint distribution of all the variables of the process) is essentially described by a single real parameter—the entropy. The concept makes sense only for finite entropy processes and flows and we restrict to that case throughout this section.

We begin with a definition for finite-valued processes or partitions in discrete time.

Definition. A process (S, P) is *finitely determined* if given $\epsilon > 0$ there is a $\delta > 0$ and an N such that whenever (\bar{S}, \bar{P}) satisfies: *close in finite distribution*

$$d\{\text{distribution } (\bigvee_0^N S^i P), \text{ distribution } (\bigvee_0^N \bar{S}^{-i} \bar{P})\} < \delta$$

and close in entropy

$$|h(\bar{S}, \bar{P}) - h(S, P)| < \delta$$

then

$$\bar{d}((S, P), (\bar{S}, \bar{P})) < \epsilon.$$

Since for fixed N the space of distributions on $\sqrt[0]{N}S^{-i}P$ is finite dimensional the specific metric d that is used is not important, one can, for example, use the l^1 -norm between the probability vectors, or the \bar{d} -metric. The basic observation is that independent processes are finitely determined. That is fairly easy to see using the fact that the only way for the entropy of a process to be close to its theoretical maximum (the entropy of the partition P itself) is for P to be nearly independent of $\sqrt[-1]{N}S^{-i}P$. After this basic observation is made, one can go on to show that the FD property is preserved under factor maps and this gives some insight into how the concept is very useful.

When working with flows or compact partitions it is best to use the α -entropy even though this definition of finitely determined becomes more cumbersome. This was given by Feldman.

Definition. A process (f_t, P) is *finitely determined* if given $\epsilon > 0$ there is an $\gamma > 0$ and functions $T(\beta), \delta(\beta)$ (that depend both on γ and the process) such that if (\bar{f}_t, \bar{P}) is any process that satisfies for some β

(i) $h(\bar{f}_t, \bar{P}, \beta) \geq h(f_t, P) - \gamma,$

(ii) $\bar{d}(\text{distribution } \sqrt[0]{T(\beta)}\bar{f}_t\bar{P}, \text{distribution } \sqrt[0]{T(\beta)}f_tP) < \delta(\beta),$

then

$$\bar{d}((\bar{f}_t, \bar{P}), (f_t, P)) < \epsilon.$$

The extra β is necessary since the closeness in distribution required in (ii) depends upon the level β at which the test process has enough entropy. For $T(\beta)$ chosen well (i) and (ii) together will imply

$$|h(\bar{f}_t, \bar{P}, \beta) - h(f_t, P)| < \gamma$$

since one can easily give upper bounds for the entropy in terms of finite distributions.

The above form of the definition is the one that is used in proving the main isomorphism theorems (the original proof involved a messy discretization of time). However, in applications it is convenient to have another version of the definition in which ordinary entropy rather than α -entropy appears.

Lemma 3.5.1. *If (f_t, P) is finitely determined and P is a finite partition, then given $\epsilon > 0$, there is an $\bar{\gamma} > 0$, $\bar{\delta} > 0$, and T such that if (\bar{f}_t, \bar{P}) is ergodic and satisfies*

- (a) $\frac{1}{|t_0|} h(\bar{f}_{t_0}, \bar{P}) \geq h(f_t, P) - \bar{\gamma}$ for some $t_0 \in (\frac{1}{2}, 2)$,
- (b) $\bar{d}(\text{dist } \vee_0^T \bar{f}_t \bar{\mathcal{P}}, \text{dist } \vee_0^T f_t P) < \bar{\delta}$,

then

$$(*) \quad \bar{d}((\bar{f}_t, \bar{P}), (f_t, P)) < \epsilon.$$

Note that $h(\bar{f}_{t_0}, \bar{P})$ refers to the entropy of the \mathbf{Z} -action generated by \bar{f}_{t_0} , while $h(f_t, P)$ refers to the entropy of the \mathbf{R} -action, f_t , $t \in \mathbf{R}$.

Here is the main theorem concerning FD and Bernoulli processes.

Theorem 3.5.1. *(B_t, P) is FD for all P . If (f_t, P) is FD for a generator P , then f_t is isomorphic to B_{ct} for some constant c .*

If S is Bernoulli (S, P) is FD for all P . If (S, P) is FD for a generator P then S is Bernoulli.

In particular if a process is not Bernoulli then we can find finite approximations to it that remain a fixed distance away in \bar{d} . A zero entropy process (S, P) cannot be finitely determined since the n -step Markov approximation (with a tiny bit of pure randomness thrown in) (S_n, P_n) are certainly close in finite distribution and in entropy but there is no nontrivial (nonindependent) matching between (S, P) and (S_n, \mathcal{P}_n) so that the \bar{d} -distance between them is bounded away from zero. Let us take a closer look at rotations of the circle. Any two rotations, R_α, R_β , are fixed distance apart in \bar{d} no matter what partition we take. Thus in any neighborhood of R_α in the topology of finite distributions (entropy is not an issue since they all have zero entropy) there are an uncountable number of processes, all of whose mutual distance in \bar{d} are greater than some fixed $\delta > 0$. This is another example of how our results are limited to chaotic systems.

An immediate consequence of the FD property for Bernoulli flows is the following result which has no analogue in discrete time:

Proposition 3.5.1. *If f_t is a Bernoulli flow and P is a fixed partition, then the mapping*

$$c \mapsto (f_{ct}, P)$$

is continuous in the \bar{d} -metric.

For non-Bernoulli flows this kind of continuity can fail dramatically. For the simplest examples consider the Kronecker flows on a torus of any dimension (including 1).

Remark. A result that is much stronger than Proposition 1 was given in §2.3.

3.6. Isomorphism theorems. We begin by stating again the basic definitions. Two measure-preserving flows²⁵ $(X, \mathcal{B}, \mu, f_t), (\bar{X}, \bar{\mathcal{B}}, \bar{\mu}, \bar{f}_t)$ are *isomorphic* if there is a measurable invertible mapping $\theta: X \rightarrow \bar{X}$ such that $\mu(\theta^{-1}(\bar{B})) = \bar{\mu}(\bar{B})$ for all $\bar{B} \in \bar{\mathcal{B}}$ and

$$\bar{f}_t \circ \theta = \theta \circ f_t$$

for all t .

A partition $P: X \rightarrow M$ is said to be a *generator* if the smallest sub- σ -algebra of \mathcal{B} with respect to which all $f_t \circ P$'s are measurable is all of \mathcal{B} (modulo μ -null sets). In general this sub- σ -algebra is denoted by $P_{-\infty}^{+\infty}$, thus P is a generator means that $\mathcal{B} = P_{-\infty}^{+\infty}$. While the existence of *finite* generators for transformations of finite entropy was a nontrivial achievement (first accomplished by W. Krieger [Kri]) it is easy to find compact generators. Indeed if X itself is compact, as is often the case, then $P = \text{identity}$ is trivially a generator.

Abstract processes $(f_t, P), (\bar{f}_t, \bar{P})$ are isomorphic if the flows $(X, P_{-\infty}^{+\infty}, f_t), (\bar{X}, \bar{P}_{-\infty}^{+\infty}, \bar{f}_t)$ are isomorphic.

Finally, let us repeat again the definitions of γ -congruence. First of all two flows f_t, \bar{f}_t defined on the same space M , are γ -congruent if they are isomorphic via a mapping $\theta: M \rightarrow M$ that satisfies

$$\mu\{x: d(x, \theta(x)) > \gamma\} < \gamma,$$

where d is a fixed metric on M . Thus γ -congruence strengthens the notion of isomorphism by requiring that the correspondence giving the isomorphism be close to the identity in a strong geometric sense.

To compare flows on different spaces from this point of view we need to fix partitions on these spaces with values in the same space (M, d) . So flows

$$(f_t, X, P), \quad (\bar{f}_t, \bar{X}, \bar{P})$$

²⁵We will denote a flow simply by f_t or by (X, f_t) , or by (X, \mathcal{B}, f_t) .

with both P and \bar{P} taking values in (M, d) are γ -congruent relative to P and \bar{P} if there is an isomorphism

$$\theta: X \rightarrow \bar{X}$$

between (f_t, X) and (\bar{f}_t, \bar{X}) such that

$$\mu\{x: d(P(x), \bar{P}(\theta(x))) > \gamma\} < \gamma.$$

Note that this notion depends on more than the abstract process (f_t, P) since we require that θ be an isomorphism even if P is not a generator.

The first definition is a special case of the second when we take $\mathcal{P} = \bar{\mathcal{P}} = \text{identity}$. Notice that if (f_t, X, P) and $(\bar{f}_t, \bar{X}, \bar{P})$ are γ -congruent for small γ , then the processes (f_t, X, P) and $(\bar{f}_t, \bar{X}, \bar{P})$ are close in \bar{d} since the map θ defines a joining between (f_t, P) and (\bar{f}_t, \bar{P}) . This joining is simply the μ measure lifted up to the graph $\{(x, \theta(x)): x \in X\}$ in $X \times X$. The fact that this measure projects onto $\bar{\mu}$ in the second coordinate is the fact that $\theta \circ \mu = \bar{\mu}$.

When processes $(f_t, P_{-\infty}^{+\infty}, P)$ and $(f_t, \bar{P}_{-\infty}^{+\infty}, \bar{P})$ are 0-congruent we will simply call them congruent. In the language of probability theory, congruent processes are identically distributed processes. We make the following convention: When we write (f_t, P) and (\bar{f}_t, \bar{P}) are congruent or identically distributed, we will always refer to the process $(f_t, P_{-\infty}^{+\infty}, P)$. γ -congruence, however, refers to the entire σ -algebra, not just the one generated by P . We will also write $d(P, \bar{P}) < \gamma$ as an abbreviation for $\mu\{x: d(P(x), \bar{P}(x)) > \gamma\} < \gamma$.

Recall that f_t is defined to be a Bernoulli flow (of finite or infinite entropy) if and only if f_{t_0} is a Bernoulli shift for some fixed t_0 . A Bernoulli shift is defined as follows. Let Y be a measure space of total measure one. Let $Y_i, -\infty < i < \infty$, be copies of Y . Let $X = \prod_{-\infty}^{\infty} Y_i$ with product measure (points in X are sequences with values in Y). The Bernoulli shift acts by shifting each of the above sequences.

Here is the weak isomorphism (or imbedding) theorem for Bernoulli flows:

Theorem 3.6.1. *Let (X, f_t) be a Bernoulli flow and P any partition defined on X .*

(a) **Imbedding.** *If $h(\bar{f}_t) \geq h(f_t, P)$ then there is a partition \bar{P} on \bar{X} such that the processes (\bar{f}_t, \bar{P}) and (f_t, P) are identically distributed.*

(b) **Stability of imbedding.** *If $h(\bar{f}_t) \geq h(f_t, P)$ and there is a partition \hat{P} on \bar{X} satisfying*

$$\bar{d}((\bar{f}_t, \hat{P}), (f_t, P)) < \alpha$$

then for any $\epsilon > 0$ there is a partition \bar{P} such that

$$d(\bar{P}, \hat{P}) < \alpha + \epsilon$$

and (\bar{f}_t, \bar{P}) and (f_t, P) are identically distributed.

The result in (b) of course is a strengthening of (a), which follows from (b) by beginning with any partition. It is also the more basic result since it gives a hint as to how the theorem is proved. What it shows is that any approximate imbedding, one which is not too distant from the goal, can be changed slightly so as to become an exact imbedding. In all of the subsequent generalizations the same pattern persists—first one establishes an embedding theorem of the type that we just gave and then when both flows are Bernoulli and the entropies are equal one obtains isomorphisms. In the next theorem, part (a) is the basic isomorphism theorem for Bernoulli flows of equal entropy and it appeared first for finite entropy in [O 6]. Part (c) is a refinement which is often useful and says that isomorphisms are prevalent. Parts (b) and (d) deal with γ -congruences and are new; we shall give their proofs in §5.

Theorem 3.6.2. (a) **Isomorphism theorem for flows.** *If (X, f_t) and (\bar{X}, \bar{f}_t) are Bernoulli flows of equal entropy, then they are isomorphic.*

(b) *If (X, f_t) and (\bar{X}, \bar{f}_t) are Bernoulli flows of equal entropy and P, \bar{P} are partitions that satisfy*

$$\bar{d}((f_t, P), (\bar{f}_t, \bar{P})) < \frac{\gamma}{10},$$

then (X, f_t, P) is γ -congruent to $(\bar{X}, \bar{f}_t, \bar{P})$.

(c) **Prevalence of isomorphisms.** *If (X, f_t) is a Bernoulli flow and R is a full entropy partition (i.e., $h(f_t, R) = h(f_t)$), then for any $\delta > 0$ there is some \tilde{R} satisfying*

(i) \tilde{R} is a generator,

(ii) $d(R, \tilde{R}) < \delta$,

(iii) *the processes (f_t, \tilde{R}) and (f_t, R) are identically distributed.*

(d) *If (X, f_t, P) is a Bernoulli flow and $\gamma, \eta > 0$ are given, then there is a $\delta > 0$ such that any Bernoulli flow $(\bar{X}, \bar{f}_t, \bar{P})$ that satisfies*

$$(i) \bar{d}((f_t, P), (\bar{f}_t, \bar{P})) < \gamma/20,$$

- (ii) $|h(f_t) - h(\bar{f}_t)| < \delta$
 is γ -congruent to (X, f_{ct}, P) for some constant c that satisfies $|1 - c| < \eta$. (6(d) differs from (b) only in the case of finite entropy.)

3.7. Factors, \bar{d} -limits, and the relativized theory. The main structural theorem concerning Bernoulli processes is:

Theorem 3.7.1. (a) *Any factor of a Bernoulli flow is a Bernoulli flow, where by a factor we mean the restriction of the flow to an invariant sub- σ -algebra.*

- (b) *If \mathcal{B}_n are σ -algebras, $\mathcal{B}_i \subset \mathcal{B}_j$ for $j > i$ and $\bigcup \mathcal{B}_n = \mathcal{B}_\infty$ and if f_t restricted to each \mathcal{B}_n is a Bernoulli flow, then f_t restricted to \mathcal{B}_∞ is a Bernoulli flow.*
- (c) *If $(f_t^{(n)}, P^{(n)})$ are Bernoulli processes that converge in \bar{d} to (f_t, P) , then (f_t, P) is also a Bernoulli process. (i.e., f_t acting on the σ -algebra generated by P is the Bernoulli flow).*

Especially useful is part (b) which reduces proving that a given flow is Bernoulli to obtaining the result for special partitions.

The next most random processes after the independent processes are the Markov chains. Already for discrete time and finite state space a Markov chain is not necessarily Bernoulli because it may be periodic. If it is not purely periodic then from the point of view of the general theory it has a simple structure: it is the direct product of a finite rotation and a Bernoulli process. The \bar{d} -limits of such processes are fairly easy to describe; they are the direct products of a Bernoulli process and an inverse limit of finite rotations (see [RudSch]).

The best way to think about an inverse limit of rotations is as follows. When a process has a periodic factor with period p_1 , there is a definite shift invariant way to divide almost every realization of the process into consecutive blocks of size p_1 . If in addition there is a periodic factor with period $p_1 p_2$, then one can group these p_1 -blocks into p_2 -groups in a definite way. Continuing in this fashion, a process has such an inverse limit of finite rotations as a factor if almost every realization of the process can be given a hierarchical nested block structure into p_1 -blocks, $p_1 p_2$ -blocks, etc. If these blocks completely determine the realization, then the process itself is an inverse limit of rotations.

It is a general consequence of the disjointness theory of H. Furstenberg [Fu] that a zero entropy factor (in particular an inverse limit of rotations) and a Bernoulli factor are independent. What is not always the case for a given process is that one can find a Bernoulli factor that generates together with the zero entropy factor. One of the consequences of J.-P. Thouvenot's relative theory is a concrete criterion for exactly when this happens. This criterion is what is linking the results mentioned above. Because of its importance we give the general formulation. We first give the formulation for discrete time.

Definition of relative Bernoulli. A process (f, P) is said to be relatively Bernoulli with respect to a factor \mathcal{A} , if there is a Bernoulli factor \mathcal{B} that is independent of \mathcal{A} and together with \mathcal{A} gives the entire σ -algebra of the process.

When \mathcal{A} is generated by a partition $Q(\mathcal{A} = \bigvee_{-\infty}^{\infty} f^n Q)$ then we look for a partition R such that

- (i) all $f^i R$ are independent,
- (ii) the σ -algebras $\bigvee_{-\infty}^{\infty} f^i R$ and $\bigvee_{-\infty}^{\infty} f^i Q$ are independent,
- (iii) $Q \vee R$ is a generator, i.e.,

$$\bigvee_{-\infty}^{\infty} f^i (Q \vee R) = \bigvee_{-\infty}^{\infty} f^i P.$$

For the notions of relatively VWB and relatively FD one modifies the usual definitions by conditioning on the σ -algebra \mathcal{A} . One can define these notions directly in terms of \mathcal{A} by means of the Rohlin representation of X as a measurable bundle of measure spaces over $X|\mathcal{A}$, the space of the factor. A more finite definition is possible if one fixes some generator Q of \mathcal{A} . We give the latter first. It is convenient to use the notion of a *relative \bar{d} -metric*. Given two spaces X, \bar{X} and partitions Q, \bar{Q} so that (f, Q) is congruent to (\bar{f}, \bar{Q}) , we define for partitions P of X, \bar{P} of \bar{X}

$$\bar{d}_{Q, \bar{Q}}((f, P), (\bar{f}, \bar{P}))$$

as the supremum of $\Pr\{P \neq \bar{P}\}$ taken over all joinings of $(f, P \vee Q), (\bar{f}, \bar{P} \vee \bar{Q})$ that identify (f, Q) with (\bar{f}, \bar{Q}) . More formally there should be a space with partitions (P', \bar{P}', Q') such that:

- (a) $(f', P' \vee Q') \sim (f, P \vee Q),$
- (b) $(f', \bar{P}' \vee Q') \sim (\bar{f}, \bar{P} \vee \bar{Q}),$

$\bar{d}_{Q, \bar{Q}}((f, P)(\bar{f}, \bar{P})) = \sup \Pr(\bar{P}' \neq P')$ where the supremum is taken over all $(f', P' \vee \bar{P}' \vee Q')$ satisfying (a) and (b).

As usual, these things may be calculated by looking for such joinings of $\bigvee_0^n f^i(P \vee Q)$, $\bigvee_0^n \bar{f}^i(\bar{P} \vee \bar{Q})$ for each n , and this gives the finite definition of the relative \bar{d} -metric.

Definition. A process (f, P) is FD relative to the factor $\mathcal{A} = \bigvee_{-\infty}^\infty f^i Q$ if given $\epsilon > 0$ there is a $\delta > 0$ and an n , such that if $(\bar{f}, \bar{P} \vee \bar{Q})$ is any ergodic process satisfying:

- (1) $(\bar{f}, \bar{Q}) \sim (f, Q)$,
- (2) $|h(f, P \vee Q) - h(\bar{f}, \bar{P} \vee \bar{Q})| < \delta$,
- (3) $\bar{d}_{Q, \bar{Q}}((\bigvee_{i=0}^n f^i(P \vee Q), \bigvee_{i=0}^n \bar{f}^i(\bar{P} \vee \bar{Q}))) < \delta$,

then

$$\bar{d}_{Q, \bar{Q}}((f, P), (\bar{f}, \bar{P})) < \epsilon.$$

This notion was first introduced by J.-P. Thouvenot [Tho 1] who proved that these are exactly the processes that are relatively Bernoulli. This work was extended by A. Fieldsteel in [Fi 1] to flows with a definition that is exactly analogous. He also proved the analogue of the statement that if f_{t_0} is Bernoulli, then f_t , as a flow, is Bernoulli, and showed that if P is Bernoulli relative to a factor for one fixed time f_{t_0} , then it is relatively Bernoulli as a flow.

To check that specific examples are relatively Bernoulli it is convenient to have the analogue of relatively VWB. In practice one usually does this when conditioning on the entire fibers of the factor (as opposed to a finite condition).

Definition. We say that (f, P) is VWB relative to \mathcal{A} if for any $\epsilon > 0$ there is some N_0 , and for ϵ -a.e. atom A of \mathcal{A} and for any L there is a set \mathcal{E} of atoms of $\bigvee_{-L}^0 f^n P$ of relative measure in A greater than $1 - \epsilon$ and then for all $N \geq N_0$

$$\bar{d}(\bigvee_0^N f^{-n} P|A \cap E, \bigvee_0^N f^{-n} P|A) < \epsilon, \quad E \in \mathcal{E}.$$

Here too variants are possible where one supposes \mathcal{A} generated by Q and looks only at the conditioning on a finite portion of $\bigvee_{-\infty}^\infty f^{-n} Q$.

Although this notion is not found explicitly in Thouvenot's work, it is straightforward to see that it too is equivalent to relatively FD and to being relatively Bernoulli. For flows, one uses

this to check that f_{t_0} is relatively Bernoulli and then applies Fieldsteel’s result as we have described above.

A further result, based on these was obtained in [ShiTho], where it is shown that the class of processes that are the direct products of zero entropy and Bernoulli is closed in the \bar{d} -metric. These results form the background to what we have said about \bar{d} -limits of Markov chains and \bar{d} -limits of semi-Markov processes (see Theorem 2.6.2 and [Pa, RudSch]). Finally we mention the striking result that any factor of a process that is Bernoulli \times zero entropy is again of that form. This shows that the class of $B \times$ zero entropy processes has the same structural properties that the class of Bernoulli processes themselves possess. We will give here one result that we use to prove the results of §2.

Theorem 3.7.2. *Suppose that (X, \mathcal{B}, f_t) is Bernoulli or the direct product of a Bernoulli flow and a flow of zero entropy. Let P be a partition such that (f_t, P) is Bernoulli. Given ϵ , there is a \hat{P} satisfying:*

- (i) $d(P, \hat{P}) < \epsilon$.
- (ii) (f_t, P) and (f_t, \hat{P}) are identically distributed.
- (iii) there is an f_t -invariant σ -algebra \mathcal{A} , independent of $\bigvee_{-\infty}^{\infty} f_t \hat{P}$ such that

$$\mathcal{A} \vee (\bigvee_{-\infty}^{\infty} f_t \hat{P}) = \mathcal{B}.$$

In case (f_t, P) has full entropy then \mathcal{A} can be taken to be the zero entropy factor alone.

4. A SURVEY OF SOME SMOOTH CHAOTIC SYSTEMS

4.1. Introduction. In this section we shall review the material from the theory of dynamical systems that forms the subject matter of our study. The material is quite standard and can be found in various places. See, for example, [Shu, Ma 1, Sm]. The basic problem can be quickly stated: Given a system of ordinary differential equations

$$(*) \quad \frac{dx_i}{dt} = F(x_1, \dots, x_d), \quad 1 \leq i \leq d$$

analyze and describe the qualitative behavior of solution curves $x_i(t)$, especially as $t \rightarrow \infty$. A more general version deals with a manifold M^d and a vector field X on the manifold and studies

the integral of the vector field, or what amounts to the same, the flow

$$f_t: M \rightarrow M$$

which gives, as $\{f_t x\}$ the integral curve that passes through a given point x . Even in the rare instances that explicit solutions of (*) can be found, their asymptotic behavior still presents a problem, while for the most part, of course, one cannot even hope for explicit solutions. The problem originates in Newtonian mechanics with the differential equations (*) describing the motion in the dynamical system phase space.

Perhaps the simplest nontrivial case is the 2-body problem which in suitable coordinates (taking the origin at the center of gravity) gives rise to the system

$$\frac{dq_i}{dt} = p_i, \quad \frac{dp_i}{dt} = -\frac{cq_i}{\|q\|^2} \quad i = 1, 2, 3.$$

Here there are basically two types of motion:

- (1) the two bodies eventually separate and grow further and further apart as time goes on;
- (2) there is a periodic motion so that one returns again and again to the starting position *exactly*.

The first represents the prototype of wandering where there is no recurrence at all, while the second represents the simplest type of recurrence, namely periodicity. As soon as one admits a third body to the problem, the situation changes dramatically. A new type of behavior occurs—recurrence without periodicity and recurrence in a highly nonregular fashion. A similar thing happens when one passes from the simple pendulum that exhibits a qualitatively simple periodic behavior to a *compound pendulum* obtained by adding a bar at the end that is free to rotate in the plane of the motion. Here once again one observes experimentally highly nonregular behavior.

One of the early attempts to organize the study of such irregular systems was via the notion of structural stability. This important idea was introduced in [AndP] motivated mainly by physical considerations. These lead one to look for systems whose essential qualitative behavior doesn't change under small perturbations of the underlying differential equations. The perturbations are supposed to reflect our imperfect knowledge of the precise equations

that govern a system. A flow $\{f_t\}$ is said to be *structurally stable* if for any $\{\hat{f}_t\}$ which is nearby in the C^1 topology (on the vector fields defining the flow) there is a homeomorphism close to the identity of the phase space that carries orbits of \hat{f}_t to orbits of f_t . Somewhat paradoxically, the phenomenon of structural stability is intimately connected with instability of individual orbits, in the sense that small changes in the initial position can lead to exponential divergence for large time values. Almost periodic flows are not structurally stable since the rotation number can be changed by very small changes in the vector field. This can be seen very clearly already for the linear flows on the two-dimensional torus. There is a good historical survey of structural stability in [Ano 2].

While there has been little success so far in rigorously explaining phenomena, such as the compound pendulum, a great deal of progress has been made in the analysis of model problems that exhibit some kind of *hyperbolicity* which seems to be the source of the highly irregular behavior of solution curves in examples like the compound pendulum. The archetypal model example is the geodesic flow on a manifold, which represents the motion of a free particle constrained to move on the surface. Here one sees highly irregular behavior when the surface has negative curvature and one has a good understanding of the asymptotic behavior through the work of many mathematicians such as Hadamard, Morse, Hedlund, Hopf, Sinai, and Anosov. ([Ano 1] contains a good account of this work.) Following a quick summary of this work we shall describe the successively more encompassing generalizations with weaker hyperbolicity assumptions: Anosov flows, axiom A flows, partially hyperbolic systems (such as billiards with convex obstacles).

A parallel story can be told concerning discrete time systems in which the iterates of a single transformation replaces the continuous time flow. In this brief survey we shall alternate between the two cases usually choosing to focus on that case for which the story is more easily told.

4.2. Geodesic flows. Fix some compact Riemann manifold M^d with a Riemannian metric and let $X = T_1M^d$ denote the unit tangent bundle, i.e., the bundle consisting of pairs (u, ξ) , where $u \in M^d$ and ξ is an element of the tangent space at u that has length one (already the metric is being used). The geodesic flow

g_t acts on X and maps (u, ξ) to the point $(\bar{u}, \bar{\xi})$ as follows:

Form the geodesic starting at u in the direction ξ , follow it to a distance t (if $t < 0$ go backwards), the point that is reached is \bar{u} while $\bar{\xi}$ is the unit tangent to that geodesic at \bar{u} (in the direction away from u , of course).

The geodesic flow preserves the natural Riemannian volume element of X which is taken to be the measure in any measure-theoretic discussion of the geodesic flow. When the manifold has positive curvature, the sphere for example, then the motion tends to be periodic or quasi-periodic. We shall concern ourselves with the case of *negative curvature*. There the first phenomenon that one encounters is the *instability* of the motion in the sense that small changes in the starting point typically lead to large deviations as time goes by. One can see this explicitly very clearly if M is a two-dimensional manifold with constant negative curvature because then the covering space is simply the hyperbolic space and typically straight lines, even if they start out very close to one another, will diverge apart.

There is, however, a compensating feature, namely for any fixed point (u_0, ξ_0) if u is near u_0 then there are special directions ξ for which $g_t(u_0, \xi_0)$ and $g_t(u, \xi)$ stay close to one another, and in fact for some fixed constant c the points

$$g_t(u_0, \xi_0), \quad g_{t+c}(u, \xi)$$

get close to one another exponentially fast. In the classic case of constant negative curvature these points lie on a *horocycle* through (u_0, ξ_0) . Let's define in general the stable and unstable sets for x in X :

$$W^s(x) = \{y: \lim_{t \rightarrow +\infty} d(g_t(y), g_t(x)) = 0\}$$

$$W^u(x) = \{y: \lim_{t \rightarrow -\infty} d(g_t(y), g_t(x)) = 0\}.$$

For the geodesic flow it turned out that these sets have the structure of submanifolds that intersect transversally at x (and are also transversal to the flow direction), and that furthermore their tangent spaces at x together with the vector field defining the flow span the tangent space at x . These foliations play a major role in proving the structural stability of the flows. The stable and unstable manifolds have an important geometric property—namely, it turns out that they are *dense* in the manifold. From the point of

view of the measure-theoretic properties of the geodesic flow the key property is that they form *absolutely continuous foliations*. To motivate this definition we first sketch Hopf's argument for the ergodicity of the geodesic flows on surfaces of variable negative curvature.

For any fixed continuous function ϕ , Birkhoff's ergodic theorem implies that

$$(*) \quad \lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t \phi(g_s x) ds = \bar{\phi}(x),$$

where $\bar{\phi}$ is the projection of ϕ onto the set of invariant functions. To prove ergodicity is equivalent to showing that for every continuous function ϕ this projection $\bar{\phi}$ is a constant a.e. Since ϕ is continuous if $y \in W^s(x)$ then since $g_s x$ and $g_s y$ get closer and closer also $d(\phi(g_s x), \phi(g_s y)) \rightarrow 0$ and then $\bar{\phi}(x) = \bar{\phi}(y)$. But the invariant functions for g_s and g_{-s} are the same; thus if $\hat{\phi}$ represents the limit in (*) when $t \rightarrow -\infty$ we have $\bar{\phi} = \hat{\phi}$ almost everywhere. We are thus in the following situation. We have two functions $\bar{\phi}$, $\hat{\phi}$ such that

- (a) for a.e. x , $\bar{\phi}(x) = \hat{\phi}(x)$,
- (b) $\bar{\phi}$, $\hat{\phi}$ are both constant along orbits,
- (c) $\bar{\phi}$ is constant on stable manifolds,
- (d) $\hat{\phi}$ is constant on unstable manifolds,

and would like to conclude that $\bar{\phi}$ is constant. Without further information we cannot draw such a conclusion. In spite of the fact that the three parts—the flow direction and stable and unstable manifolds—locally span the whole space since we only know (a) for a.e. x we cannot conclude that ϕ is locally a constant.

Here is a schematic diagram of what may happen (we have eliminated the flow direction by projecting onto a plane—in reality the stable and unstable manifolds never form such a two-dimensional surface). The vertical curves represent the unstable foliation and the horizontal lines the stable foliation. A function ϕ taking two values ± 1 is defined in such a way that it is constant on horizontal lines and is $+1$ on the upper half of the strip and -1 on the lower half. $\hat{\phi}$ (Figure 3, see p. 74) will be constant on vertical curves and will almost equal $\bar{\phi}$.

Iterating this kind of picture again and again allows one to replace most by almost everywhere (a.e.). Essentially, absolute continuity eliminates this pathology requiring that the mapping between different leaves of the unstable foliation, given by mapping

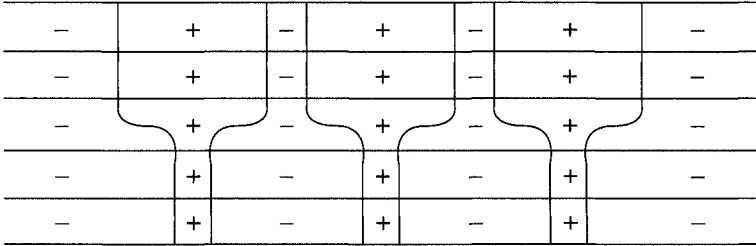


FIGURE 3

along the stable manifolds is absolutely continuous with respect to the Lebesgue measure on the leaves. This allows one to conclude that $\bar{\phi}$ and $\hat{\phi}$ are locally constant functions. A further argument of a topological nature is required to finish the proof. Here by other methods one shows that the flow had dense orbits—or what is the same, is topologically transitive. This latter means that for any nonempty open sets U, V there is some t_0 such that $g_{t_0}U \cap V \neq \emptyset$.

The above considerations apply to a wider class of flows and actually prove more than ergodicity. There is an alternative, depending on whether or not the stable and unstable vector fields are jointly integrable. If yes, then there is a rotation factor for the flow and the following discussion is valid for the Poincaré map to any such leaf. If not, then one sees that measurably any set foliated by both stable and unstable leaves has measure zero or one. Then one proves that the flow is a K -flow as follows (for geodesic flows such as we have been considering it is known that the first alternative cannot arise²⁶): For this one notices that for any smooth partition P of the partition $\bigwedge_n \bigvee_n^\infty f_{kt_0}P$ is essentially W^u . The Pinsker–Rohlin–Sinai theorem [Ro, Si] implies that measure-theoretically the following two σ -algebras are the same

$$\bigcap_n \bigvee_{n=-\infty}^{-n} g_{kt_0}P = \bigcap_n \bigvee_n^\infty g_{kt_0}P \quad \text{up to sets of measure zero.}$$

Thus a measurable function which is constant along leaves of W^u is almost surely constant along leaves of W^s and as we have just seen such functions must be constant. We use the P-R-S theorem once again to conclude that our flow is K .

This argument, and extensions of it, represent the contributions

²⁶For more details, see the discussion in §4.4 below.

made by Sinai and Anosov to the geodesic flows in the 1960s after the theory of entropy was developed. With the further advances made in the theory of Bernoulli processes we were able to show in fact the geodesic flow is isomorphic to the Bernoulli flow [OW 3]. We shall sketch later on how one carries this out for a discrete time analogue, namely a hyperbolic toral automorphism.

4.3. Anosov flows. An *Anosov flow* is one that possesses a global hyperbolic structure similar to the one present in the geodesic flow. Here is the standard definition:

A flow $f_t: M^d \rightarrow M^d$ is called *Anosov* if there is an f_t -invariant decomposition of the tangent bundle TM into a Whitney sum

$$TM = E^s \oplus E^u \oplus E,$$

where E is the direction of the flow and E^s, E^u satisfy

$$(*) \quad \begin{aligned} \|Df_t(u)\| &\leq Ce^{-\lambda t} \|u\| \quad u \in E^s, \quad t \geq 0, \\ \|Df_{-t}(v)\| &\leq Ce^{-\lambda t} \|v\| \quad v \in E^u, \quad t \geq 0 \end{aligned}$$

for constants $C, \lambda > 0$ with $\| \cdot \|$ representing the norm with respect to some Riemannian metric. As usual, Df represents the differential of the map f . It is technically somewhat easier to analyze an *Anosov diffeomorphism* which is a diffeomorphism $f: M \rightarrow M$ such that TM decomposes into $TM = E^s \oplus E^u$ and the analogue of (*) holds (replacing t by an integer and understanding f_n to be the n th iterate of f). Note that the time one map of an Anosov flow is not an Anosov diffeomorphism because of the presence of the flow direction. One reduces Anosov flows to Anosov-like maps by considering the first return map to a codimension one cross-section transversal to the flow. The simplest example of an Anosov diffeomorphism is a hyperbolic toral automorphism, i.e., an algebraic automorphism of T^d none of whose eigenvalues has modulus one. We shall use these to illustrate the results about flows.

The first important result is the integration of E^s, E^u to obtain stable and unstable manifolds.

There are foliations $\mathcal{E}^s, \mathcal{E}^u$ of the manifold M which are invariant under f_t and whose leaves are tangent to E^s, E^u , respectively. If the flow is $C^k, k \geq 2$ then each leaf is a submanifold with smoothness C . As C. Pugh has recently shown [Pu] one cannot obtain such a result with merely C^1 , although $C^{1+\alpha}$ for some $\alpha > 0$ does suffice. Furthermore there is a metric on the manifold

such that if x, y belong to the same leaf of \mathcal{E}^s

$$d(f_t x, f_t y) \leq e^{-\lambda t} \quad \text{for all } t \geq 0$$

and some fixed $\lambda > 0$ and similarly for \mathcal{E}^u with f_t replaced by f_{-t} . A similar result holds for Anosov diffeomorphisms. When f is a toral automorphism these foliations are simply the affine subspaces parallel to the two basic linear spaces E^s, E^u which group together the eigenvectors corresponding to modulus strictly less than one and strictly greater than one, respectively.

A straightforward consequence of this is the fundamental shadowing of pseudo-orbits. We'll describe this for diffeomorphisms.

Definition. A sequence $\{x_n\}_{n=-\infty}^{\infty}$ is an ϵ -pseudo-orbit for f (with respect to a fixed metric d) if

$$d(fx_n, x_{n+1}) < \epsilon \text{ all } n.$$

Theorem 4.3.1. *If f is an Anosov diffeomorphism there is some $\epsilon_0 > 0$ such that if $\{x_n\}_{n=-\infty}^{\infty}$ is an ϵ -pseudo-orbit with $\epsilon < \epsilon_0$ there is a unique point \bar{x} such that*

$$d(f^n \bar{x}, x_n) < c(\epsilon) \text{ all } n,$$

where f^n is defined by $f^0 = \text{identity}$, $f^{n\pm 1} = f^{\pm 1} \circ f^n$ and $c(\epsilon)$ tends to zero with ϵ .

It might be helpful to compare this purely topological theorem when the pseudo orbit is produced by a perturbation, with shadowing that comes from α -congruence. Here all perturbed orbits have a shadow orbit that is close all of the time. α -congruence produces a shadow orbit for almost every perturbed orbit that is close most of the time. The shadowing orbits produced by α -congruence, however, are generic; i.e., they are typical in the sense that their statistics reflect the statistics of diffeomorphisms or flows (see §2.6) while the orbits in the above theorems need not be generic (in fact they will all be periodic if the pseudo-orbits arise from simulating by a finite state machine, i.e., a computer).

In the case of continuous time there is an additional difference. In the topological case, the shadowing orbit may have to be reparametrized in a way that varies from point to point along the orbit while in the measure-theoretic case we only have to reparametrize by a constant rescaling of time (which is the same for all orbits).

Here is a sketch of a proof of this shadowing theorem due to C. Conley [C]. One considers squares fibered by the stable and unstable manifolds, say take a piece of the unstable manifold through x , \mathcal{E}_x^u , and then through every y in that piece consider a small piece of \mathcal{E}_y^s . When f is applied to such a rectangle, one direction is stretched and the other is contracted, so that one gets a picture like

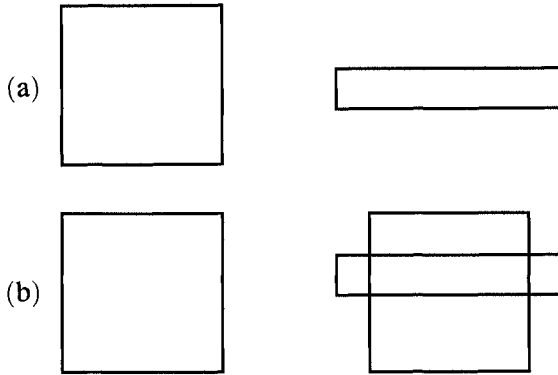


FIGURE 4

The point is that such a rectangle moves completely through an ordinary shaped rectangle centered at $f(x)$.

The ϵ -pseudo-orbit defines an infinite sequence of rectangles, each of which is mapped to the next as just described, and then one obtains a continuum joining the top and bottom of the first rectangle that maps into all of the rectangles. This proves a one-sided version of the shadowing theorem. Then reversing time one gets a continuum going across the other direction and their intersection point is the desired \bar{x} .

Now for structural stability one can proceed as follows. First we get a result of P. Walters [Wa] about topological stability in the sense that if g is C^0 close to f then there is a continuous map h such that

$$f \circ h = h \circ g.$$

For this notice that a true g -orbit is an ϵ -pseudo-orbit for f if g is sufficiently close to f . Define $h(x)$ to be the unique point whose true orbit follows closely upon the g -orbit of x . Then one works a little to see that h is continuous and onto. If g is C^1 close to f then one first proves that g is also Anosov and that

the $C(\epsilon)$ can be chosen uniformly in a neighborhood of f , then reversing the argument one sees that each is a factor of the other and finally that they are actually topologically conjugate.

The ergodic theory of these diffeomorphisms or flows is understood best when f or f_t preserves a smooth measure. Anosov and Sinai showed [AnoSi] that they were K -automorphisms and later on, with the development of the Bernoulli theory, they were shown to be Bernoulli [Az, Bun, Rat]. These results depend upon the measure-theoretic properties of the foliations $\mathcal{E}^s, \mathcal{E}^u$ that we discussed in the previous section (absolute continuity). When f or f_t doesn't preserve a smooth measure there are several natural measures that present themselves for consideration. There is the measure of maximal entropy (which Bowen showed was unique, in general) and the measures defined by the natural geometric measures associated with the foliations $\mathcal{E}^s, \mathcal{E}^u$. These latter are the SBR measures which we discuss below.

For diffeomorphisms these issues are best studied via the symbolic dynamics for the Anosov diffeomorphisms. This beautiful subject that goes back to the work of Hadamard and Morse became a very powerful tool for the analysis of hyperbolic systems. The key idea is that of a Markov partition, first used for classifying automorphism of the 2-torus in [AdW 1, AdW 2] and then extended by Ya. G. Sinai [Si 5] and R. Bowen [B 1]. A discussion of these ideas would take us too far afield and so we return to the dynamics and take up the next generalization: axiom A flows.

The main ideas and results of this section are due to Smale [Sm].

4.4. Axiom A flows. If $\{f_t\}$ is a flow on a manifold M the nonwandering set Ω is defined by

$$\Omega = \left\{ x \in M : \begin{array}{l} \text{for every neighborhood } U \text{ of } x, \text{ there are} \\ \text{arbitrarily large positive } t\text{'s such that} \\ U \cap f_t(U) \neq \emptyset \end{array} \right\}.$$

Basically axiom A flows are those with hyperbolic structure on Ω . The point is that only on Ω are there any interesting long-range dynamics and so that is where the hyperbolic structure is needed to recover results like those obtained for Anosov flows. Define in general a hyperbolic structure on an f_t -invariant set Λ to be a continuous f_t -invariant splitting of the tangent bundle $T_\Lambda(M) = E^s + E^u + E$ where E is the one-dimensional tangent bundle to

the flow and there are constants $C, \lambda > 0$ so that

- (i) $\|Df_t(v)\| \leq Ce^{-\lambda t}\|v\|$ for $v \in E^s, t > 0,$
- (ii) $\|Df_t(u)\| \leq Ce^{-\lambda t}\|u\|$ for $u \in E^u, t < 0.$

Definition. The flow f_t is an axiom A flow if $\Omega = \Lambda \cup F$ where

1. F is the set of fixed points of the flow each of which is hyperbolic, (in particular F is finite).
2. Λ is the closure of the nontrivial periodic orbits of f_t and carries a hyperbolic structure.
3. Λ and F are disjoint.

The reason for the separation between F and the rest of Ω is that the only way a point x in F can be hyperbolic is for the sum

$$\dim E_x^s + \dim E_x^u$$

to equal the dimension of the manifold, since there is no flow direction at a fixed point, and this can't mesh continuously with E^s, E^u that satisfy

$$\dim E^s + \dim E^u = \dim M - 1 .$$

Anosov flows are axiom A flows. The nonwandering set is the whole manifold and there are no fixed points so the one thing to check is that periodic orbits are dense and indeed a variant of the argument that shows that ϵ -pseudo-orbits are tracked by orbits does just that.

For the axiom A case one establishes the existence of stable and unstable foliations integrating E^s and E^u except that now these manifolds extend *beyond* Ω , and so the theory becomes much more technical. The main result here can be formulated as follows:

Theorem 4.4.1. *If f_t is an axiom A flow then for each point $x \in \Omega$ we have injectively immersed manifolds $W^u(x), W^s(x)$ tangent to E_x^u, E_x^s , respectively. They are characterized by the property*

$$W^s(x) = \{y: \text{for some } c, \lambda > 0, d(f_t x, f_t y) \leq ce^{-\lambda t}, t \geq 0\}$$

$$W^u(x) = \{y: \text{for some } c, \lambda > 0, d(f_{-t} x, f_{-t} y) \leq ce^{-\lambda t}, t \geq 0\}.$$

Two leaves $W^u(x), W^u(y)$ either coincide or are disjoint. The families $\mathscr{W}^u = \{W^u(x): x \in \Omega\}, \mathscr{W}^s = \{W^s(x): x \in \Omega\}$ are

invariant under f_t , the leaves are as smooth as is the vector field that defines the flow and vary continuously with x .

For this formulation compare [PuShu]. We reiterate the point that these W^u , W^s 's extend in general beyond Ω which might in fact have the structure of a Cantor set crossed with \mathbf{R} .

Next using this result one goes on to establish that Ω decomposes into a finite number of components $\Omega = \Omega_0 \cup \Omega_1 \cup \dots \cup \Omega_m$, on each of which f_t is *topologically transitive*. This means that most orbits (in the sense of: with the exception of a set of the first category) are dense. In terms of these components one can formulate the *no cycle condition*. A cycle is a finite chain

$$\Omega_{i_0}, \Omega_{i_1}, \dots, \Omega_{i_k} = \Omega_{i_0},$$

where for each $0 \leq j < k$ the i_j 's are distinct and we have that there is a point $x \in M$ whose α -limit set lies in Ω_{i_j} and whose ω -limit set lies in $\Omega_{i_{j+1}}$. Recall that the α -limit set of x is the set of limit points of $f_t x$ as $t \rightarrow -\infty$ and the ω -limit set of x is the set of points of $f_t x$ as $t \rightarrow +\infty$. The orbit through x provides a kind of transition between Ω_{i_j} and $\Omega_{i_{j+1}}$. The flow satisfies the *no cycle condition* if there are no cycles with $k \geq 2$.

The form that the structural stability theorem takes here is the following:

Theorem 4.4.2. *An axiom A flow f_t with the no cycle condition is Ω -stable in the sense that for any flow \bar{f}_t sufficiently close to it in the C^1 topology there is a homeomorphism of Ω into $\bar{\Omega}$, the nonwandering set for \bar{f}_t , that maps directed trajectories for f_t to directed trajectories for \bar{f}_t .*

It follows in particular that the perturbed flow \bar{f}_t is also axiom A. In the decomposition above, the sets Ω_i are called *basic hyperbolic sets*. They have the property that there is an open set $U_i \supset \Omega_i$ such that $\Omega_i = \bigcap_{t \in \mathbf{R}} f_t U_i$. A special case of these basic sets is formed by those where for some open neighborhood U_i and t_0 , for all $t \geq t_0$, $f_t(U_i) \subset U_i$. It then follows that $\Omega_i = \bigcap_{t \geq 0} f_t(U_i)$, and such systems are called *axiom A attractors*. One can prove that for a C^2 axiom A flow these *basins of attraction* with those of point attractors, cover the manifold up to a set of Lebesgue measure zero. For these attractors a special measure is singled out by the following theorem. Denote by W the basin

of the attractor Ω which is the union of the stable manifolds of points in Ω . Thus W consists of all the points x that are asymptotic to some point $y \in \Omega$ in the sense that $d(f_t x, f_t y) \rightarrow 0$ as $t \rightarrow +\infty$. Let m denote the volume measure on the manifold.

Theorem 4.4.3. *There is a probability measure μ , invariant under f_t , such that for m -a.e. $x \in W$, the basin of attraction of Ω , we have ($m \equiv$ Lebesgue measure)*

$$(*) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \phi(f_t x) dt = \int_{\Omega} \phi d\mu$$

for all continuous functions ϕ .

The main point here is that for most points in W the orbit behaves in a statistically regular way. The attractor, which typically has measure zero and cannot be seen precisely, is what governs this regularity through the mediation of the invariant measure μ ; i.e., the dynamics on the attractor govern the long-term behavior of all points near the attractor. (The orbit of a.e. point in W is forward asymptotic to the orbit of a generic point on the attractor. Generic means that the time average of any continuous function converges along the orbit to its integral.)²⁷ Conversely the points of the large set W —which always contains an open set—can be used to get good information about the hidden attractor.

For the nature of the flow f_t on the attractor with this special measure (which we shall refer to as the Sinai–Bowen–Ruelle measure, cf. [Si 6, BRue]) we first have to discuss the basic dichotomy for flows and it’s best to begin with Anosov flows (a special case of an axiom A attractor). A simple example will illustrate the possibilities. Take an Anosov diffeomorphism f of a compact manifold M . Consider the constant time suspension flow of f defined as follows. First take $M \times [0, 1]$ and identify points $(x, 1)$ with $(f(x), 0)$. Then flow along the vertical fibers, i.e., $f_t(x, r) = (x, r + t)$. Where $r + t$ is greater than 1, one reduces using the identification just made. This flow is never mixing because one has eigenfunctions. Indeed if one collapses M to a point then our flow is just the periodic flow around the circle. Nonetheless the flow is seen to be an Anosov flow. Thus Anosov flows need not be mixing and *a fortiori* need not be K or Bernoulli. However, in this example, after one accounts for

²⁷Thus the long-term behavior of certain dissipative systems is governed by a finite invariant measure.

the periodic behavior, the resulting transformation is Anosov and one has a well-understood ergodic behavior. In general, this is essentially the only way that an Anosov flow can fail to have good mixing properties. To formulate the next theorem, due to J. Plante [Pl] (an earlier version is due to Anosov who uses a smooth invariant measure), recall that strongly stable and strongly unstable foliations \mathcal{E}^{ss} , \mathcal{E}^{uu} are the ones that integrate the strictly stable and strictly unstable directions.

Dichotomy Theorem. *If f_t is an Anosov flow of a compact manifold M such that there are no wandering points then either*

- (I) *Each strongly stable and strongly unstable manifold is dense in M , or*
- (II) *f_t is the suspension (constant time) of an Anosov diffeomorphism of a compact C^1 submanifold of codimension one in M .*

In [B 1], R. Bowen proves an analogous dichotomy theorem for axiom A flows. The resulting structure in the case that leaves are not dense is not quite as smooth as in the above and one gets on the cross section not an axiom A diffeomorphism but a homeomorphism that behaves qualitatively like an axiom A diffeomorphism (i.e., the global, integrated versions of the properties hold). For case (I), the symbolic dynamics of the SBR measure enable one to establish that the flows are K and even Bernoulli (cf. Bunimovich [Bun] and M. Ratner [Rat]). In case (II) the flow is measure-theoretically a Bernoulli times a pure rotation.

The cases (I) and (II) also correspond to the weakest kind of mixing—namely topological mixing. That is to say that case (I) holds if and only if the flow is topologically mixing. It then turns out that this implies the strongest kind of measure-theoretic mixing.

4.5. Partially hyperbolic systems and billiards. The next class of flows for which a good analogue of much of the theory above can be still established consists of those flows for which one has some hyperbolicity but it is not uniform and it expresses itself in the long-range behavior. Here one finds such simple physical systems as billiards with convex scatterers that can model the hard sphere gas, geodesic flows on manifolds with some positive curvature, and so on. The most basic idea here is that of Lyapunov exponents and the theorem of Oseledec and we proceed to explain that next, beginning with the case of diffeomorphisms.

An important feature of a diffeomorphism $f: M \rightarrow M$ is that one can linearize in an infinitesimally small neighborhood of any point. Doing this to f^n and applying the chain rule we get

$$(*) \quad Df^n(x_0) = (Df)(x_{n-1}) \cdot Df(x_{n-2}) \cdots Df(x_0),$$

where $x_i = f^i(x_0)$ and $Df(x)$ is of course a linear mapping from $T_x M$ to $T_{f(x)} M$. This expression is a multiplicative analogue of a sum $g(x_0) + g(x_1) + \cdots + g(x_{n-1})$ which forms the matter of the classical ergodic theorems and the *multiplicative ergodic theorem* of Oseledec is a general kind of limit theorem for expressions such as $(*)$. The general setup is as follows. A finite measure-preserving transformation $T: X \rightarrow X$ is given and a vector space V_x of fixed dimension is attached measurably to each $x \in X$. Finally one is given a linear transformation $A(x)$ at each $x \in X$ mapping V_x to V_{Tx} . By fixing bases in a measurable way one can think of the $A(x)$'s as being matrices. Define

$$A_n(x) = A(T^{n-1}x)A(T^{n-2}x) \cdots A(Tx)A(x).$$

To simplify the formulation we assume that T is ergodic, in the general case the various constants appearing below become general functions that are invariant under T .

Theorem 4.5.1. [Os, Rag] *If (X, \mathcal{B}, μ, T) is an ergodic measure-preserving transformation and $A(x)$ a $d \times d$ matrix-valued measurable function such that $\log(\|A(x)\|)$ is integrable then there are constants $\lambda_1 < \lambda_2 < \cdots < \lambda_l$ and subspaces $0 = V_0 \subset V_1(x) \subset V_2(x) \subset \cdots \subset V_{l-1}(x) \subset V_l = \mathbf{R}^d$ that depend measurably on x such that for a.e. x :*

- (i) $A(x)V_i(x) = V_i(Tx), \quad 1 \leq i \leq e,$
- (ii) for $v \in V_i(x) \setminus V_{i-1}(x)$

$$(*) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \|A(T^{n-1}x)A(T^{n-2}x) \cdots A(x)v\| = \lambda_i.$$

The λ_i are called the *Lyapunov exponents*, they describe the possible exponential growth rates of vectors under $A_n(x)$. Their *multiplicity* is $\dim V_i - \dim V_{i-1}$. In case $A^{-1}(x)$ also satisfies the integrability assumption and T is invertible then one can replace the lattice of subspaces V_i by eigenspaces $\Lambda_i(x)$ such that

$$\mathbf{R}^d = \Lambda_1(x) \oplus \Lambda_2(x) \oplus \cdots \oplus \Lambda_e(x),$$

$$A(x)\Lambda_i(x) = \Lambda_i(Tx) \text{ and for } v \in \Lambda_i(x) \setminus \{0\}, (*) \text{ holds.}$$

The basic application of the theorem is to a diffeomorphism f on a manifold M . The vector space associated to a point x is the tangent space to M at x , $T_x M$, and the linear map $A(x)$ is the differential $f'(x)$. For each invariant measure the multiplicative ergodic theorem applies and gives a set of full measure for which exponents exist. Asymptotic hyperbolicity expresses itself in the condition that all Lyapunov exponents differ from zero. In that case Ya. Pesin [Pe] showed that much of the theory of axiom A systems carries over. He treats the case when a smooth measure, i.e., absolutely continuous with respect to the volume element, is preserved and develops a theory of locally stable manifolds which form measurable functions rather than smooth ones, but are still absolutely continuous. These in turn are used to establish the basic ergodicity properties. The main results are:

- (I) The diffeomorphism has at most a countable number of ergodic components (after ignoring an invariant set of measure 0).
- (II) On each ergodic component some iterate of the diffeomorphism is Bernoulli.

For flows f_t the results are quite similar, the exponents now capture the possible values of

$$\limsup_{t \rightarrow +\infty} \frac{1}{t} \log \|df^t v\|,$$

where v is in the tangent space to the manifold. In the proofs of these results an important role is played by the *entropy formula* which computes the entropy of the system in terms of the exponents:

$$h(f) = - \int_M \sum_{\lambda_i(x) \geq 0} \lambda_i(x) d\mu(x).$$

Here we must take account of the multiplicity and repeat each λ_i according to the dimension of the associated eigenspaces.

It follows that if a diffeomorphism of a two-dimensional manifold has positive entropy, then at least one exponent is positive. Then the fact that the diffeomorphism preserves a smooth measure shows that the other exponent must be negative and so the theory applies and this is how the general result concerning positive entropy diffeomorphisms of 2-manifolds is obtained. For flows there is always one neutral direction, and the theory applies when all

exponents not in the flow direction are different from zero. This gives the result for positive entropy flows on 3-manifolds.

Even before this theory, Anosov and Sinai [AnoSi] succeeded in extending the analysis to some physical systems that were not completely smooth. The archetypal example was the two-dimensional billiards with a convex scatterer. The inherent hyperbolicity enabled them to construct stable and unstable manifolds, establish absolute continuity, and prove the K property for these systems. Then in [GO] the Bernoulli property was established and some perturbation results of [E] enable us to extend some of our results to such systems.

In [KatoStr] the theories of Pesin and nonsmooth systems are combined to obtain quite general results for partially hyperbolic nonsmooth systems. There is much room here for further work.

An important difference between the Pesin or billiard foliations, and those of axiom A , is that their lack of uniformity destroys the proofs of ϵ -shadowing and structural stability. It is hoped that much of this gap will be filled by statistical stability.

4.6 Proving that smooth systems are Bernoulli. The first example of a Bernoulli flow was given in [O 5]. This was the flow built on an independent process on two equiprobable symbols with a height function taking constant values on the two sets of the basic partition. The first smooth flow that was shown to be Bernoulli was the geodesic flow on a two-dimensional manifold with negative curvature [OW 3]. The method developed in that paper forms the basis for almost all of the subsequent proofs that smooth or otherwise physically defined systems (such as billiards) are Bernoulli.

The situation that appeared there, and indeed the most frequently occurring one in which we can prove that a system is Bernoulli is that of hyperbolicity (i.e., if E_x^s, E_x^u represent the asymptotically contracting and expanding directions at x , then for a set of full measure $\dim E_x^s + \dim E_x^u = \dim$ manifold for diffeomorphisms) $\dim E_x^s + \dim E_x^a + 1 = \dim$ manifold for flows. In most cases where this happens we can prove:

(A)

- (1) A countable number of ergodic components exhaust the given invariant measure;
- (2) On each such component the flow is either Bernoulli or a Bernoulli flow times a pure rotation (finite periodic transformation for diffeomorphisms).

Once (A) has been established, further special arguments are needed to establish ergodicity. This has been carried out in some special cases like the geodesic flow on a negatively curved manifold, certain billiards with convex scatterers, axiom A attractors, and others. In these cases it also turns out that in the alternative in (2) above the rotation part does not occur so that these flows are indeed Bernoulli.

Here is a precise condition that implies (A). Denote by $\mathcal{E}^s, \mathcal{E}^u$ foliations that integrate E^s and E^u ; the leaves of these need not be complete manifolds, but there are enough of them so that a set of full measure can be covered by *hyperbolic blocks*. For a single transformation, a *hyperbolic block* is obtained from a bundle B of parts of leaves from \mathcal{E}^s and another such bundle C from \mathcal{E}^u , such that each contracting piece intersects each expanding piece in exactly one point. The hyperbolic block is the set of all these intersection points, and thus has the abstract structure of a product space.

(HB)

- (a) A set of full measure can be covered by a countable number of hyperbolic blocks;
- (b) In each of which the invariant measure μ is equivalent to a product measure.

For flows we define a transverse hyperbolic block by requiring that bundles B and C from \mathcal{E}^s and \mathcal{E}^u , when projected by the flow onto a manifold transverse to the flow, form a hyperbolic block in the previous sense. (We define a hyperbolic block to be a transverse hyperbolic block thickened by the flow.)

(HB)

- (a) There are a countable number of transverse hyperbolic blocks that sweep out a set of full measure under the flow.
- (b) On each transverse hyperbolic block, the transverse invariant measure is equivalent to a product measure (transverse measure is defined by thickening the transverse manifold in the flow direction).
- (c) If O is an open subset of a manifold transverse to B (or C), then the part of B (or C) that intersects O is measurable.

In the following we assume that the invariant measure is separable and that continuous functions are dense in L_1 . The only property

of the leaves of \mathcal{E}^s (and \mathcal{E}^u) that we need once we have (HB) is that they contrast under $f_t(f_{-t})$.

Theorem. *The condition (HB) implies (A).*

It is worth pointing out that once (HB) has been established, then the smoothness of the flow is no longer needed to prove the theorem. Thus the nonsmooth systems such as billiards or bouncing balls are covered by this format.

In the literature one does not usually see (HB) (b) proven explicitly; however, it can be established given the conditions:

- (i) In the disintegration of the invariant measure μ along the leaves of \mathcal{E}^u , the conditional measures are absolutely continuous with respect to Lebesgue measure on these leaves;
- (ii) The foliations \mathcal{E}^u and \mathcal{E}^s are absolutely continuous.

If the invariant measure is smooth, then (ii) alone suffices since it implies (i) in that case.

- (iii) Fix one leaf in a hyperbolic block, and let O be an open subset of this leaf. Then the collection of leaves (from the other bundle) that intersect O is measurable.

This formulation is new, and a detailed proof will appear in a later publication of ours. We give here an outline of the proof of the theorem.

The proof of (A)(1) follows the outline sketched above in 4.2. The assumption that μ is equivalent to produce measure on a hyperbolic block shows that if the invariant set has positive intersection with a hyperbolic block, then it completely fills it whence the fact that the invariant σ -algebra is atomic and hence a countable number of ergodic components exhaust the measure.

For the next step we want to investigate the Pinsker algebra—or algebra of zero entropy factors for the flow. It may not be easy to establish the existence of a smooth generating partition; but, nonetheless, there are smooth partitions P_n such that $\bigcup P_n$ generates and using the fact that the full Pinsker algebra can be approximated by the Pinsker algebras of the factors defined by the P_n 's one sees that sets in the Pinsker algebra contain entire leaves of \mathcal{E}^u and \mathcal{E}^s (up to μ -measure zero of a set of leaves and conditional measures on leaves). We now want to show that the flow on the factor defined by the Pinsker algebra contains a single orbit (having fixed one ergodic component that gave positive measure to some hyperbolic block). For any fixed hyperbolic block

B we take some T_0 sufficiently large and then verify that if u is a subset of the Pinsker algebra whose intersection with B has positive measure, then $\bigcup_{|t| \leq T_0} f_t(u)$ fills all of the block B . It is easy to see that this implies that f_t restricted to the Pinsker algebra consists of a single orbit. If the orbit is a point, then the flow is a K -flow, otherwise it is the flow built under a constant function over a measurable cross section. This cross section is completely fibered by \mathcal{E}^s and \mathcal{E}^u .

Next we have to show that the K -flow is Bernoulli. For this we use the method of our paper [OW 3]. The new observation is that when μ is equivalent to the product measure its Radon–Nikodym derivative with respect to product measure on a fixed hyperblock is measurable and therefore locally almost constant. If we restrict to small hyperbolic blocks, we get that the R–N derivative is constant but for a set of small measure; and, therefore, the map along leaves of \mathcal{E}^{s+1} between leaves of \mathcal{E}^u is close to measure-preserving for most pairs of leaves.

It is not easy in general to get a nice disjoint covering of the space by such hyperbolic blocks. We get around this difficulty by looking at $f_n(B)$ for many n 's and one fixed good hyperbolic block of the type that we just described. The ergodic theorem says that for N large these $\{f_n(B) : 0 \leq n \leq N\}$ form an almost even covering of the space, and we use this almost even covering to map fixed past atoms along leaves of \mathcal{E}^{s+1} onto the whole space. When this is done in an almost measure-preserving way, this construction verifies the VWB condition for a fixed smooth partition. For a smooth partition nearby points on the same leaf of \mathcal{E}^{s+1} will have almost the same future name, and this gives the closeness in \bar{d} of the future conditioned on the past with the unconditional future. The K -property is used in an essential way in this argument to see that most past atoms are well distributed among the mapping boxes.

When the flow is not K , first we see that the Poincaré map on the measurable cross section is K . This is done by extending the cross section by f_t , $0 \leq t \leq t_0$ with t_0 small. Then an argument similar to the one above for maps rather than flows shows that the Poincaré map is a Bernoulli shift, and it follows then that the whole flow is a Bernoulli flow times a pure rotation.

For some specific systems, other methods have been used. Even before the isomorphism theory of Bernoulli shifts, Markov partitions were introduced in [AdW 1,2] and used there to classify automorphisms of the 2-torus by special finite state Markov shifts

which were further classified via methods of symbolic dynamics by their entropy. Later Azencott [Az] used these Markov partitions to verify the VWB condition and prove Bernoulliness. For nonhyperbolic toral automorphisms VWB was verified using some diophantine estimates [Katz]. In some constructions of smooth systems the Bernoulli property is obtained by using an additional ingredient from the abstract theory—namely the result of Rudolph [Rud 4] that mixing compact extensions of Bernoulli flows are Bernoulli, see [BrFeKato].

We should point out that the method for verifying the K -property is due to Sinai and predates the Bernoulli theory. The importance of strengthening K to Bernoulli is that at the level of abstraction of isomorphism the Bernoulli property is definitive—there is only one Bernoulli flow.

A large collection of systems that have been shown to be Bernoulli (or K) can be found in [Si 7].

4.7 Smooth models for abstract systems. The most important question on the interface between smooth dynamics and abstract ergodic theory is what properties, if any, are forced upon a dynamical system in virtue of its coming from a smooth flow or transformation. There is one basic restriction that comes from the fact that the manifold is finite dimensional and that is the finiteness of the entropy. Are there any further restrictions? If one admits non-smooth invariant measures, then the Smale horseshoes show that one has all processes that can be represented on a finite shift space and then since every finite entropy process has a finite generator [Kri] we get all finite entropy processes. The way in which these sit on the manifold makes it clear that we should not be considering these as smooth systems and so henceforth we shall restrict to smooth invariant measures.

Next it is worthwhile observing that without compactness the problem becomes much easier. Here it was shown in [ArOW] that any ergodic flow has a model as a C^∞ -flow preserving a smooth measure on a two-dimensional manifold that is of course not compact. Having focused the problem the results become much more scarce. It has been shown that on all compact manifolds there are smooth Bernoulli diffeomorphisms (cf. [BrFeKato])²⁸. In the

²⁸Related results are: Any compact closed manifold of dimension 2 or 3 has a C^∞ Riemannian metric whose geodesic flow is Bernoulli [Ka 2]. Any compact real analytic surface has a real analytic metric whose geodesic flow is Bernoulli [BurGe].

other direction, there are some non-Bernoulli K -automorphisms that have been given smooth representations [Ka 1, Rud 7] but no essential progress has been made on the general question.

It is worth pointing out that for a pair of commuting transformations entropy provides further restrictions. Here we are really thinking about a \mathbf{Z}^2 -action rather than a \mathbf{Z} -action and again ask when does an abstract \mathbf{Z}^2 -action have a smooth realization. This means a pair of commuting diffeomorphisms preserving a smooth measure such that the \mathbf{Z}^2 -action defined by $f^k g^l$ is isomorphic to the given \mathbf{Z}^2 -action. It was pointed out by M. Herman that for a pair of commuting transformations of that type the entropy inequality

$$h(fg) \leq h(f) + h(g)$$

is a consequence of the Pesin theory. Since there are finite entropy examples of abstract commuting transformations for which the inequality is false (cf. the discussion in [OW 1, Appendix A(b)]), one sees a further restriction arising from the entropy theory.

In low-dimensional manifolds smooth actions have special properties. In dimension one, homeomorphisms cannot have very complicated dynamics, and essentially one gets either trivial dynamics or things like circle rotations. In dimension two, homeomorphisms can be quite general but diffeomorphisms are limited. In particular if a diffeomorphism of a compact two-dimensional manifold preserves a smooth invariant measure then (by Pesin) on every ergodic component of positive entropy it must be Bernoulli times a finite rotation. Similarly if f_t is a smooth flow on a three-dimensional compact manifold preserving a smooth invariant measure, then on every ergodic component of positive entropy f_t is Bernoulli times a rotation.

5. PROOFS OF THE NEW RESULTS

5.1. Proofs for strong stochastic stability (§2.1). Theorem 2.1.1 is a special case of Theorem 2.1.3 once one observes that a geodesic flow on a manifold of negative curvature is an Anosov flow in which the expanding and contracting foliations are not jointly integrable. This may be found in [Ano 1].

Theorem 2.1.3. *A C^2 axiom A attractor is SSS if and only if it is not the C^1 limit of C^2 suspensions.*

According to Proposition 5.4 of [BRue], if the perturbation \tilde{f}_t converges to f_t in the C^1 topology.

- (1) The SBR measures $\tilde{\mu}$ of \tilde{f}_t converge weakly to the SBR measure μ of f .
- (2) $h(\tilde{f}_t, \tilde{\mu})$ converges to $h(f_t, \mu)$.

According to [Rat] the flow (f_t, μ) is Bernoulli. By the stability of axiom A attractors, $(\tilde{f}_t, \tilde{\mu})$ is again an axiom A attractor, of course, and $(\tilde{f}_t, \tilde{\mu})$ is also Bernoulli. Letting P be a partition of the manifold into sets of small diameter, P will generate under both f_1 and \tilde{f}_1 (this follows from the local product structure and structural stability). Relation (1) will imply that for any T

$$\bar{d} \left(\text{dist} \bigvee_0^T f_t P, \text{dist} \bigvee_0^T \tilde{f}_t P \right) \rightarrow 0$$

as the perturbation \tilde{f}_t tends to f_t . Together with (2) we get from the version of FD given by the Lemma following the definition of FD that the processes converge in \bar{d} . Now the Theorem follows from Theorem 3.6.2(d). \square

The proof of Theorem 2.1.2 follows along the same lines. The fact that the entropy of the perturbed billiards is close to the original entropy was established by K. Eloranta [E]. He also showed that for a fine enough partition (even though it may not generate now with respect to f_{t_0} for fixed t_0) its entropy varies continuously with respect to the perturbation. The invariant measure is geometrically defined and is clearly stable under the kind of perturbation that we are looking at. Finally the fact that these billiards are Bernoulli follows from [GO] and thus once again Theorem 3.6.2(d) and Lemma 3.5.1 give the result.

Proposition. *Let f_t be the direct product of a Bernoulli flow and a flow of zero entropy (in particular a rotation). Let P be a partition such that (f_t, P) has full entropy and is Bernoulli. Given ϵ , there is a \hat{P} satisfying*

- (i) $d(P, \hat{P}) < \epsilon$,
- (ii) (f_t, P) is congruent to (f_t, \hat{P}) ,
- (iii) f_t is the direct product of the zero entropy factor and the σ -algebra generated by \hat{P} under f_t .

Proof. Iterating Lemma 5 of [Fi 1] gives a \hat{P} that satisfies (i), (ii), and in addition generates together with the zero entropy factor.

The fact that (iii) holds, i.e., that (f_t, \widehat{P}) is independent of the zero entropy factor follows from the basic disjointness result of H. Furstenberg [Fu].

Theorem 2.1.4. *If f_t is an axiom A attractor that is not a suspension, even if it is a C_1 -limit of suspensions, then one still has a lot of stability: the perturbation \overline{f}_t is α -congruent either to f_{ct} or to $R_t \times f_{ct}$, $|c - 1| < \alpha$.*

Proof. Here the only novelty occurs when the perturbation \tilde{f}_{ct} is the direct product of a Bernoulli flow and a rotation. Having chosen c , so that

$$h(\tilde{f}_{ct}) = h(f_t)$$

we use the Lemma following the definition of FD to get closeness in \overline{d} and then we apply Theorem 3.6.1(b) to $(\tilde{f}_{ct}P)$ to get \widehat{P} , close to P such that $(\hat{f}_{ct}, \widehat{P})$ is congruent to (f_t, P) . Now we can apply Theorem 3.7.2 to conclude the proof.

Proposition. *Let f_t be the flow on an axiom A attractor M . Then given $\alpha > 0$, if \overline{f}_t is a C^2 flow, sufficiently close C^1 to f_t , there is an invertible measurable²⁹ map from M to M that takes orbits to orbits and moves all points by $< \alpha$ except for a set of measure $< \alpha$.*

Proof. If f_t is an axiom A attractor and \tilde{f}_t a small perturbation, then the Bowen–Ruelle result still applies. However, neither f_t nor \tilde{f}_t is necessarily Bernoulli.

(1) Make a small ($< \epsilon$) change in the parametrization of f_t to get \hat{f}_t . We can ensure that \hat{f}_t is Bernoulli and, in fact, we can make \hat{f}_t axiom A, so its SBR measure will be a scalar multiple of the original SBR measure.

(2) With \hat{f}_t fixed, we can reparametrize each \tilde{f}_t to $\tilde{\tilde{f}}_t$ so that the $\tilde{\tilde{f}}_t \rightarrow \hat{f}_t$ in entropy and finite distribution. We make another small change in $\tilde{\tilde{f}}_t$, calling it $\overline{\tilde{\tilde{f}}}_t$ so that $\overline{\tilde{\tilde{f}}}_t$ is Bernoulli and $\overline{\tilde{\tilde{f}}}_t$ still converges to \hat{f}_t in entropy and finite distributions. As in the previous proof, we can conclude at this point that \hat{f}_t and $\overline{\tilde{\tilde{f}}}_t$ are ϵ -congruent. This gives that f_t is ϵ -congruent to a small reparametrization of \tilde{f}_t . \square

²⁹If ϕ is not an isomorphism it cannot take an invariant measure to an invariant measure.

Proof of the dichotomy for axiom A attractors. In other words an axiom A attractor is either Bernoulli or the direct product of a Bernoulli and a rotation. Ratner proved Bernoulliness in the C -dense case (see [BRue], Remark 3.5). Bowen [B 1], Theorem 3.2, shows that an axiom A attractor is either C -dense or a suspension (of a homeomorphism satisfying axiom A). The suspension is the suspension of a Bernoulli shift or the product of a Bernoulli shift and a finite cyclic permutation. One way to see this is to thicken the base and use absolute continuity of the expanding and contracting foliation that are in the base (this is proved independently whether we have a suspension) as in [OW 2] to prove Bernoulliness. Because a Bernoulli shift can be imbedded in the Bernoulli flow, we can write the suspension as the direct product of the Bernoulli flow and a rotation. \square

Remark. In a similar vein we establish our version of Pesin's theorem that we gave in §1.5. In [Pe] Theorem 9.7, Pesin states that the set on which we have nonzero Lyapunov exponents is the union of a countable number of ergodic components and on each of these we have either a Bernoulli flow or a suspension. He also proves that positive entropy implies positive Lyapunov exponents and obtains expanding and contracting fibers [Pe, Theorem 5.1]. The absolute continuity of the expanding and contracting fibers is proved independently of whether the flow is a suspension and therefore the base, which contains complete expanding and contracting fibers, is a Bernoulli shift or the product of a Bernoulli shift and a finite cyclic permutation. Because Bernoulli shifts can be imbedded in a flow, the flow we started with is a Bernoulli times a rotation.

5.2. Proofs for §2.4: Bernoulli flows and Markov processes.

Theorem 2.4.1. *If f_t is a Bernoulli flow on a metric space M then, given α , there is a (special) semi-Markov process on M that is α -congruent to f_t .*

Proof. The original flow that was shown to be Bernoulli was the flow built under a function, where the base transformation T has an independent generator (P_1, P_2) , and the height function k was constant on the atoms P_1, P_2 with values k_1, k_2 so that k_1/k_2 is irrational. With α given, choose these k_i so that f_t does not move points in M by more than $\frac{1}{100}\alpha$ for $0 \leq t \leq \max(k_1, k_2)$. We then choose probabilities for P_1 and P_2 that give the correct entropy. Then use the fact that f_t is isomorphic to the Bernoulli

flow described above to find a cross section $Y \subset X$ for the flow f_t so that the return time function will take the values k_1, k_2 on sets P_1, P_2 in Y , with (P_1, P_2) being an independent generator for (T, Y) . Here $T: Y \rightarrow Y$ is the return map to the cross section. The fact that $P = (P_1, P_2)$ is a generator implies that for N sufficiently large, the atoms of $\bigvee_{-N}^N T^j P = Q$ are mostly contained in sets of diameter less than $\frac{1}{100}\alpha$. The process (T, Q) is a Markov process, so that the flow (f_t, \widehat{Q}) is a semi-Markov flow (\widehat{Q} is the partition of M defined by assigning a point x of M to that atom of Q that the point first encounters when flowing back to the cross section Y). Choose now for each atom Q_i of Q a point x_i so that we minimize

$$e_i = \mu \{y \in Q_i: d(y, x_i) \geq \frac{1}{100}\alpha\}.$$

If the N is sufficiently large then we can arrange things so that

$$\sum_Q e_i < \frac{1}{100}\alpha.$$

Replacing the sets of \widehat{Q} by points x_i we get a concrete semi-Markov flow that is isomorphic to (f_t, \widehat{Q}) and the isomorphism doesn't move most points by more than α and thus we have the desired α -congruence.

For \overline{B}_t one uses the same proof except that we take k_1/k_2 rational. We thus can prove Theorem 2.4.1 for \overline{B}_t instead of B_t . \square

Here are other versions of this result:

Theorem 5.2.1. *Let P be any finite-valued observation on a system isomorphic to B_t (i.e., P is a finite function on the state space X). Then we can change P (without changing the range) on an arbitrarily small subset of X so that the resulting observation is N -semi-Markov and generates (i.e., the doubly infinite sequence of observations determines the point in X).*

An N -semi-Markov process has a finite number of possible outcomes r_i . Each r_i lasts for time t_1 or t_2 but not both. The next output is determined by a probability distribution \prod_l (we spin a roulette wheel), which is determined by what the process did between time $-N$ and 0 for some fixed N . (There are a finite number of possible \prod_l .)

Proof.

Step 1. We modify P (call it P_1) on an arbitrarily small set so that $h(f_t, P) = h(f_t)$. (We could even make P_1 generate.)

Step 2. We can make an arbitrarily small change in P_1 (call it P_2) so that every orbit can be partitioned into intervals of length k_1 or k_2 (for some k_1, k_2 small enough and k_1/k_2 irrational). P_2 is constant on each of these intervals and each value in the range of P_2 (range of $P_2 =$ range of P) is taken on intervals of length k_1 or k_2 but not both. In other words: f_t is represented as a flow built under a function g that takes values k_1 and k_2 . P_2 is constant on each vertical line between the base and g and defines the partition according to whether a point lies below k_1 or k_2 . Step 2 follows from the argument of Rudolph [Rud 1].

Step 3. Let Q be the partition of the points x in the base X , according to which atom of P_2 , contains the vertical line above x . Let T be the transformation on X . Take a sequence of multi-step Markov processes T_i, Q_i that converge to T, Q in entropy and finite distributions. Let ${}_i f_t$ be the corresponding flows and P_i the partitions corresponding to Q_i as P_2 corresponds to Q . $({}_i f_t, P_i)$ will also converge in entropy and finite distributions. Furthermore if we pick a t_0 small compared to k_1 and k_2 , P_i will generate under t_0 (this follows from the argument in [O 1, Lemma 1, p. 58]). By the Lemma 3.5.1 ${}_i f_t, \mathcal{P}_i$ converges in \bar{d} to f_t, P . We now use Theorem 3.6.2 to complete the proof. \square

Theorem 5.2.2. *Let T be a Bernoulli shift and P a finite-valued function such that $h(P, T) = h(T)$. We can change P (to \bar{P}) on a set of arbitrarily small measure (without changing the range of P) so that \bar{P} generates under T and T, \bar{P} is multistep Markov. (k -step Markov means that the probability of the next output depends only on the last k outputs.)*

This theorem will follow immediately from the discrete time version Theorem 3.6.2(b) and the following lemma:

Lemma 1. *A Bernoulli process (T, P) , P finite, is the \bar{d} -limit of k -step mixing Markov processes of the same entropy.*

To prove Lemma 1, we will need the following:

Lemma 2. *Let P, T be a finite state process. Given ϵ we can find $\bar{P}_t, 0 \leq t \leq 1$ such that $|\bar{P}_t - P| < \epsilon, 0 \leq t \leq 1, \bar{P}_t$ is continuous in t and $h(T, \bar{P}_1) < h(T, P)$ and $P_0 = P$.*

Proof. We will start by defining \bar{P}_1 . Pick a Rohlin tower of height n and pick a set E in the base such that $m(\bigcup_n^{n-1} T^i E) = \frac{1}{2}\epsilon$. We get \bar{P}_1 by changing P on $\bigcup_n^{n-1} T^i E$ so that it takes on only one value on this set. It is easy to see that if n is large enough we have dropped the entropy by at least $(1 - \frac{1}{4}\epsilon)$. To form \bar{P}_t we take instead of E a part of E of relative measure ϵ . \square

To prove Lemma 1 we pick a large n and take the n -step canonical Markov approximation to $T, P_1, (T_M, P_1)$. If n is large enough $h(T_M, P_1) < h(T, P)$. By continuity we can find a t_0 such that if we take the Markov approximation (T_M, P_{t_0}) to T, \mathcal{P}_{t_0} then $h(T_M, P_{t_0}) = h(T, P)$. We can make the finite distribution of T_M, P_{t_0} as close to those of T, P by taking ϵ small enough and since T, P is finitely determined T, P and T_M, P_{t_0} are close in \bar{d} . \square

5.3. Proof for §2.2: Random perturbations.

Proof of Theorem 2.2.1. By the theorem of [OShi] the random perturbation is a Bernoulli transformation with infinite entropy. Fix a finite partition P of the space into sets of diameters less than $\frac{1}{10}\alpha$. For this partition P , the results of [Ki 1, 2] imply that the random perturbation (S, \bar{P}) is close in entropy and finite distribution to (f, P) . By the FD property of f this implies that (f, P) is close in \bar{d} to (S, \bar{P}) . By Theorem 3.6.1 there is a \hat{P} which is close to \bar{P} so that (S, \hat{P}) is congruent to (f, P) . Now Theorem 3.7.2 will give us the theorem.

Proof of Theorem 2.2.2. We will postpone stating a lemma which would be unnecessary if we began by assuming an additional kind of uniformity for the distribution $\mu(x)$ of our ending position after the diffusion. Namely, the measures $\mu(x)$ are uniformly absolutely continuous with respect to the flow invariant measure $d\nu$. This means that for every $\epsilon > 0$ there is a $\delta > 0$ such that any set E with $\nu(E) < \delta, \mu(x)(E) < \epsilon$ for all x . With this assumption it is fairly easy to see what happens to the perturbed process. When the quiet-active process is fixed (for all time) the orbit of the perturbed process consists of random pieces of orbits of the f_t flow. This differs from the independent concatenation of equal blocks from the (f_t, M) process in two ways. The length of the blocks is variable and the distribution of each new block is not independent of the previous blocks but rather depends upon them in the simplest way possible, namely through the end position of the last block concatenated.

Observe that fixing the quiet-active process realization we have a Markovian (nonstationary) process on the manifold. The fact that each $\mu(x)$ dominates a fixed constant times the uniform distribution on a disc of fixed radius enables one to show, via a coupling argument, that fixing any two past atoms, which means of course, fixing two points on the manifold M , say u_0, v_0 , the positions of u_t and v_t for large t will have essentially the same distribution. The mixing property of the Bernoulli flow makes this coupling argument easy because after a long, fixed time, the uniform distribution on a fixed ball will be spread throughout the manifold in a uniform way and thus a fixed proportion of the distribution of u_t and v_t can be matched the next time an active period occurs. Starting with Lebesgue measure in the arbitrarily distant past we thus get an invariant distribution for the perturbed process conditioned on this fixed quiet-active orbit. Averaging this distribution over the various possible quiet-active realizations gives us an invariant measure for the perturbed process and completes the definition of the process. (Formally, our description in the text only defined transition probabilities for the process.) The additional uniformity assumption at the beginning of the proof or the lemma that follows implies that the invariant measure will be absolutely continuous and that this absolute continuity will be uniform over our perturbations.

We check the relative VWB condition by the coupling argument above and hence our process is relatively Bernoulli. Note that the entropy of the perturbed process is infinite since $\mu(x)$ has a continuous distribution.

If we assume the active-quiet process is Bernoulli we could apply Theorem 3.7.2 once we have checked that the perturbed flow is \bar{d} close to the original flow. If we do not make this assumption we have to apply Theorem 5.7.2. In this case we must check that the perturbed flow is \bar{d} close to the original flow relative to the active-quiet process. We check both of these using FD and checking finite distributions and entropy on each quiet-active orbit. Discretize time and space (this is harmless since we are only interested in the \bar{d} -distance) uniform absolute continuity of the invariant measure and the ergodic theorem imply the finite distribution property, since the starting point of the long orbit pieces is not likely to be in a small set of the invariant measure for the Bernoulli flow. The same property of the invariant measures combined with the conditional independence of the successive blocks and the Shannon–McMillan theorem enables one to see that

the entropy is close to that of the original Bernoulli flow. Then the FD property gives us the remaining hypothesis of Theorem 5.7.2 and an application of that theorem gives us Theorem 2 under our stronger assumption.

To deal with the weaker assumption that we made in formulating Theorem 2 in the text we have to establish the requisite uniformity as a consequence of this weaker assumption. Abstractly we have the following situation.

Lemma. *Suppose we have a transition kernel $P(x, dy)$ on the manifold that has the property that*

$$(*) \quad P(x, dy) \geq c \, dy \quad \text{for } y \in B_r(x),$$

where dy is the volume element and c and r are fixed positive constants, and in addition a collection of point transformations S_{t_i} , that represent f_{t_i} where the t_i 's are the times between successive active periods in a fixed quiet-active realization. Fixing a starting measure μ (Lebesgue measure), the measures

$$\mu_k = \mu P S_{k_1} P S_{k_2} \cdots P S_{k_k},$$

(as $k \rightarrow \infty$) become uniformly absolutely continuous with respect to the volume element.

Proof. Suppose not, then for some $a > 0$, and arbitrarily large k , there will be sets A_k with $\lambda(A_k) \rightarrow 0$, where λ is the Lebesgue measure defined on the manifold, and

$$\mu_k(A_k) \geq a.$$

Let a_0 be the supremum of the values of a for which such a sequence exists, and suppose that a is close to this supremum. Now the fact that the invariant measure for f_t is absolutely continuous with respect to Lebesgue measure means that if $\lambda(A_k) \rightarrow 0$ also $\lambda(S_{k_k}^{-1} A_k) \rightarrow 0$.

Since (*) holds, we can conclude that for some constant $b < 1$

$$P(x, S_{k_k}^{-1} A_k) \leq b \quad \text{all } x \text{ and } k > k_0.$$

Choose δ small enough so that $(1 - \delta)/b > 1$, and then observe that since $P(x, dy)$ is absolutely continuous for all x , if B_k is the set of x 's for which

$$P(x, S_{k_k}^{-1} A_k) \geq \delta a$$

then $\lambda(B_k) \rightarrow 0$ as $k \rightarrow \infty$. Now since

$$a \leq \mu_k(A_k) = \int \bar{\mu}_{k-1}(dx)P(x, S_{k_k}^{-1}A_k) \leq \delta a + \bar{\mu}_{k-1}(B_k)b$$

we can conclude

$$\left(\frac{1-\delta}{b}\right)a \leq \bar{\mu}_{k-1}(B_i), \quad \bar{\mu}_k = \mu PS_{k_1} \cdots PS_{k_{k-1}}.$$

If a is close enough to the supremum, $a(1-\delta)/b > a_0$, a contradiction which proves the required uniformity. \square

Proof of Theorem 2.2.3. (1) The difficult analysis of these random perturbations has been carried out by Y. Kifer [Ki 1]. The results of his that we need are the following:

1. $\mu^\epsilon \rightarrow \mu$ weakly (Proposition 1 of §2).
2. $P_{\mu^\epsilon} \rightarrow dm$ weakly. (This follows easily from (1) above.)
3. For any fixed partition P into sets of small diameter in M (viewed as a partition of the zero'th coordinate in X) and any fixed τ_0 ,

$$\lim_{\epsilon \rightarrow 0} h(\sigma_{\tau_0}, P, P_{\mu^\epsilon}) = h(\sigma_{\tau_0}, P, dm) = h(f_{\tau_0}, \mu).$$

Here σ_{τ_0} represents the shift on X by τ_0 , and the entropy is taken with respect to the invariant measure indicated in the argument. The last equality holds because if P is sufficiently fine it is a generator for f_{τ_0} .

(2) It follows from (1) and from the fact that the Bernoulli shift (f_{τ_0}, μ) is finitely determined that for any τ_0 if ϵ is sufficiently small the process $(\sigma_{\tau_0}, P, P_{\mu^\epsilon})$ is converging in \bar{d} to (f_{τ_0}, P, μ) . We want to conclude that the flows $(\sigma_t, P, P_{\mu^\epsilon})$ are also close in \bar{d} . For this, one only needs to remark that the uniformities on the diffusion involved guarantee that with probability close to one the full trajectory of the Markov process x_t^ϵ is close in \bar{d} to the trajectory determined by its τ_0 skeleton (if τ_0 is small enough). After this observation Theorem 3.6.1 will give us the imbedding of $(f_t; \mu)$ into the x_t process by an imbedding that is close to the identity.

As before we are now in the position to apply Theorem 3.7.2 to complete the proof of the theorem. \square

5.4. Proofs for §2.5: Long-term versus short-term behavior.

Theorem 2.5.1. *Given f_t, X, P , where (f_t, X) is Bernoulli and $\alpha > 0$, then for a.e. $x \in X$, and T large enough the following*

algorithm will produce a process $(\overline{f}_t, \overline{X}, \overline{P})$ that will be α -congruent to the direct product of (f_t, X, P) and another flow (g_t, Y) relative to the functions \overline{P} and \tilde{P} where $\tilde{P}(y, x) = P(x)$, $x \in X$, $y \in Y$. If P identifies X with a manifold M , then we can think of (g_t, Y) as a viewer and if \hat{P} is the image of \overline{P} under the α -congruence then we see $\hat{P}(y, x)$ when the viewer is in state y and we are looking at x in $M \sim X$.

The major part of the work needed to prove this theorem is contained in [OW 4] where we prove that the processes constructed by the algorithm described in the text converge in \bar{d} to the flow (f_t, P) . In order to apply Theorem 3.7.2 it remains only to see that the process obtained by independent concatenations is a Bernoulli times a pure rotation. This follows readily by the relativized theory described in §3.7. There is an obvious rotation factor due to the fixed length of the blocks we are concatenating and the random spacers of fixed lengths. Fixing this factor it is trivial to check relative VWB and thus Theorem 3.7.2 applies to prove the theorem.

Discussion for remark. Let us first describe the simulation more precisely. We fix a t_0 sufficiently small so that f_t , the flow that integrates the given vector field, doesn't move points on the manifold by more than $\alpha/10$ for $|t| \leq t_0$. Then we take L large enough so that if P is a fixed smooth finite generator under t_0 , the distribution on L -blocks given by $\bigvee_{k=0}^{L-1} f_{-kt_0} P$ concatenated independently will be within $\alpha/100$ in \bar{d} of the process (f_{t_0}, P) .

This follows from the definition of FD.³⁰ Next we discretize time finely enough and approximate the vector field well enough so that the numerical integration of the vector field will approximate to within $\alpha/100$ the true orbit out to time $t_0 L$. Finally space is discretized to sets A_j such that

1. all the A_j 's have equal Lebesgue measure,
2. if x and y belong to the same A_j , $f_t x$ and $f_t y$ stay within $\alpha/100$ of each other for $0 \leq t \leq t_0 L$.

Now points $x_j \in A_j$ are chosen and a computer program is written that will move x_j at times kt_0 , $0 \leq k < L$ to within $\alpha/50$ of

³⁰Note, that while t_0 can be estimated in advance from the vector field, L cannot since it depends upon the ergodic properties of the flow f_t with respect to a presumptive smooth invariant measure. Thus various L 's must be attempted. Given L and t_0 , the further approximations and discretizations needed can be determined purely in terms of the vector field.

where f_{kt_0} would really move x_j to. If the invariant measure were Lebesgue measure, then the distribution on L -blocks that we get by randomizing uniformly the starting point of the computation over the x_j 's would clearly be very close to the invariant measure distribution on L -blocks with respect to P . Thus the independent concatenation is close in \bar{d} to the process (f_{t_0}, P) and if we add entropy as in part A above then Theorem 3.7.2 will complete the proof.

If the invariant measure has a bounded density with respect to Lebesgue measure, or if we randomize only over a fixed small piece of the space, we invoke the extremality property. When choosing L this will provide a further restriction, and then the proof continues as before.

5.5 . Proofs §2.3: Scaling time.

Theorem 2.3.1. *Let (B_t, Y, P) be the Bernoulli flow where P identifies Y with the manifold M . Given α , if $\gamma > 0$ is small enough then $(B_{(1+\gamma)t}, Y, P)$ is α -congruent to $(B_t \times B_{\gamma t}, Y \times Y, \bar{P})$ where $\bar{P}(r_1, r_2) = P(r_1)$.*

In the flow $(B_{(1+\gamma)t}, Y, P)$, think of $(B_{(1+\gamma)t}, P)$ as being an approximation to (B_t, P) . Lemma 3.5.1 shows that as γ tends to zero the \bar{d} -distance between them tends to zero. Thus there is a \bar{P} such that $d(\bar{P}, P)$ is small and $(B_{(1+\gamma)t}, \bar{P})$ is congruent (B_t, P) , by Theorem 3.6.1(b). Now we are in a position to apply Theorem 3.7.2 to conclude the proof. \square

5.6. Proofs for instability (§2.6). Theorem 2.6.1 follows at once from Theorem 3.7.1(c), since, if there were no such α , f_t would be Bernoulli.

For the discussion of the results of §2.2 we need to verify that indeed our random perturbations are Bernoulli.

First consider the setup in Theorem 2.2.1. The compactness of the manifold, and the fact that we jump with a uniform distribution in a ball of radius r around $f(x)$, imply that the Markov process has a unique finite invariant measure absolutely continuous with respect to Lebesgue measure. Now one checks the VWB property just like we checked the relative VWB in §5.3 (see also [McCSh, OSh]). A similar argument will apply to the discrete time skeleton of the diffusion in Theorem 2.2.3. For Theorem 2.2.2, after we added the hypothesis that the interference producing process

is Bernoulli, the relative Bernoulli properties that we checked in §5.3 give that the perturbations are Bernoulli.

5.7. Some proofs for §3. Our main task here is to prove Lemma 3.5.1 and to deduce the infinite entropy case of some of our theorems from the finite entropy case which can be found in the literature.

Lemma 3.5.1. *If (f_t, P) is finitely determined and P is a finite partition then given $\epsilon > 0$, there is an $\bar{\gamma} > 0$, $\bar{\delta} > 0$, and T such that if (\bar{f}_t, \bar{P}) is ergodic and satisfies*

- (a) $\frac{1}{|t_0|} h(\bar{f}_{t_0}, \bar{P}) \geq h(f_t, P) - \bar{\gamma}$ for some $t_0 \in \left(\frac{1}{2}, 2\right)$
- (b) $\bar{d}(\text{dist} \vee_0^T \bar{f}_t \bar{P}, \text{dist} \vee_0^T f_t P) < \bar{\delta}$,

then

$$(*) \quad \bar{d}((\bar{f}_t, \bar{P}), (f_t, P)) < \epsilon.$$

Note that $h(\bar{f}_{t_0}, \bar{P})$ refers to the entropy of the \mathbf{Z} -action generated by \bar{f}_{t_0} , while $h(f_t, P)$ refers to the entropy of the \mathbf{R} -action, f_t , $t \in \mathbf{R}$.

Proof of the lemma. The fact that (f_t, P) is finitely determined means that given $\epsilon > 0$, we have γ , $\delta(\beta)$, $T(\beta)$ and that $(*)$ follows if (i) and (ii) of the definition of FD above hold. Our main task is to see how the entropy condition (i) can be made to follow from (a) in the presence of (b) and where β is determined by (f_t, P) , $\bar{\gamma}$, and $\bar{\delta}$ without further knowledge of (\bar{f}_t, \bar{P}) . For this we have to be able to control continuous names through the discrete skeleton and this is done using the measurability of (f_t, P) and the fact that P is finite. Together these imply that for any $\theta > 0$, by making $\eta > 0$ sufficiently small, we can guarantee that the set of $x \in X$ for which

$$(**) \quad |\{t \in [-\eta, \eta]: P(f_t x) = P(x)\}| \geq \frac{99}{100}(2\eta)$$

has measure greater than $1 - \theta$. Replacing θ by 2θ and $\frac{99}{100}$ by $\frac{9}{10}$ this will continue to hold for (\bar{f}_t, P) if $\bar{\delta}$ is sufficiently small and T sufficiently large. We let $E \subset \bar{X}$ be the set of \bar{x} 's where this analogue of $(**)$ is valid.

The ergodic theorem now ensures for us that for most \bar{x} and N large enough

$$\{0 \leq n < N: \bar{f}_{nt_0} \bar{x} \in E\} \geq (1 - 3\theta)N.$$

Finally if two such \bar{x} 's have their $f_t, \bar{P}, t \leq N$ names close enough (this is one of our restrictions on β) then in fact their $(\bar{f}_{nt_0}, \bar{P})_{n \leq N}$ names will agree up to an error of at most 10θ . For θ sufficiently small and β as above (note that the choice of β is independent of the nature of (\bar{f}_t, \bar{P}) and depends only on θ, η which were obtained from f_t, P) this in turn implies that most of the typical $(\bar{f}_t, \bar{P}, \beta, N)$ balls can be covered by fewer than $2^{\gamma N/4} (\bar{f}_{t_0}, \bar{P}, N)$ balls. Thus if

$$\frac{1}{|t_0|} h(\bar{f}_{t_0}, \bar{P}) \geq h(f_t, P) - \frac{1}{2}\gamma$$

then we will get as well

$$h(\bar{f}_t, \bar{P}, \beta) \geq h(f_t, P) - \gamma.$$

The final choices are $\bar{\gamma} = \gamma/2, T = T(\beta), \bar{\delta} = \delta(\beta)$ and (a) and (b) will imply (i) and (ii) and hence (*) will follow. \square

Theorem 3.7.2 is easily deducible from results in [Fi 2]. Fieldsteel's results are for *finite entropy* and for convenience we record here the one we need in our notation.

A. *Prevalence of factors with Bernoulli complements.* Let (f_t, X) be a flow with R a partition such that $\bigvee_{-\infty}^{\infty} f_t R$ has a Bernoulli complement. If Q is any partition, for all ϵ there is a \bar{Q} such that

- (1) $d(Q, \bar{Q}) < \epsilon,$
- (2) $(f_t, R \vee Q)$ and $(f_t, R \vee \bar{Q}),$ are identically distributed,
- (3) $\bigvee_{-\infty}^{\infty} f_t (R \vee \bar{Q})$ has a Bernoulli complement in $(f_t, X).$

This is Lemma 3 of [Fi 2].

Theorem 3.7.2. *Suppose that (X, \mathcal{B}, f_t) is a Bernoulli flow or the direct product of a Bernoulli flow and a flow of zero entropy. Let P be a partition such that (f_t, P) is Bernoulli. Given $\epsilon,$ there is a \hat{P} satisfying:*

- (i) $d(P, \hat{P}) < \epsilon,$
- (ii) (f_t, P) and $(f_t, \hat{P}),$ are identically distributed
- (iii) there is an f_t -invariant σ -algebra $\mathcal{A},$ independent of $\bigvee_{-\infty}^{\infty} f_t \hat{P}$ such that

$$\mathcal{A} \vee \left(\bigvee_{-\infty}^{\infty} f_t \hat{P} \right) = \mathcal{B}.$$

Note that in case (f_t, P) has full entropy, \mathcal{A} can be taken to be the full zero entropy factor alone. (If in addition f_t is a Bernoulli flow then \widehat{P} can be taken to generate. (This is theorem 3.6.2 c).

Proof. For finite entropy the theorem follows at once from (A), and the fact that we have already noted that any Bernoulli factor is independent of any zero entropy factor.

In case of infinite entropy for X , but finite entropy for P we first cut down to a large finite entropy factor that has a Bernoulli complement. (Model the Bernoulli flow by the direct product of countably many finite entropy flow.) In that factor, if it is large enough, there is a partition P_1 that is close to P and thus (f_t, P_1) and (f_t, P) are very close in \overline{d} -metric. Therefore we can imbed an exact copy of (f_t, P) in the finite entropy factor that is close to P , apply the result already proved there, and then include in \mathcal{A} also the remaining infinite entropy Bernoulli complement, and the 0-entropy factor.

When (f_t, P) itself has infinite entropy we have to first find in that process a finite entropy factor P_1 such that $d(P, P_1)$ is very small and $\bigvee_{-\infty}^{\infty} f_t P_1$ has a Bernoulli complement in $\bigvee_{-\infty}^{\infty} f_t P$, with a generator P_2 . We now apply the previous step to (f_t, P_1) to obtain \widehat{P}_1 such that $d(P_1, \widehat{P}_1)$ is small, and $\bigvee_{-\infty}^{\infty} f_t \widehat{P}_1$ has a complementary invariant sub- σ -algebra A_1 . In A_1 there is a Bernoulli complement to the zero entropy factor (by the way, we obtained A_1) and there we find a model for (f_t, P_2) , say \widehat{P}_2 , that generates this complement (since there is only one Bernoulli flow of infinite entropy). Now in $\bigvee_{-\infty}^{\infty} f_t (\widehat{P}_1 \vee \widehat{P}_2)$ we can find a copy of P , say \widehat{P} , which is a generator and $d(\widehat{P}, \widehat{P}_1) = d(P, P_1)$. This \widehat{P} will satisfy our requirements. \square

Theorem 3.6.2.

(b) *If (X, f_t) and $(\overline{X}, \overline{f}_t)$ are Bernoulli flows of equal entropy and P, \overline{P} are partitions that satisfy*

$$\overline{d}((f_t, P), (\overline{f}_t, \overline{P})) < \frac{\gamma}{10}$$

then (X, f_t, P) is γ -congruent to $(\overline{X}, \overline{f}_t, \overline{P})$.

(d) *If (X, f_t, P) is a Bernoulli flow and $\gamma, \eta > 0$ are given then there is a $\delta > 0$ such that any Bernoulli flow $(\overline{X}, \overline{f}_t, \overline{P})$ that satisfies*

$$(i) \overline{d}((f_t, P), (\overline{f}_t, \overline{P})) < \gamma/20,$$

- (ii) $|h(f_t) - h(\bar{f}_t)| < \delta$ is γ -congruent to (X, f_{ct}, P) for some constant c that satisfies $|1 - c| < \gamma$.

Part (d) has no analogue in discrete time since it depends upon a rescaling of the given flow f_t so as to get it to have an entropy which equals the entropy of (\bar{f}_t, \bar{P}) . That simple observation together with Proposition 3.5.1 of the preceding section which says that for c close to 1, (f_{ct}, P) is close to (f_t, P) in \bar{d} , reduces part (d) to part (b). In turn when P and \bar{P} have full entropy then one proves part (b) as follows:

When $h(f_t, P) = h(\bar{f}_t)$: Apply part (c), a special case of Theorem 3.7.2, to (f_t, P) to get a P_1 that is a generator and satisfies $d(P, P_1) < \gamma/10$ and (f_t, P_1) and (f_t, P) are identically distributed. Then apply Theorem 3.6.1 and imbed (f_t, P_1) in \bar{f}_t regarding \bar{P} as the initial attempt. Thus there is a \bar{P}_1 such that $d(\bar{P}_1, \bar{P}) < 3\gamma/10$ and (\bar{f}_t, \bar{P}_1) and (f_t, P_1) are identically distributed. Now apply part (c) again to move \bar{P}_1 to \hat{P}_1 which still has the same distribution as (\bar{f}_t, \bar{P}_1) , and now generates under \bar{f}_t . The fact that (f_t, P_1) and (\bar{f}_t, \hat{P}_1) are identically distributed and both P_1 and \hat{P}_1 are generators gives an isomorphism θ between f_t and \bar{f}_t which sends P_1 to $\hat{P}_1(P_1(x) = \hat{P}_1(\theta(x)))$, $d(P, P_1) < \gamma/2$ and $d(\hat{P}_1, \bar{P}) < \gamma/2$ translates to γ -congruences between (X, f_t, P) and $(\bar{X}, \bar{f}_t, \bar{P})$.

When P and \bar{P} don't have full entropy use Theorem 3.6.1 (b) to change \bar{P} to \tilde{P} by a small amount so that (\bar{f}_t, \tilde{P}) has the same distribution as (f_t, P) . Now Theorem 3.7.2 may be applied to both \tilde{P} and P to get \tilde{P}_1 and P_1 such that the canonical isomorphism between the processes $(f_t, (P_1)_{-\infty}^{+\infty})$ and $(\bar{f}_t, (\tilde{P}_1)_{-\infty}^{+\infty})$ extends to an isomorphism of the whole flow. This isomorphism gives the required γ -congruence.

We will need an analogue of Theorem 3.7.2 for the proof of Theorem 2.2.2 when the active quiet process is not Bernoulli. To this end it is convenient to define what we mean by saying that in a flow (f_t, X) a partition P defines a process (f_t, P) that is ϵ -close in \bar{d} , relative to a factor \mathcal{A} , to another process (g_t, Q) . We mean that for a.e. fiber C of the factor \mathcal{A} (picture X as a measurable bundle of measure spaces over $X|\mathcal{A}$, the space of the factor).

$$(*) \quad \bar{d}((f_t, P)|C, (g_t, Q)) < \epsilon.$$

Note that the process $(f_t, P)|C$ is not stationary so that in defining \bar{d} in this context we are of course not restricted to stationary joinings.

In order to verify this kind of hypothesis one usually uses the FD property of a Bernoulli flow (g_t, Q) . Here is a finite version of this FD property which is useful. Given $\epsilon > 0$, there is a K and a $\delta > 0$ so that if N is large enough and if λ_N is any distribution on N -blocks satisfying

(1) the distribution of K -blocks (averaged over the different starting points) is within δ of $\text{dist } \bigvee_1^K g^i Q$,

(2)

$$\left| \frac{\text{entropy of } \lambda_N}{N} - h(g_t, Q) \right| \leq \delta,$$

then there is a joining between $\bigvee_1^N g^i Q$ and λ_N which is on average within ϵ of the identity, i.e.,

$$\bar{d}(\lambda_N, \text{dist } \bigvee_1^N g^i Q) < \epsilon.$$

The sharp version of the basic imbedding theorem says that if (*) holds for (g_t, Q) and (f_t, X, P) , relative to the factor \mathcal{A} , where (g_t, Q) is Bernoulli, and if (f_t, X) has enough entropy relative to \mathcal{A} , i.e., if

$$(**) \quad h(f_t, X) \geq h(f_t, X, \mathcal{A}) + h(g_t, Q),$$

then there is a partition \hat{P} within ϵ of P such that

(i) $\bigvee_{-\infty}^{\infty} f_t \hat{P}$ is independent of \mathcal{A} ,

(ii) (f_t, \hat{P}) is congruent to (g_t, Q) .

The relative isomorphism theorem says that if in addition (f_t, X) is Bernoulli relative to \mathcal{A} and if equality holds in (**) then in addition we can arrange that

(iii) $(\bigvee_{-\infty}^{\infty} f_t \hat{P}) \vee \mathcal{A}$ is the whole σ -algebra of (f_t, X) .

We will need a variant of this when in (**) equality doesn't hold. It is a generalization of Theorem 3.7.2.

Theorem 5.7.1. *Suppose that (f_t, X) is relatively Bernoulli with respect to a factor \mathcal{A} . Let (g_t, Q) be a Bernoulli flow and P a partition of X such that (f_t, P) is ϵ -close in \bar{d} to (g_t, Q) relative to \mathcal{A} . If*

$$h(f_t, X) - h(f_t, \mathcal{A}) > h(g_t, Q)$$

we can find a partition \hat{P} and an invariant α -algebra $\hat{\mathcal{A}}$ satisfying

(i) $d(P, \hat{P}) < \epsilon$,

- (ii) (f_t, \widehat{P}) is congruent to (g_t, Q) ,
- (iii) $\bigvee_{-\infty}^{\infty} f_t \widehat{P}, \mathcal{A}, \widehat{\mathcal{A}}$ are mutually independent and span the whole σ -algebra of (f_t, X) .

Note that we do not assume that \mathcal{A} has zero entropy or is Bernoulli.

For finite entropy the above theorem is Corollary 1 of [Fi 1]. The case where \mathcal{A} has infinite entropy can be proved in the same way but we see no quick reduction to the finite case.

APPENDIX BY DAVID FRIED, BOSTON UNIVERSITY³¹

The interior of Bernoulli. Let Λ be a C^2 hyperbolic attractor, $\Lambda = \bigcap_{t \geq 0} f_t U$ where $f_t: U \rightarrow U$ is a C^2 semiflow transverse to ∂U , U is a connected C^∞ manifold. Assume Λ is nontrivial, i.e., consists of more than one orbit. We say Λ is *circular* if there is a map $\theta: \Lambda \rightarrow \mathbf{R}/\mathbf{Z}$ such that $X\theta = (d/dt)\theta(f_t x) > 0$ for all $x \in \Lambda$. If $X\theta$ is constant for some such θ , then we say Λ is a *suspension*.

Bowen has shown that Λ is a suspension if and only if Λ is not topologically mixing [B 1]. In turn Λ is topologically mixing if and only if the SBR measure is Bernoulli [Rat]. Thus if Λ is not circular then all C^1 -near C^2 semiflows $g_t: U \rightarrow U$ define Bernoulli attractors $\Lambda_g = \bigcap_{t \geq 0} g_t U$ (Λ_g is topologically conjugate to Λ hence not a suspension). We will examine a circular Λ and determine whether it is a C^1 limit of C^2 suspensions. This will give a converse to Theorem 2 of Ornstein and Weiss and identify precisely which Λ are SSS.

Proposition A. *For a nontrivial C^2 hyperbolic attractor Λ the following are equivalent:*

- (a) f is the C^1 limit of C^2 suspensions.
- (b) f is the C^2 limit of reparametrizations that are suspensions.
- (c) The distribution $E^U \oplus E^S$ on Λ is integrable.
- (d) There is a cohomology class $\xi \in H^1(U; \mathbf{R})$ such that the period $l(\gamma)$ of any closed orbit $\gamma \subset \Lambda$ and its homology class $[\gamma] \in H_1(U; \mathbf{R})$ satisfy

$$\xi[\gamma] = l(\gamma).$$

³¹David Fried was partially supported by the National Science Foundation and the Sloan Foundation.

Proof.

(b) \Rightarrow (a): Trivial

(a) \Rightarrow (d): Suppose ϕ is a C^2 suspension flow C^1 near to f . Then there is a θ, c with $(d/dt)\theta(\phi_t x) = c > 0$. The closed 1-form $(1/c)d\theta = \omega$ satisfies $\omega(X) = 1$, where X is the infinitesimal generator of f . Choose $\phi_i, \theta_i, c_i, X_i, \omega_i$ so that $\phi_i \rightarrow f$ as $i \rightarrow \infty$ in the C^1 topology. By the Lemma below, the sequence of cohomology classes $[\omega_i]$ stays bounded as $i \rightarrow \infty$ so we may suppose it converges to a class $\xi \in H^1(M; \mathbf{R})$. Let γ_i denote the closed orbit of ϕ_i corresponding to γ under the topological conjugacy of ϕ_i to f provided by Smale's Ω -stability theorem. Then

$$\xi[\gamma] = \lim_{i \rightarrow \infty} \int_{\gamma_i} \omega_i = \lim_{i \rightarrow \infty} l(\gamma_i) = l(\gamma).$$

(d) \rightarrow (c): Let ω be a closed 1-form that represents ξ . For any closed orbit ω ,

$$\int_{\gamma} \omega = \xi[\gamma] = l(\gamma)$$

so the function $\omega(X) - 1$ has integral zero over every such γ . By Livshitz's theorem [Liv], there is a continuous function g on Λ differentiable along the flow such that

$$\omega(X) - 1 = Xg.$$

For nearby points $x, y \in \Lambda$ define

$$c(x, y) = \int_x^y \omega + g(x) - g(y),$$

where the integral is taken over any short path in U from x to y (two such paths are homotopic if x, y are close enough, relative to some Riemannian metric on U). Then

$$\begin{aligned} c(x, y) + c(y, x) &= 0, \\ c(x, y) + c(y, z) &= c(x, z), \end{aligned}$$

so c defines a continuous Alexander 1-cocycle on Λ . Note that for $x \in \Lambda$

$$c(x, f_t x) = \int_0^t (\omega(X(f_s x)) - Xg(f_s x)) ds = t.$$

Thus if $y, x \in \Lambda$ and $y \in W_\epsilon^s(x)$ (respectively, $y \in W_\epsilon^u(x)$) then

$$c(x, y) = c(f_t x, f_t y) \rightarrow 0$$

as $|t| \rightarrow \infty$ with $t > 0$ (respectively, $t < 0$) and so $c(x, y) = 0$. For $x_0 \in \Lambda$ one sees that the locally defined function $h(y) = c(y', x_0)$, $y \in W_\epsilon^s(y')$, $y' \in \Lambda$, is constant on local stable and unstable manifolds near x_0 and satisfies

$$h(y) - h(f_t y) = t$$

for $y \in U$, y near x_0 . Thus $E^U \oplus E^S$ is integrable.

(c) \Rightarrow (b): On a U -neighborhood of $x_0 \in \Lambda$ we have a continuous function h constant on local stable and unstable manifolds but nowhere constant. One can arrange that $h(y)$ measures the flow time between x_0 and the codimension one manifold M_y containing y formed by taking the point $y' \in \Lambda$ with $y \in W_\epsilon^s(y')$ and setting

$$M_y = \bigcup_z W_\epsilon^s(z), \quad z \in W_\epsilon^u(y').$$

Then $h(y) - h(f_t y) = t$. We define $c(x, y) = h(x) - h(y)$ so that c is a continuous Alexander 1-cocycle on a neighborhood V of Λ and

$$c(x, f_t y) = c(x, y) + t$$

for x near y , $|t|$ small. Define $\tilde{c}(x, y)$ for all pairs $x, y \in U$, x near y , by

$$\tilde{c}(x, y) = c(\phi_T x, \phi_T y)$$

with $T > 0$ chosen so $\phi_T U \subset V$. Then \tilde{c} is also a 1-cocycle and

$$\tilde{c}(x, f_t y) = \tilde{c}(x, y) + t.$$

Let $\xi \in H^1(U; \mathbf{R})$ be the cohomology class of \tilde{c} . Then we choose a class ξ_0 near ξ with rational periods and represent $\xi_0 - \xi$ by a C^∞ small closed 1-form ω_0 . The 1-cocycle

$$\alpha(x, y) = \tilde{c}(x, y) + \int_x^y \omega_0$$

(where the integral is taken over a short path) represents ξ_0 . Let $Y_0 = X/(1 + \omega_0(X))$ and define a C^2 flow ϕ_t by integrating Y_0 . Then ϕ is C^2 close to f and we have

$$\alpha(x, \phi_t y) = \alpha(x, y) + t.$$

Define the Abel map

$$\theta: U \rightarrow \mathbf{R}/\Gamma, \quad \theta(x) = \int_b^x \alpha,$$

where b is a basepoint in U , $x \in U$, $\Gamma \subset \mathbf{Q}$ denotes the set of periods of ξ_0 and where the integral $\int_b^x \alpha$ denotes the finite sum

$$\alpha(b, x_1) + \alpha(x_1, x_2) + \cdots + \alpha(x_n, x)$$

over some sequence $x_0 = b, x_1, \dots, x_n, x_{n+1} = x$ of closely spaced points. Changing the choice of sequence affects the sum by a period of ξ_0 so θ is well defined. As ξ_0 has rational periods and as $H_1(U; \mathbf{Z})$ is finitely generated (U is compact) Γ is discrete. Thus we may identify \mathbf{R}/Γ with \mathbf{R}/\mathbf{Z} by rescaling. We see that

$$\frac{d}{dt} \theta(\phi_t x) = 1$$

so that ϕ is a suspension.

Now we turn to the question of constructing Bernoulli attractors that are the C^1 limit of non-Bernoulli C^2 flows. As noted above, the flows f_t must be circular for this to hold. We let \mathcal{A} denote the C^2 hyperbolic attractors on U and $\mathcal{B} \subset \mathcal{A}$ the Bernoulli attractors, each with the C^1 topology.

Proposition B. \mathcal{B} is open in $\mathcal{A} \Leftrightarrow$ either \mathcal{A} contains no circular flows or $\dim H^1(U; \mathbf{R}) = 1$.

Proof. \Leftarrow) If \mathcal{A} has no circular flows then $\mathcal{B} = \mathcal{A}$. Assume $\dim H^1(U; \mathbf{R}) = 1$ and $\phi_i \rightarrow f, \phi_i \in \mathcal{A} - \mathcal{B}, f \in \mathcal{A}$. Then take ξ as in the proof of (a) \Rightarrow (d) of Proposition A and take the cocycle c as in the proof of (d) \Rightarrow (c). Then since $\dim H^1(U; \mathbf{R}) = 1$, $H_1(U; \mathbf{Z})$ has rank 1 and so the group $\Gamma \subset \mathbf{R}$ of periods of ξ is discrete. Thus the proof of (c) \Rightarrow (b) (with $\xi = \xi_0, \omega = 0, \alpha = \tilde{c}$) shows that f is a suspension, i.e., $f \notin \mathcal{B}$.

\Rightarrow) Assume $f \in \mathcal{A}$ is a circular flow and $\dim H^1(U; \mathbf{R}) = \beta > 1$. After reparametrizing we may assume that f is a suspension. Near f we will find $\psi \in \overline{\mathcal{B} \cap \mathcal{A} - \mathcal{B}}$.

Let $\xi \in H^1(U; \mathbf{R})$ be the cohomology class defined by the $\theta: U \rightarrow \mathbf{R}/\mathbf{Z}$ arising in the definition of suspension. Then near ξ we choose ξ_1 so that the group $\Gamma_1 \subset \mathbf{R}$ of periods of ξ is not discrete (this is possible since $\beta > 1$). Then we represent $\xi_1 - \xi$ by a small C^∞ closed 1-form ω_1 and use

$$Y_1 = X / (1 + \omega_1(X))$$

to define a C^2 flow ψ near f . Proposition A shows $\psi \in \overline{\mathcal{A} - \mathcal{B}}$ is not in the interior of Bernoulli. To see $\psi \in \mathcal{B}$ we use

Lemma. *The homology classes of closed orbits of any $\phi \in \mathcal{A}$ span $H_1(M; \mathbf{R})$.*

This lemma and our choice of ξ_1 implies the $\xi_1(\gamma)$, γ a closed orbit of ψ , are not all multiples of some positive constant and so ψ is not a suspension.

To prove the lemma one argues just as in the transitive Anosov case [Fr]. If the orbits didn't span there would be a surjective homomorphism $\alpha: \pi_1 U \rightarrow \mathbf{Z}$ with $\alpha(\omega) = 0$ for all γ . Take $N > 0$ and let U_N be the N -sheeted connected cyclic cover of U corresponding to

$$\pi_1 U \xrightarrow{\alpha} \mathbf{Z} \rightarrow \mathbf{Z}/N\mathbf{Z}$$

with the lifted flow ϕ^N . Then $\Lambda^N = \bigcap_{t \geq 0} \phi_t^N U_N$ is a connected hyperbolic attractor. But if N is large enough Λ^N does not contain a dense orbit (for this step we refer to Theorem H in [Fr]) which is a contradiction.

Thus we have characterized the interior of \mathcal{B} . For the sake of completeness we mention that examples of circular hyperbolic attractors with $\beta > 1$ abound: for instance one can take an Anosov suspension flow in dimension 3 and DA $\beta + 1$ closed orbits (i.e., replace them by sources) to get an attractor with $\dim H^1(U; \mathbf{R}) = \beta$. For this choice of U , \mathcal{B} is not open.

For an Anosov suspension with $\beta = 1$ the openness of \mathcal{B} was shown by Bowen [B 2] (no Anosov suspension with $\beta > 1$ is known).

BIBLIOGRAPHY

- [AdShSmo] R. Adler, P. C. Shields, and M. Smorodinsky, *Irreducible Markov shifts*, Ann. Math. Stat. **43** (1972), 1027–1029.
- [AdW 1] R. Adler and B. Weiss, *Entropy a complete invariant for automorphisms of the torus*, Proc. Nat. Acad. Sci. U.S.A. **57** (1967), 1573–1576.
- [AdW 2] —, *Similarity of automorphism of the torus*, Mem. Amer. Math. Soc. **98** (1970).
- [AndP] A. Andronov and L. Pontriagin, *Coarse systems*, Dokl. Akad. Nauk SSSR **14** (1937).
- [Ano 1] D. V. Anosov, *Geodesic flows on closed Riemannian manifolds with negative curvature*, Proc. Steklov Inst. Math. **90** (1967).
- [Ano 2] —, *Structurally stable systems*, Proc. Steklov Inst. Math. **169** (1985), 59–93.
- [AnoSi] D. Anosov and Ya. G. Sinai, *Some smooth ergodic systems*, Russian Math. Surveys **22** (5) (1967), 103–167.
- [ArOW] P. Arnoux, D. Ornstein, and B. Weiss, *Cutting and stacking, interval exchanges and geometric models*, Israel J. Math. **50** (1985), 160–168.

- [Az] R. Azencotte, *Diffeomorphismes d'Anosov et schémas de Bernoulli*, C. R. Acad. Sci. Paris **270** (1970), A1105–A1107.
- [B 1] R. Bowen, *Periodic orbits for flows*, Amer. J. Math. **94** (1972), 1–30.
- [B 2] —, *Mixing Anosov flows*, Topology **15** (1976), 77–79.
- [BrFeKato] M. Brin, J. Feldman and A. Katok, *Bernoulli diffeomorphisms and group extensions of dynamical systems with non zero characteristic exponents*, Ann. of Math. **113** (1981), 159–179.
- [BrG] M. Brin and M. Gromov, *On the ergodicity of frame flows*, Invent. Math. **60** (1980), 1–7.
- [BrKato] M. Brin and A. Katok, *On local entropy*, Geometric Dynamics, Lecture Notes in Math., vol. 1007, Springer-Verlag, Berlin-New York, 1982, pp. 30–38.
- [BRue] R. Bowen and D. Ruelle, *The ergodic theorem for Axiom A flows*, Invent. Math. **29** (1975), 181–202.
- [Bun] L. Bunimovitch, *On a class of special flows*, Math. USSR-Izv. **8** (1974), 219–232.
- [BurGe] K. Burns and M. Gerber, *Real analytic geodesic flows on S^2* , Ergodic Theory Dynamic Systems **9** (1989), 27–45.
- [C] C. Conley, *Hyperbolic invariant sets and shift automorphisms*, Dynamical Systems Theorem and Applications, Lecture Notes in Phys., vol. 38, 1975, pp. 539–549.
- [E] K. Eloranta, *Alpha-congruence for billiards and Markov processes*, Ph.D. Thesis, Stanford, 1988.
- [Fe 1] J. Feldman, *New K -automorphisms and a problem of Kakutani* Israel J. Math. **24** (1976), 16–38.
- [Fe 2] —, *r -Entropy, equipartition, and Ornstein's isomorphism theorem in \mathbf{R}^n* , Israel J. Math. **36** (1980), 321–345.
- [FeO] J. Feldman and D. S. Ornstein, *Semi-rigidity of horocycle flows over compact surfaces of variable negative curvature*, Ergodic Theory Dynamic Systems **7** (1987), 49–72.
- [Fi 1] A. Fieldsteel, *The relative isomorphism theorem for Bernoulli flows*, Israel J. Math. **40** (1981), 197–215.
- [Fi 2] —, *Stability of the weak Pinsker property for flows*, Ergodic Theory Dynamical Systems (to appear).
- [Fr] D. Fried, *Flow equivalence, hyperbolic systems and a new zeta function for flows*, Comment. Math. Helv. **57** (1982), 237–259.
- [Fu] H. Furstenberg, *Disjointness in ergodic theory, minimal sets and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1–49.
- [GO] G. Gallavotti and D. S. Ornstein, *Billiards and Bernoulli schemes*, Comm. Math. Phys. **38** (1974), 83–101.
- [Ka 1] A. Katok, *Smooth non-Bernoulli K -automorphisms*, Invent. Math. **61** (1980), 291–300.
- [Ka 2] —, *Invariant cone families and stochastic properties of smooth dynamical systems*, Calif. Inst. Tech., preprint.
- [KatoStr] A. Katok, J.-M. Strelcyn, F. Ledrappier, and F. Przytycki, *Invariant manifolds, entropy and billiards, smooth maps with singularities*, Lecture Notes in Math., vol. 1222, Springer-Verlag, Berlin-New York, 1986.

- [Katz] Y. Katznelson, *Ergodic automorphisms of T^n are Bernoulli shifts*, Israel J. Math. **10** (1971), 186–195.
- [KeSmo] M. Keane and M. Smorodinsky, *Bernoulli shifts on the same entropy are finitarily isomorphic*, Ann. of Math. **109** (1979), 397–406.
- [Ki 1] Y. Kifer, *General random perturbations of hyperbolic and expanding transformations*, J. Analyse Math. **47** (1986) 111–150.
- [Ki 2] ———, *Ergodic theory of random transformations*, Birkhauser, Boston, 1986.
- [Ki 3] ———, *Random perturbations of dynamical systems*, Birkhauser, Boston, 1988.
- [Ko 1] A. N. Kolmogorov, *A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces*, Dokl. Akad. Nauk SSSR **119** (1958), 861–864. (Russian) MR **21** # 2035a.
- [Ko 2] ———, *Entropy per unit time as a metric invariant of automorphisms*, Dokl. Akad. Nauk SSSR **124** (1959), 754–755. (Russian) MR **2** # 2035b.
- [Kre] U. Krengel, *Ergodic theorems*, W. de Gruyter, New York, 1985.
- [Kri] W. Krieger, *On entropy and generators of measure-preserving transformations*, Trans. Amer. Math. Soc. **149** (1970), 453–464.
- [Led] F. Ledrappier, *Propriétés ergodiques des mesures de Sinai*, Publ. Math. Institut des Hautes Etudes Scientifiques **59** (1984), 163–188.
- [Lig] T. M. Liggett, *Interacting particle systems*, Springer-Verlag, Berlin, 1985.
- [Lin] D. Lind, *The structure of skew products with ergodic group actions*, Israel J. Math. **28** (1977), 205–248.
- [Liv] A. N. Livshits, *Homology properties of Y -systems*, Math. Notes **10** (1971), 758–763.
- [McCSh] R. McCabe and P. C. Shields, *A class of Markov shifts that are Bernoulli*, Adv. in Math. **6** (1971), 323–328.
- [Ma 1] R. Mané, *Ergodic theory and differential dynamics*, Springer-Verlag, Berlin, 1987.
- [Ma 2] ———, *A proof of the c' stability conjecture*, Pub. Math. Institut des Hautes Etudes Scientifiques **66** (1981), 161–210.
- [MiTh 1] G. Miles and R. K. Thomas, *The breakdown of automorphisms of compact topological groups*, Studies in Probability, Adv. Math. Suppl. Studies **2** (1978), 207–218.
- [MiTh 2] ———, *On the polynormal uniformity of translations of the n -torus*, Studies in Probability, Adv. Math. Suppl. Studies **2** (1978), 219–229.
- [MiTh 3] ———, *Generalized torus automorphisms are Bernoullian*, Studies in Probability, Adv. Math. Suppl. Studies **2** (1978), 231–249.
- [O 1] D. S. Ornstein, *Ergodic theory, randomness and dynamical systems*, Yale Univ. Press, New Haven, 1974.
- [O 2] ———, *Bernoulli shifts with the same entropy are isomorphic*, Adv. in Math. **4** (1970), 337–352.
- [O 3] ———, *Two Bernoulli shifts with infinite entropy are isomorphic*, Adv. in Math. **5** (1970), 339–348.
- [O 4] ———, *Factors of Bernoulli shifts are Bernoulli shifts*, Adv. in Math. **5** (1970), 349–364.
- [O 5] ———, *Imbedding Bernoulli shifts in flows*, Contributions to Ergodic Theory and Probability, Lecture Notes in Math., Springer-Verlag, Berlin, 1970, pp. 178–218.

- [O 6] —, *The isomorphism for Bernoulli flows*, Adv. in Math. **10** (1973), 124–142.
- [O 7] —, *An example of a Kolmogorov automorphism that is not a Bernoulli shift*, Adv. in Math. **10** (1973), 49–62.
- [O 8] —, *A K -automorphism with no square root and Pinsker's conjecture*, Adv. in Math. **10** (1973), 89–102.
- [O 9] —, *A mixing transformation for which Pinsker's conjecture fails*, Adv. in Math. **10** (1973), 103–123.
- [O 10] —, *Factors of Bernoulli shifts*, Israel J. Math. **21** (1975), 145–153.
- [OShi] D. S. Ornstein and P. C. Shields, *Markov shifts of kernel type are Bernoulli*, Adv. in Math. **10** (1973), 143–146.
- [OSu] D. S. Ornstein and L. Sucheson, *An operator theorem on L_1 convergence to zero with applications to Markov kernels*, Ann. of Math. Statistics **5** (1970), 1631–1639.
- [Os] V. I. Oseledec, *A multiplicative ergodic theorem*, Trans. Moscow Math. Soc. **19** (1968), 197–231.
- [Otl] J.-P. Otal, *Le spectre marqué des longueurs des surfaces à courbure négative*, preprint.
- [OW 1] D. S. Ornstein and B. Weiss, *Entropy and isomorphism theorems for actions of amenable groups*, J. Analyse Math. **48** (1987), 1–141.
- [OW 2] —, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94–104.
- [OW 3] —, *Geodesic flows are Bernoullian*, Israel J. Math. **14** (1973), 184–198.
- [OW 4] —, *How sampling reveals a process*, Ann. Probab. **18** (1990).
- [ORudW] D. Ornstein, D. J. Rudolph and B. Weiss, *Equivalence of measure preserving transformations*, Mem. Amer. Math. Soc. **262** (1982).
- [Pa] K. Park, *A flow built under a step function with a multi-step Markov partition on a base*, Ph.D. Thesis, Stanford, 1981.
- [Pe] Ya. B. Pesin, *Characteristic Lyapunov exponents and smooth ergodic theory*, Russian Math. Surveys **32** (4) (1977), 55–114.
- [Pl] J. Plante, *Anosov flows*, Amer. J. Math. **94** (1972), 729–754.
- [Pu] C. Pugh, *The $C^{l+\alpha}$ -hypothesis in Pesin theory*, Publ. Math. IHES **59** (1984), 143–161.
- [PuShu] C. Pugh and M. Shub, *The Ω -stability theorem for flows*, Invent. Math. **11** (1970), 150–158.
- [Rag] S. Raghunathan, *A proof of Oseledec's multiplicative ergodic theorem*, Israel J. Math. **32** (1979), 356–362.
- [Rat 1] M. Ratner, *Anosov flows with Gibbs measures are also Bernoullian*, Israel J. Math. **17** (1974), 380–391.
- [Rat 2] —, *Bernoulli flows over maps of the interval*, Israel J. Math. **31** (1978), 298–314.
- [RoSi] V. Rohlin and Ya. Sinai, *Construction and properties of invariant measurable partitions*, Dokl. Akad. Nauk SSSR **141** (1961), 1038–1041.
- [Rud 1] D. J. Rudolph, *A two-valued step coding for ergodic flows*, Math. Z. **150** (1976), 201–220.
- [Rud 2] —, *Restricted orbit equivalence*, Mem. Amer. Math. Soc. **323** (1985).

- [Rud 3] —, *If a finite extension of a Bernoulli shift has no rotation factor it is Bernoulli*, Israel J. Math. **30** (1978), 193–206.
- [Rud 4] —, *Classifying the compact extensions of a Bernoulli shift*, J. Analyse Math. **34** (1978), 36–59.
- [Rud 5] —, *An isomorphism theory for Bernoulli free z -skew-compact group actions*, Adv. Math. **47** (1983), 241–257.
- [Rud 6] —, *An example of a measure-preserving map with minimal self-joinings, and applications*, J. Analyse Math. **35** (1979), 97–122.
- [Rud 7] —, *Asymptotically Brownian co-cycles give non-loosely Bernoulli K -automorphisms*, Invent. Math. **91** (1988), 105–128.
- [RudSch] D. J. Rudolph and G. Schwarz, *The limits in \bar{d} of multi-step Markov chains*, Israel J. Math. **28** (1977), 103–109.
- [ShiTho] P. C. Shields and J.-P. Thouvenot, *Entropy zero \times Bernoulli processes are closed in the \bar{d} -metric*, Ann. Probab. **3** (1975), 732–736.
- [Shu] M. Shub, *Global stability of dynamical systems*, Springer-Verlag, Berlin, 1987.
- [Si 1] Ya. G. Sinai, *On the notion of entropy of a dynamical system*, Dokl. Akad. Nauk SSSR **124** (1959), 768–771.
- [Si 2] —, *A weak isomorphism of transformations with an invariant measure*, Dokl. Akad. Nauk SSSR **147** (1962), 797–800. (Soviet Math. Dokl. **3** (1962), 1725–1729.) MR **28** # 5164a: **28** # 1247.
- [Si 3] —, *Geodesic flows on compact surfaces of negative curvature*, Dokl. Akad. Nauk SSSR **136** (3) (1961), 549–552.
- [Si 4] —, *Dynamical systems with elastic reflections*, Uspekhi Mat. Nauk **27** (1972), 137.
- [Si 5] —, *Gibbs measures in ergodic theory*, Russian Math. Surveys **166** (1972), 21–69.
- [Si 6] —, *Markovian partitions and U -diffeomorphisms*, Functional Anal. Appl. **2** (1968), 64–89.
- [Si 7] —, ed., *Dynamical Systems II*, Encyclopaedia of Math. Sci., vol. 2, Springer-Verlag, Berlin, 1988.
- [Sm] S. Smale, *Differential dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 744–817.
- [Smo] M. Smorodinsky, *β -automorphisms are Bernoulli shifts*, Acta Math. Acad. Sci. Hungar. **24** (1973), 3–4.
- [Ste] J. Steif, *The ergodic structure of interacting particle systems*, Ph.D. Thesis, Stanford, 1988.
- [Sz] W. Szlenk, *An introduction to the theory of smooth dynamical systems*, John Wiley, New York, 1984.
- [Tho 1] J.-P. Thouvenot, *Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli*, Israel J. Math. **21** (1975), 177–207.
- [Tho 2] —, *Une classe de systèmes pour lesquels la conjecture de Pinsker est vraie*, Israel J. Math. **21** (1975), 208–214.
- [Tr] S. Troubetzkoy, *Extreme instability of the horocycle flow*, Ph.D. Thesis, Stanford, 1987.
- [Wa] P. Walters, *Anosov diffeomorphisms are topologically stable*, Topology **7** (1970), 71–78.

[Wo 1] M. Wojtkowski, *A system of one-dimensional balls with gravity*, preprint.

[Wo 2] ———, *A system of one-dimensional balls in an external field, II*, preprint.

STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305-2125

HEBREW UNIVERSITY, JERUSALEM, ISRAEL 91905