

Strange Attractors, Chaotic Behavior, and Information Flow *

Robert Shaw

Physics Department, University of California, Santa Cruz, California 95064, USA

Z. Naturforsch. **36a**, 80–112 (1981); received October 15, 1980

Simple system equations displaying turbulent behavior are reviewed in the light of information theory. It is argued that a physical implementation of such equations is capable of acting as an information source, bringing into the macroscopic variables information not implicit in initial conditions. The average rate of information production $\bar{\lambda}$ is a system state function, and is given for simple cases by a “Liapunov characteristic exponent”, developed by Oseledec. The transition of a system from laminar to turbulent behavior is understandable in terms of the change of $\bar{\lambda}$ from negative to positive, corresponding to the change of the system from an information sink to a source. The new information of turbulent systems precludes predictability past a certain time; when new information accumulates to displace the initial data, the system is undetermined.

The observed geometry of strange attractors is seen to arise naturally from a rule allowing joining but not splitting of trajectories in phase space. The phenomenology of strange attractors in three dimensions is discussed, and a basis for their classification suggested. A comment is made on the commonplace occurrence of information producing systems in the real world, and on their possible relation to $1/f$ noise.

Table of Contents

I. Introduction
II. Relevance of Attractors to Dissipative Systems
III. Informational Aspects of Nonconservative Flows
IV. Maps on the Unit Interval
V. Analogy with Phase-Transitions, Overview
VI. Relation of One-Dimensional Maps to Dynamic Flows
VII. Calculation of $\bar{\lambda}$ for the Lorenz Attractor
VIII. Experimental Verification
IX. Determinism and Reversibility in Phase Space
X. Steps Toward a Classification of Strange Attractors
XI. Relation Between 3D Flows, 2D Maps, 1D Maps
XII. Strange Attractor Phenomenology
XIII. Informational Aspects of Turbulence
XIV. $1/f$ Noise, Concluding Speculations
Acknowledgements
Appendix I Dynamical System as a Noise Amplifier
Appendix II Evolution of Probability Densities
References

I. Introduction

Appearing recently in fields as disparate as meteorology, physiology, ecology and mathematics have been descriptions of an odd class of mathematical objects, those which display so-called “chaotic” behavior.

Each of these objects specifies an iteration procedure; given an initial number or set of numbers, a string of new numbers can be generated. Such a scheme can model the evolution of some system

from given initial data. The remarkable fact is that despite the deterministic nature of the governing equations, the string of numbers so produced is in any practical sense unpredictable.

Included in this class are the sets of first-order equations of Lorenz [1], Rössler [3], and others, the simple difference equations treated by May [4], Guckenheimer et al. [5], and others, and mappings of the plane onto itself discussed by Hénon [6], and Stein and Ulam [7], among others. The qualitative behavior of the solutions to these various equations is similar, notably their unpredictable nature, and their extreme sensitivity to initial conditions. Particularly striking is the change in behavior which several of these systems show as a parameter is

* Supported by a grant from the National Science Foundation.

Reprint requests to Robert Shaw, Physics Department, University of California, Santa Cruz, California 95064.

0340-4811 / 81 / 0100-0080 \$ 01.00/0. — Please order a reprint rather than making your own copy.



Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung“) beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition “no derivative works”). This is to allow reuse in the area of future scientific usage.

increased. Often we see the solutions first converge from any initial condition to a static value, then display periodic motion, and then, at some critical value of this parameter, make a transition into a chaotic regime. Indeed, the physical systems of convection and fluid flow often show just this behavior as a function of R , the dimensionless Rayleigh or Reynolds number. This similarity has inspired new hope for a mathematical description and better understanding of the elusive nature of turbulence [8]. As solid links have yet to be forged between equations of motion and “chaotic” solutions, these similarities remain in the realm of “mathematical metaphors”.

Still we might ask, what is the common thread which is responsible for the same qualitative behavior in these systems? This is the question which this paper addresses. We will attempt to answer it relying on a few simple physical notions, notably those of an uncertainty length, and information flow, to guide us past the large areas of the subject which have not yielded to attempts at mathematical analysis.

Classical mechanics successfully treats many systems ignoring completely the fact that any macroscopic system consists of an enormous number of particles, each of which has at least a limited freedom to move independently. The energy contained in a system seems to be able to be divided into two categories:

- 1) that contained in the “macroscales”, the relatively few large-scale correlated motions of which we claim classically to have *complete knowledge*, and
- 2) that contained in the heat degrees of freedom, or “microscales”, the vast multitude of small-scale motions of which we are *completely ignorant*, except for the temperature T , a single number measuring the average energy per degree of freedom.

The macroscales do participate in this thermal motion, and thus their energies are uncertain by roughly kT , but this is small compared to the amount of energy per degree of freedom they usually contain, and is successfully neglected. Even though the total energy in the microscales may be far larger than that contained in the macroscales,

it has proved possible to predict essentially completely the future behavior of the macroscopic degrees of freedom using only the complete initial knowledge of these same scales: one does not have to know the temperature to do a classical mechanics problem.

It is the thesis of this paper, however, that there exist many nonconservative systems for which this implicit separation breaks down; instead of our complete knowledge or ignorance being confined to the macro- or microscales, there may be an active flow of information between them, one way or the other, and a more careful auditing of information is required for an understanding of their behavior. Information theory, as developed by Shannon, provides a framework for the quantification of this notion. We construct a parameter $\bar{\lambda}$, the units of which are “bits” of information per unit time, which measures the average amount of information flowing into or out of the macroscopic scales of which we have, again classically, complete knowledge. This is an invariant quantity, essentially identical, in the simple cases we will study, to the “Liapunov characteristic exponent” developed by Oseledec. Information entering the macroscales from the heat bath is reflected in a positive value of this parameter, and this provides a basis for understanding the unpredictable nature of dissipative systems such as turbulent flow.

If a system is abstractly represented by a “flow” in “phase space”, this transfer of information is seen to be accomplished by parts of the flow which have definite geometrical forms, of which “strange attractors” are examples. We claim that the many examples of systems exhibiting “chaotic” behavior which are appearing and will no doubt continue to appear, are different guises of what are actually a limited number of basic forms. Furthermore, these forms arise naturally in a geometry where irreversibility is taken as a *postulate*, in contrast to the reversibility implicit in the usual description of systems in phase space. After a brief exposition on the relevance of attractors in phase space to dissipative systems, and of the application of information theory to the description of these systems, we will examine the simplest of these forms, and their relation to the parameter $\bar{\lambda}$ mentioned above. We will conclude with a few remarks about turbulence, and a speculation on the origin of $1/f$ noise.

II. Relevance of Attractors to Dissipative Systems

Dynamical systems are conveniently represented in “phase space”, a space with as many dimensions as the system has total degrees of freedom. Thus, the complete state of a system at any time is specified by a point in phase space, and the motion of a particular system as time passes generates an orbit through the space. The time evolution of the set of orbits originating from all possible initial conditions generates a “flow” in this space, governed by a set of n first-order differential equations,

$$dX_i/dt = F_i(X_1, X_2, \dots, X_n), \quad i = 1, \dots, n \quad (1)$$

where n is the dimensionality of the space.

If the functions F_i are suitably “well-behaved”, the extensive mathematical formalism treating vector fields can be applied, and various theorems invoked, notable among them being that orbits never cross, thus the uniqueness of a given orbit arising from a given initial condition is guaranteed. *Implicit* in the use of this formalism for the description of dynamical systems is the *assumption of reversibility*, i.e., no information is lost as time passes.

A “system” is a conceptual boundary between some part of reality and the rest of the world, the localization implied by this boundary means that we will be interested in *compact* flows in the phase space. A law of Nature, such as Conservation of Energy, will reduce the dimensionality of the phase space available to a given system, in general there will be a reduction by one dimension for each constraint equation.

In an “open system”, energy is allowed to cross in and out over this conceptual boundary, typically entering the macroscopic variables in the form of “work”, and leaving in the form of “heat”, with a net entropy increase. In these systems, orbits are not constrained to an energy surface of one less dimension. Nevertheless, we feel that a “law of Nature” is operating, we notice that although we can prepare the system in a variety of initial states, after a time only some of them are visited by the system, in contrast to the conservative situation. For instance, in a convection problem we could in principle start the system off in a checkerboard pattern of hot and cold fluid, but we know that we will *not* see this state subsequently. So in dissipative systems, we expect to see the set of all orbit converge as time passes to a compact region of phase

space and remain there. This is just the mathematical definition of an “attractor”.

A simple example is the damped harmonic oscillator:

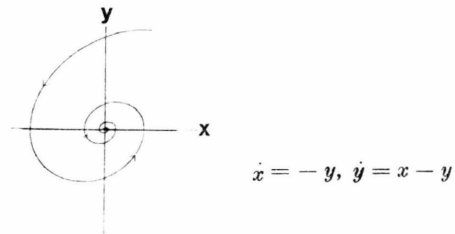


Fig. 1

Here the set of all initial conditions form a *plane*, and, as the oscillator ultimately stops, all solutions tend toward a *fixed point*, the simplest attractor. The dimensionality of the solutions is reduced from two to zero.

Another example is the Van der Pol oscillator:

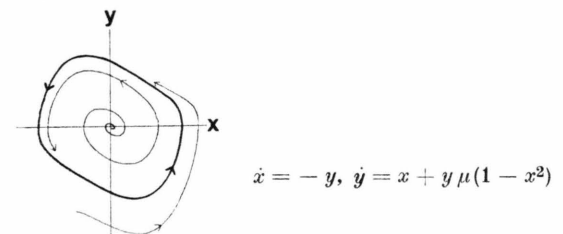


Fig. 2

In this system all solutions tend toward a stable periodic orbit, or *limit cycle*, another example of an attractor. Here, the dimensionality of the solution set is reduced from two to one.

This behavior is not found in classical conservative systems, where Conservation of Energy and Liouville's Theorem hold. Conservation of volume under the flow in phase space means the dimensionality of the solutions stays fixed. Any state visited by the system will be revisited in a characteristic “Poincaré recurrence time”. The phenomenon of irreversible contraction onto an attractor does not occur, there are no “transients” in conservative systems.

Rigorously speaking, the mathematical solutions to these examples never actually arrive *at* their attractor, but only approach it exponentially. So formally the solutions maintain their identity with respect to their individual initial conditions. However, in any practical sense the dimensionality is

reduced in *finite time*, and whole sets of solutions become identified.

As soon as any two trajectories approach within some distance Δh of each other, they will become indistinguishable. In any physical implementation of the system Δh may vary depending on the accuracy of the instrument measuring the system position, the thermal motion of the system, or many other factors. However, even with “perfect” instruments and at absolute zero, Δh can *never* be reduced to zero. The Uncertainty Principle assures us that there is a minimum block size in phase space, which is a *physical constant of Nature*. Should two orbits arrive within such a block, they are no longer distinguishable, and the information represented in their separate origins is lost. Thus the dimensionality is reduced, and the seeming “rigor” of the formal solutions will have to be modified if they are to represent observed physical reality.

III. Informational Aspects of Noneconservative Flows

In this section we would like to review information theory and its application to the measurement of dynamical systems, and to suggest a concept of *information creation and destruction* for use in describing the behavior of dissipative dynamical systems.

Information, like energy, is a profound primitive concept and as such cannot be defined as a combination of elemental constituents. It can, however, be defined operationally, and given a measure. A formalism for doing this was developed after World War II, largely by Shannon [9], although its roots extend back to Boltzmann, Gibbs, and earlier.

Information occurs in the contexts of *messages* and *measurements*. In both these cases, we have an a priori knowledge of the set from which the actual message or measurement will come. We might know, for instance, that a message will consist of English words, or that an observation will be reflected in the position of a pointer somewhere on a dial. *If* we can assign probabilities p_i to each of the possible outcomes, then we can define the information associated with the outcome as:

$$H \equiv - \sum_i P_i \log_2 P_i. \quad (2)$$

This definition has the property that if we have a completely determined outcome, with probability unity, the information content is zero. We learn

nothing new from the message, thus it contains no information. It is sometimes said that information is the measure of *surprise* of an occurrence, the less a priori knowledge we have, the more information it contains.

The units of information in this definition are “bits”, the information contained in the outcome of an even binary experiment, such as an unbiased coin flip. Logarithms will usually be taken to the base 2 in this paper, thus the informational parameters appearing later will have these units.

It is clear that increasing the accuracy of a measurement increases the information obtainable. Say we are measuring the position of a point on a line of some length, and that as far as we know it is equally likely to appear anywhere along the line. We might proceed by dividing the line up into n equal intervals and seeing which one contains the point. The probability that it is in a particular interval is just $1/n$, so the information content of the measurement is:

$$H = \sum_i 1/n \log n = \log n. \quad (3)$$

As we increase our resolution, we increase the information, and, classically, we can raise the information to an arbitrarily high value by taking fine enough scale divisions on our ruler.

Practically, however, we soon reach a limit, in a given measurement situation the Uncertainty Principle tells us the theoretical resolution limit which can be approached but never surpassed. *The physical nature of reality* limits the information we can learn about a given system to a *particular number*.

The notion of an information cell of a minimum size, or “logon”, resulting from the Uncertainty Principle was suggested by Gabor [10] in the context of a communication channel. One form of the Uncertainty Principle is $\Delta E \Delta t \geq \hbar$. The frequency of a photon is related to its energy by $E = \hbar \omega$, so the minimum product of bandwidth and time which is resolvable into a particular bit, or unit of information is $\Delta \omega \Delta t \geq 1$.

Similarly, phase space can be divided up into blocks of minimum size, $\Delta x \Delta p \geq \hbar$, and these blocks are referred to in statistical mechanics simply as “states”. Once again, the main consequence of the finite minimum block size is that a system which has access to a finite volume of phase space can be found in one of only a *finite* number of states.

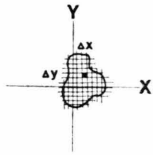


Fig. 3

Now let us re-examine our simple examples in this context. Say we have two observers who independently make measurements of the damped harmonic oscillator at two different times. The earlier observer comes by and determines that the system is in some small volume of phase space, the block size determined by either the practical or theoretical resolution of his measuring instrument.

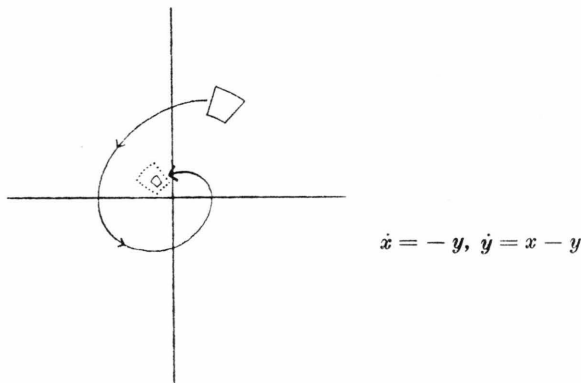


Fig. 4

Later the second observer comes by and, with an identical measuring instrument, determines that the system is within the block indicated by dotted lines. However, the flow in this dissipative system has mapped the original block into a smaller volume, smaller than the resolution of which the later observer is capable.

Now we ask: If the two observers were to communicate by mail, is there anything which one could tell the other?

Yes, the early man could tell the late man, to better accuracy than the late man could determine, the initial state of the system. *Some accessible information has been destroyed* by the contracting flow in phase space.

For a second example, let us have the early observer measure the position of a Van der Pol oscillator prepared *well inside* its limit cycle ($x \ll 1$).

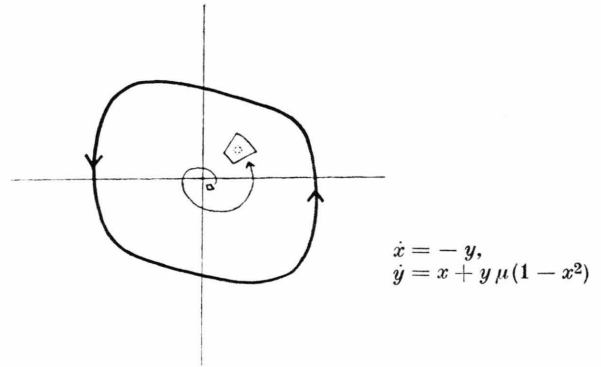


Fig. 5

This time, the volume of phase space determined to contain the system *expands* as energy is fed into the system, and the later measurement more accurately localizes the system. It is now the late man who has information to communicate to the early man, *information has been created* by the expanding flow. Some hitherto unobservable information, such as a random fluctuation of the heat bath, has been brought up to macroscopic expression.

These notions are easily quantified. In general the change in observable information is given by the log of the ratio of states distinguishable before and after some time interval:

$$\Delta H = \log(\Omega_f / \Omega_i). \quad (4)$$

For these examples, the ratio of distinguishable states is the same as the ratio of block volumes before and after the time interval:

$$\Omega_f / \Omega_i = V_f / V_i. \quad (5)$$

The *rate* of information creation or destruction is:

$$\frac{dH}{dt} = \frac{d}{dt} (\log \Omega) = \frac{1}{V} \frac{dV}{dt}. \quad (6)$$

Note that we don't have to know the minimum block size (volume in phase space per state) to determine *changes* in observable information produced by the flow, but if we want to know the *absolute* number of bits in a measurement, we must know the number of states from which a measurement will select one. Even in the classical limit, ΔH and dH/dt remain well-defined.

An important parameter in the description of nonconservative flows is the divergence of the flow in a co-moving frame, the *Lie derivative* with respect to the volume element:

$$\frac{1}{V} \frac{dV}{dt} = \frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \dots \quad (7)$$

In those cases where the number of distinguishable states is simply proportional to the volume, the Lie derivative gives directly the information creation or destruction rate. For the damped simple harmonic oscillator, $dH/dt = -b$, a constant rate of information destruction; for the Van der Pol oscillator, $dH/dt = \mu(1 - x^2)$, a constant rate of information creation when the system is well inside its limit cycle ($x \ll 1$).

However, the number of distinguishable states $\Omega(t)$ arising from some initial block volume *need not* be directly proportional to the volume change under the flow. In the operational approach taken here, the best resolution of some *measuring instrument* will define the minimum block volume, and this instrument will have some minimum uncertainty along *each dimension*. Thus, as illustrated below, a fixed volume block can be distorted so that it effectively occupies a larger volume:

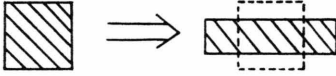


Fig. 6

This is the usual mechanism whereby uncertainty grows in a volume-preserving conservative system. A familiar example is the localized free particle:

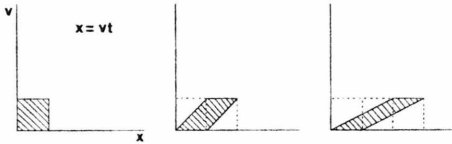


Fig. 7

Even a free particle, which one would think is the most predictable of systems, suffers an eventual delocalization in the presence of *any* uncertainty.

But the number of blocks in which the particle could be found, $\Omega(t)$, grows only linearly with time as the phase space is sheared by the flow, and a power law is typical for this delocalization rate for “simple” conservative systems, i.e., those examples which appear in present-day mechanics textbooks. If the effective volume in phase space occupied by some given initial block grows as a power law, then

$$\Omega \sim t^n \quad \text{and} \quad \frac{dH}{dt} = \frac{n}{t}. \quad (8)$$

The information creation rate dH/dt of such a system approaches zero as time passes. This statement is equivalent to the statement that the system behavior is *predictable* for an indefinite time into the future.

The fundamental distinction between what we would call a “predictable” or “unpredictable” trajectory seems to be whether $\Omega(t)$ is a *polynomial* or *exponential* function of time. If $\Omega(t)$ is a polynomial, the information obtainable from repeated observations of the system saturates with time, i.e., the system is “predictable”. If $\Omega(t)$ is an exponential, dH/dt remains positive, and the system continues to be an *information source*. There are, in fact, fairly simple conservative systems which, on the same energy surface, can show either behavior depending on initial conditions [11].

The main concern of this paper, though, is non-conservative flows, where the phenomenon of attraction is possible. In both examples of attractors examined so far, fixed points and limit cycles, once the initial transient period is over and the flow has contracted a family of nearby trajectories “onto” the attractor, it becomes an informationally static object. The system either remains in a single state, or visits repetitively an ordered sequence of states. Are any more complicated flows possible? The answer, in two dimensions, is no (see however Appendix 1). The Poincaré-Bendixon Theorem assures us that in two dimensions the only stable attractors possible are fixed points and limit cycles.

It has only recently become clear that in higher dimensions, much more is possible. In 1963 Edward Lorenz [1] found what was probably the first example of a “strange attractor”, a flow in phase space in which orbits converge to an object which is neither a fixed point nor a limit cycle, and in 1967 Stephen Smale [12] discussed strange attractors in the context of abstract dynamical systems.

In three and higher dimensions it is possible to have flows which in a compact region *continuously* expand volumes of phase space in some dimension or dimensions while contracting them in others. It is the contention of this paper that the effect of these flows is to systematically *create new information* which was *not* implicit in the initial conditions on the flow. Given *any* finite error in the observation of the initial state of some orbit, and the Uncertainty Principle guarantees such an error, the position of an orbit will be *causally disconnected*

from its initial conditions in finite time, thus any prediction as to its position after that time is *in principle impossible*.

This paper will attempt to make the point clear by studying the examples which are now available.

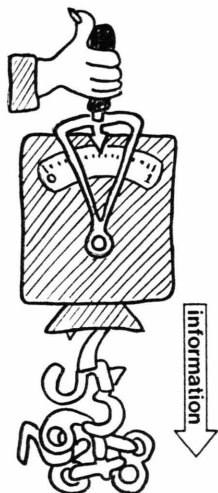


Fig. 8

IV. Maps on the Unit Interval

Perhaps the simplest mathematical objects which display “chaotic” behavior are a class of maps of the unit interval onto itself recently reviewed in an article by May [4]. Much of the behavior of flows in three dimensions is present in a simple form in one dimensional maps. Our plan of attack will thus be to make some observations in the context of one dimensional maps, and then show how the Lorenz and other flows in three dimensions can be discussed in terms of these.

It is useful to think of these maps in terms of a machine, perhaps an actual physical mechanism, which takes a *single* input x_0 and produces a stream of numbers as an output (see Figure 8).

The difference equation $x_{n+1} = F(x_n)$ amounts to a “specification” of the machine, and the “solution” $x_n = f(n, x_0)$ to the difference equation, if we could find one, would give us a prediction for all time of the machine output.

What does information theory tell us about the output of this machine? First, if the machine is producing repetitively a single digit, or a repeating string of any period, the machine is giving us *no new information* after the first period has passed;

we could leave the room for a while and miss nothing. If, on the other hand, it is producing a “random” string of digits, each is *new to us*, we must attentively observe each digit to completely describe the behavior of the machine.

The *amount* of information supplied in one iteration is once again given by

$$H = - \sum_i P_i \log P_i$$

where we sum over the ensemble of possible outcomes for each digit. For a string of random binary digits, the information supplied is one bit per digit.

Shown below is the map for the recursive equation:

$$x_{n+1} = \text{Mod}_1(2x_n). \quad (9)$$

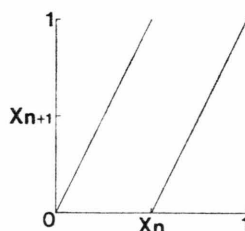


Fig. 9

If our initial value is, say, a number in binary form, and the output we take is whether each iterate falls on the left (zero) or right (one) half, the machine will simply read off the digits of the initial value in order. If it is rational, the string of digits will terminate, or become periodic. If it is irrational, the string will be infinite.

Let us say we can specify the initial value by setting the position of a pointer on the machine between zero and one. Then successive iterations will give us more and more information as to its position. The termination of information output in the rational case simply reflects the fact that the pointer position can be completely expressed in terms of a ratio between the two parts into which the pointer divides the dial.

If we have imperfect knowledge of the pointer position, how far into the future can we predict the output of the machine? This question is answered by calculating the information known at some initial time moment, and the amount of this information lost per iteration of the map.

If the pointer position is known to within an interval Δh , this knowledge represents $H = -\log(\Delta h)$

bits of information. The information change per iteration is computed from the *slope* dF/dx of the map. In the example given above, the slope is everywhere two, thus each small interval is mapped into one twice its size. After one iteration, the pointer could be anywhere within an interval $2\Delta h$. The *change* in information is:

$$\Delta H = (\log 2(\Delta h) - \log(\Delta h)) = 1 \text{ bit} \quad (10)$$

independent of the uncertainty block Δh . There is, after each iteration, *one new bit* available for measurement, correspondingly, our knowledge of the pointer position has decreased one bit. This is also clear from the fact that the map is symmetrically two onto one. After n iterations of the map, where n is given approximately by $n = -\log(\Delta h)$, the initial uncertainty Δh casts its shadow over the entire interval, and all the initial data is lost.

In this particularly simple case, the mapping is *ergodic*, that is, all trajectories which are not on the “measure zero” set of periodic orbits approach every point arbitrarily closely over the course of time. Furthermore, the probability density $P(x)$ for finding the trajectory at x is a *constant*. In more complicated mappings neither need be true. In particular, if the slope of a map is *not* everywhere greater than one, there will be regions of the map which will take a given uncertainty interval into a *smaller* interval. If a given uncertainty interval is mapped after some number of iterations back *inside* of itself, the mapping will be periodic, and stable under small perturbations. The periodic orbits of the preceding example, originating from rational initial values, are not stable.

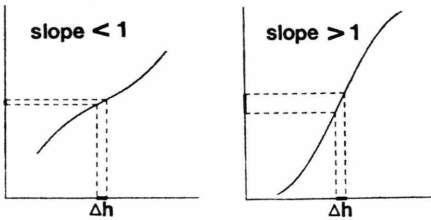


Fig. 10

Examination of the above diagram will make plausible the claim that if a given iterated mapping $y = F(x)$ has a *slope everywhere greater than one*, it will be *ergodic*, at least over some part of the unit interval. Any uncertainty interval will be *amplified*

by a factor dy/dx , and this will happen at *every point* of the map. Consequently, an uncertainty interval Δh around any point x_0 will eventually shadow all parts of the unit interval accessible to the action of the map. Nearby trajectories, wherever they exist, will be *further separated* with each iteration.

If we *assume* that a given iterated map is ergodic over at least some parts of the unit interval, we can write a functional equation for the probability density:

$$P(y) = \frac{P(x_1)}{\left| \frac{dy}{dx} \right|_{x_1}} + \frac{P(x_2)}{\left| \frac{dy}{dx} \right|_{x_2}} + \dots \quad (11)$$

The various points x_1, x_2, \dots are the inverse images of the point y under the map, i.e.,

$$y = F(x_1) = F(x_2) \dots$$

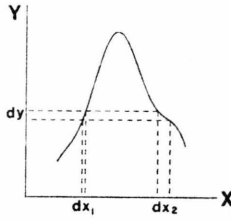


Fig. 11

As the diagram above shows, the number density of trajectories impinging on some small interval in y is equal to the densities at the inverse points of the mapping weighted by the slope at those points. Equation (11) follows by conservation of probability. The absolute value is taken as only the *spreading* of trajectories is important to these considerations. This is like a “master equation”, with a topology given by the map $y = F(x)$.

The probability density $P(x)$ can be computed numerically by the following procedure. First assume some initial probability density, say $P(x) = 1$, and then compute the probability at points y using Eq. (11). The resulting $P(y)$ corresponds to the original $P(x)$ transformed by the action of the map. Successive iterations of this procedure will produce closer and closer approximations to the correct equilibrium value of $P(x)$. If the map has a stable periodic solution, $P(x)$ will converge to sharp spikes on the periodic points, reflecting the fact that $P(x)$ is a series of delta functions in this case. This proc-

ess is discussed in somewhat greater length in Appendix 2.

The information change of a particular point x of some map $y = F(x)$ is determined by the slope at that point:

$$\Delta h = \log \left| \frac{dy}{dx} \right|. \quad (12)$$

The appropriate *average* information change over the entire interval is given by an integral weighted by the probability density $P(x)$. This important parameter we denote by $\bar{\lambda}$:

$$\bar{\lambda} \equiv \Delta H_{\text{average}} = \int_0^1 P(x) \log \left| \frac{dy}{dx} \right| dx. \quad (13)$$

If $P(x)$ is not known, $\bar{\lambda}$ can still be determined operationally, simply by iterating the map, keeping track of the average log of the slope:

$$\bar{\lambda} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \log \left| \frac{dy}{dx} \right|. \quad (14)$$

We expect that for an ergodic map, the sum will be appropriately weighted by $P(x)$ during the iteration operation itself.

The parameter $\bar{\lambda}$ can be easily computed for a few very simple maps, take for example the series of "tent" maps below:

$$y = \begin{cases} \frac{x}{x_1}, & \text{for } x < x_1, \\ \frac{1-x}{1-x_1}, & \text{for } x > x_1. \end{cases}$$

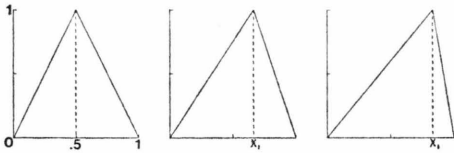


Fig. 12

The first, with $x_1 = 0.5$ is essentially identical to our early example Fig. 9, and each of these has a constant probability density $P(x) = 1$ (examine Figure 11). $\bar{\lambda}$ here reduces to:

$$\bar{\lambda} = -[x_1 \log x_1 + (1 - x_1) \log(1 - x_1)]. \quad (15)$$

This will be recognized as the entropy function $H(p)$ for a binary channel with outcome probabilities p and $(1 - p)$, familiar from elementary information theory.

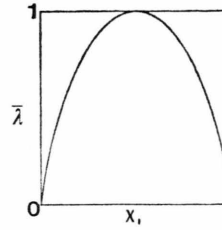


Fig. 13

The maximum information output of one bit per iteration is from the symmetric map, just as the maximum information capacity binary channel is one with equal a priori outcomes.

Another easily computable example is given below:

$$y = \begin{cases} mx + (2 - m) & \text{for } x < x_1 \\ -mx + m & \text{for } x > x_1 \end{cases} \quad \left(x_1 = \frac{m-1}{m} \right)$$

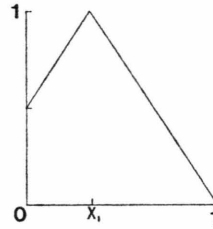


Fig. 14

Here the absolute value of the slope is the same everywhere, and Eq. (14) reduces to $\bar{\lambda} = \log m$. $\bar{\lambda}$ increases from zero to one as the map changes from one-to-one to two-onto-one.

The equilibrium probability density $P(x)$ for the case $m = 0.5$ appears in Figure 15. Each iteration of Eq. (11) adds a finite discontinuity to the distribution. Although the changes soon become too

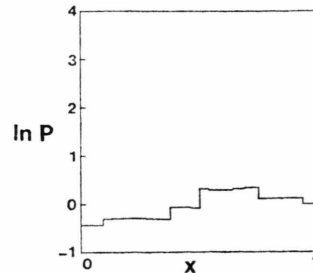


Fig. 15

small to be seen in the diagram, in the limit this function contains an infinite number of discontinuities. Fortunately, $P(x)$ is not needed to compute $\bar{\lambda}$ in this case.

A third example is provided by the mapping $y = rx(1-x)$ which often appears in discussions of mappings of the unit interval onto itself.

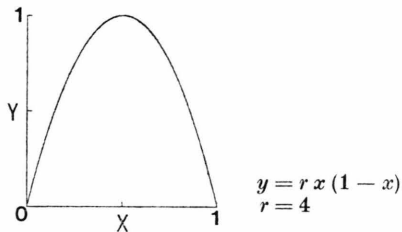


Fig. 16

For the case $r=4$, illustrated above, it is known that the map is ergodic, and that the normalized probability density is:

$$P(x) = (1/\pi)[x(1-x)]^{-1/2}. \quad (16)$$

Thus the integral to be evaluated is:

$$\bar{\lambda} = \frac{1}{\pi} \int_0^1 \frac{\log_2 |4(1-2x)|}{[x(1-x)]^{1/2}} dx \quad (17)$$

which equals one. $\bar{\lambda}$ again equals one bit per iteration. We could have saved ourselves the trouble of doing the integral by noting that this too is a symmetric two-onto-one mapping, so one bit is produced per iteration.

The criterion for complete loss of some initial data is that the “new” information produced at each iteration accumulates to equal the initial data information:

$$H_{\text{initial}} - n\bar{\lambda} = 0, \quad (18)$$

where $H_{\text{initial}} = -\log(\Delta h)$. Thus n , the number of iterations required (on the average) to erase the initial data is:

$$n = \frac{-\log(\Delta h)}{\bar{\lambda}}. \quad (19)$$

We now demonstrate this experimentally with a computer simulation of the mapping $y = 4x(1-x)$. We simulate an uncertainty interval by starting off a family of 10,000 trajectories spaced 10^{-15} apart

and following their paths as we iterate the map. As illustrated below, we define a “target” interval of width 0.01 and ask: how many of the original 10,000 have *not* hit the target after n iterations:

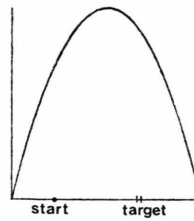


Fig. 17

We term those that have not the “survivors”, and plot the experimental results in Figure 18.

As observed by Yorke in a similar demonstration [13], we obtain a nice exponential

$$\text{Survivors} \sim e^{-n/\tau}, \quad (20)$$

where the “lifetime” τ is simply given by the probability of hitting the interval assuming a random process. In this example, using the known $P(x)$, Eq. (16), and choosing a target interval of 0.39 to 0.40,

$$\begin{aligned} \tau &\approx 1/0.01 P(x) = 100 (\pi(0.4(0.6))^{1/2}) \\ &= 153 \text{ iterations} \end{aligned} \quad (21)$$

which fits the observed slope to high accuracy.

Now according to Eq. (19) the initial data should persist for

$$n = -\log(10^{-11}) \approx 36 \text{ iterations}. \quad (22)$$

Note that the data *deviate* from the probabilistic curve for iteration numbers smaller than $n = 36$. We chose an initial value ($x_0 = 3/8$) which didn’t fall in the target interval right away, and this information was preserved for about 36 iterations.

The map $y = 4x(1-x)$ does not have a slope greater than one at all points, hence there is no guarantee that there are not stable periodic orbits. Exactly what proportion of the parameter range $r < 4$ leads to periodic orbits and under what set of assumptions remains unclear. Digital computer observations are that as r is varied the orbits change in a complicated manner, with periodic orbits of various orders appearing in some parameter windows, but with the orbits being at least “effectively” ergodic over much of the range of r greater than 3.5.

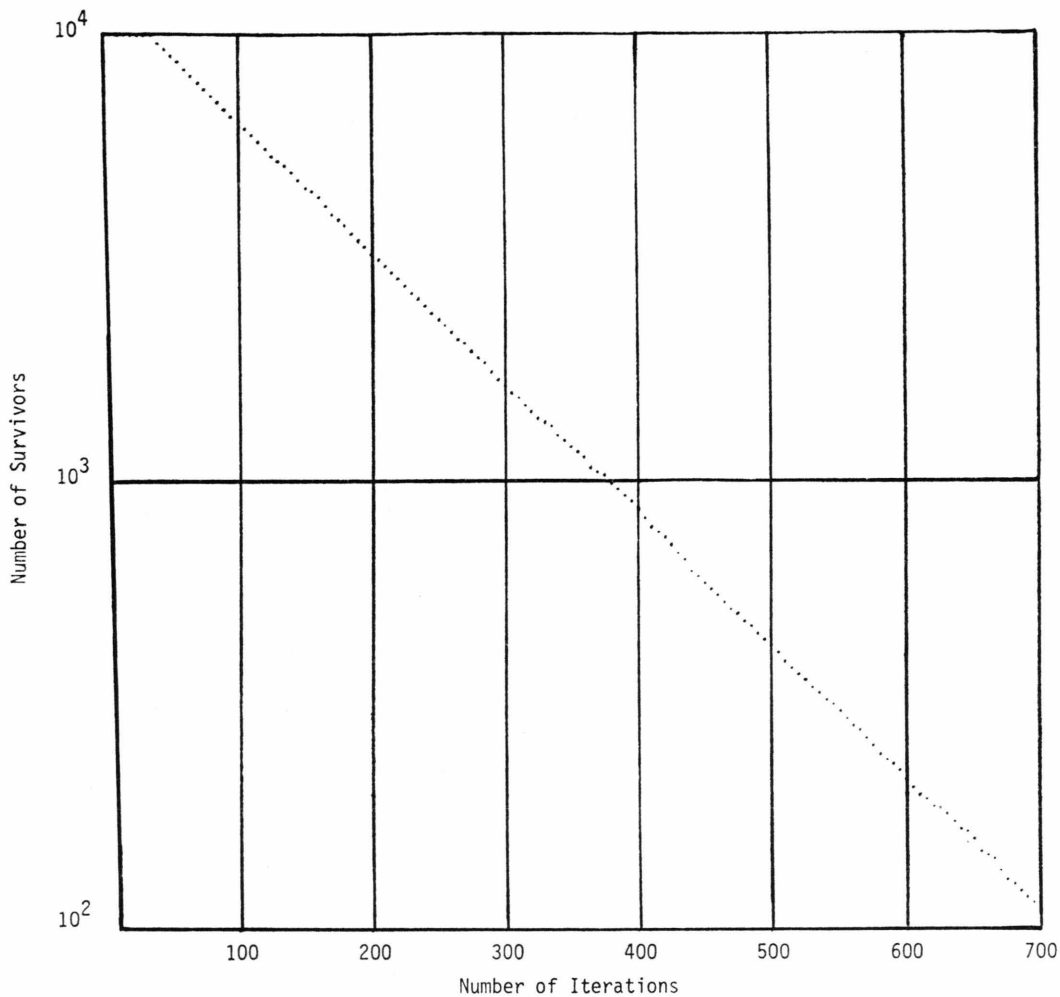


Fig. 18

Presented below are the results of a numerical experiment on the above map in the region $r < 4$. Plotted in Fig. 19 is $\bar{\lambda}$ versus r , where $\bar{\lambda}$ has been

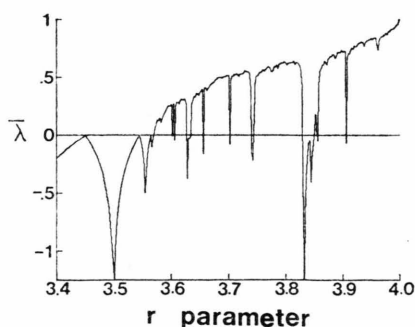


Fig. 19

computed operationally using Eq. (14). The curve is composed of 300 points, spaced 0.002 apart, each representing 100,000 iterations on a digital computer. The low-order attracting periodic orbits which have relatively wide parameter windows are visible as points where $\bar{\lambda}$ is negative. Other higher period attracting trajectories are invisible, presumably because their parameter windows are narrower than the grid spacing of the computation, or perhaps because the "transient" period before the orbit settles down is longer than 100,000 iterations. In general, the curve is smoothly rising as the steepness parameter r is increased, reaching finally one bit per iteration as the map becomes strictly two-onto-one. Figure 19 strongly resembles recent results of Sidnie Feit [14] in computer studies of the Hénon map.

The relationship between the Hénon map and maps of the interval will be discussed in a later section.

Figure 20 shows the probability density, computed using Eq. (11), for the value $r = 3.825$, which is in the effectively ergodic regime. The computed $P(x)$ is in good agreement with the data of Hoppensteadt and Hyman [15] for the same parameter value. They approximated $P(x)$ by accumulating a histogram over many thousand iterations of the map.

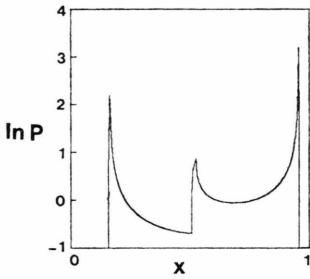


Fig. 20

We have seen that stable periodic orbits have a $\bar{\lambda}$ which is *negative*, thus the sign of this parameter determines whether orbits are periodic or aperiodic. This is clarified by looking at the composite map F^n formed by applying F to the interval n times. It follows from the chain rule that the slope of this composite map at some point x_0 is equal to the *product* of the slopes of the original map of the image points x_1, x_2, \dots, x_n where $x_i = F^i(x_0)$.

$$\left. \frac{dF^n}{dx} \right|_{x=x_0} = \left(\left. \frac{dy}{dx} \right|_{x_1} \right) \left(\left. \frac{dy}{dx} \right|_{x_2} \right) \cdots \left(\left. \frac{dy}{dx} \right|_{x_n} \right). \quad (23)$$

By taking the log of this equation, we see that Eq. (14) just computes the running average of the size of the uncertainty interval Δh . The sign of $\bar{\lambda}$ thus answers the question: does an initial uncertainty interval get mapped into a larger or smaller interval after some large number of iterations?

A negative $\bar{\lambda}$ implies a periodic orbit as the following argument shows. Assume the opposite, a negative $\bar{\lambda}$ but a non-periodic orbit. If no points are revisited, there must be a point or points which are approached arbitrarily closely in the passage of time. But if $\bar{\lambda}$ is negative, then a small interval taken around such a point is being mapped to a smaller and smaller size, and will eventually fall within the original interval. The criterion for a stable periodic orbit is just that some arbitrarily small interval be mapped inside itself.

In the periodic case Eq. (14) reduces to:

$$\bar{\lambda} = \frac{1}{n} \sum_1^n \log \left| \frac{dy}{dx} \right|, \quad (24)$$

where n is the order of the period. How negative $\bar{\lambda}$ is measures the degree of stability of the periodic orbit against small perturbations. If a trajectory is initiated off of some periodic orbit, but within its basin of attraction, the initial data information is *lost* as the orbit damps to its stable values. The parameter $\bar{\lambda}$ governs the rate at which this information is lost to the macroscopic world.

$\bar{\lambda}$ is closely related to a mathematical concept known as *topological entropy*, which, roughly speaking, is the log of the number of subsets created by each iteration. An iteration of a two-onto-one map divides the interval in two, so the number of subsets grows as 2^n , and the topological entropy is $\log 2$. However, as the example of Fig. 12 shows, this gives only an upper limit to the possible information production. This concept does not generalize to negative values.

However, $\bar{\lambda}$, as defined as a limit in Eq. (14), is identical in the one-dimensional case to a *Liapunov characteristic exponent* λ defined by Oseledec [16]. This quantity is a measure of the rate of divergence of nearby trajectories, which in the low dimensional systems studied here is the same as the information production. In his 1968 paper, Oseledec shows that this quantity is *invariant* under coordinate transformations. The information generated or destroyed by a given iterated map is a *topological property* associated with the connectivity of that map, and we would expect it to be coordinate independent. The importance of the Liapunov exponent has been stressed by Ruelle [17], [18]*.

In the context of one-dimensional mappings $y = F(x)$, $\bar{\lambda}$ is invariant under transformations of the form

$$\begin{aligned} x' &= f(x) \\ y' &= f(y) \end{aligned} \Rightarrow y' = F'(x'), \quad (25)$$

where f is applied to *both* coordinates, so long as f is not so violent as to change the topology of the mapping. As an example, apply the transformation

* These recent papers of Ruelle constitute a review of what is presently rigorously known about Liapunov exponents. Reference 17 contains a comment on these as governing the rate of loss of initial data. See also Ref. [45].

$f(x) = \sqrt{x}$ to both axes of the iterated mapping $y = rx(1-x)$. The map of Fig. 16 gets transformed to the one appearing below:

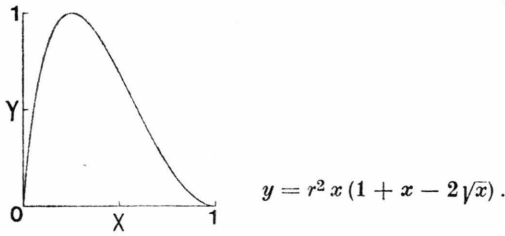


Fig. 21

Despite the changed appearance, the “spectrum” $\bar{\lambda}(r)$ is identical to the one appearing in Figure 19.

We are now in a position to make a few general statements about the expected behavior of an iterated map as some steepness parameter is increased. At low values of this parameter, the average slope is much less than one, and neighboring trajectories are strongly focused toward a stable attracting trajectory. $\bar{\lambda}$ is negative, the informational interpretation being that this reflects an ability of the system to absorb information, i.e., an imposed perturbation is not communicated far into the future. As the steepness parameter becomes larger, slopes approach unity, and $\bar{\lambda}$ approaches zero. The system becomes less able to damp out perturbations, and transients persist for longer and longer periods of time. Finally, $\bar{\lambda}$ reaches zero, and “transients” persist indefinitely. Only at this critical value does the information contained in some initial conditions affect the state of the system far into the future. The effect of maps which have a $\bar{\lambda}$ greater than zero is to systematically increase the uncertainty associated with *any* physical realization of them, until the influence of any initial data is lost.

The relationship of $\bar{\lambda}$ to some map parameter may be very complicated, as Fig. 19 shows. Bifurcations of periodic orbits in the stable region to the left of the figure are marked by points of tangency of the $\bar{\lambda}(r)$ curve to zero, and to the right of the figure the map shows many “resonances” as the behavior changes from a strictly periodic to strictly aperiodic regime. The interval may even be divided into disjoint basins, trajectories initiated from different parts of the interval being attracted to sets characterized by different values of $\bar{\lambda}$, as trivially illustrated below:

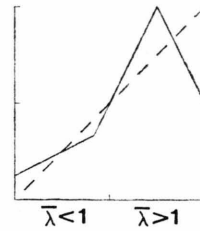


Fig. 22

On general grounds though, if we have a parameter which controls the overall steepness of the map, we expect the correspondence: low steepness, negative $\bar{\lambda}$, periodic or stationary motion; and high steepness, positive $\bar{\lambda}$, chaotic motion.

V. Analogy With Phase Transitions, Overview

It has been remarked that the transition of a flow from a laminar to turbulent state is like a phase transition, having a more or less abrupt change of behavior as some parameter reaches a critical value. Here we seek to expand somewhat on this analogy, starting from the context of one-dimensional maps.

Phase transitions are characterized by a divergence in a coherence length near some critical point. The corresponding quantity in a dynamical system is the *correlation time*, the length of time perturbations are propagated into the future. Let us recall the two observers and have them observe some system at two different times:



Fig. 23a

Now have the early man at A attempt to send a message to the late man at B, using the system as a medium. If the system happens to be periodic or stationary, let him perturb the system slightly from its stable positions. If the system is nonperiodic, let him localize it in some fashion. How far downstream can any information be propagated?

As discussed in Appendix 2 for one-dimensional maps, the loss of initial data can be described as the relaxation of an initial probability distribution to an equilibrium distribution characteristic of the system. The informational parameter $\bar{\lambda}$ governs the rate of this damping. In both the periodic and

the aperiodic case the characteristic number of iterations for total loss of information is:

$$n = \frac{H_{in}}{|\bar{\lambda}|}, \quad (26)$$

where H_{in} is the amount of initial data and the absolute value of $\bar{\lambda}$ is taken. We will soon generalize $\bar{\lambda}$ to be an information production rate in time,

$$\bar{\lambda}_t \equiv \left. \frac{dH}{dt} \right|_{\text{average}} \quad (27)$$

and the characteristic number of iterations becomes a *characteristic time*:

$$\hat{t} = \frac{H_{in}}{|\bar{\lambda}_t|}. \quad (28)$$

Now note the behavior of \hat{t} as $\bar{\lambda}_t$ approaches zero. In the periodic case, the system is less and less able to damp out perturbations. In the turbulent case, localized probability distributions take longer and longer to relax. In both cases, the observer at A is able to have an effect farther and farther into the future. Most of our academic experience is based on the highly singular case of a divergent correlation time, this has perhaps hindered appreciation of these simple points.

The analogy is not complete, however. In the flows to be discussed later in this paper, discontinuous and hysteretic changes of $\bar{\lambda}$ as a function of some parameter occur often. Also, the zeros in $\bar{\lambda}$ at bifurcations of periodic orbits do not have a counterpart in simple phase transitions.

We have distinguished in these problems a “specification”, which is a rule for getting from one point to the next, be it an iterated map, or a set of differential equations like those to be studied in the next section, and a “solution”, the actual trajectory. The specification corresponds to the blueprint of a mechanism, it can be deterministic, and can exist in the realm of pure mathematical thought.

It should be clear by now that closed form solutions predicting the long-time behavior of one-dimensional maps and other chaotic generators are impossible. Say we had such a solution:

$$x_n = f(x_0, n). \quad (29)$$

This would be an *algorithm* taking as inputs the initial position and the iteration number, and producing the position of the n th iterate. However, no amount of algebraic manipulation of known infor-

mation can produce new information, the two sides of the equation are qualitatively different. Closed form predictions are impossible because the *information* they would represent simply *does not exist* prior to the operation of the mechanism*. A more operational way of stating this is to say that the solution algorithm for the case $\bar{\lambda} > 0$ is not “efficient”, i.e., the input required to determine a trajectory position to given accuracy grows exponentially with time [19].

Another immediate consequence of this interpretation is that all these chaotic generators are *irreversible* in the ordinary sense. Imagine someone entering a room where the mechanism of Fig. 8 has been operating for some time, and attempting to recover the initial value. He will fail, even if he has complete knowledge of the mechanism, because any map which produces information, that is, has a positive $\bar{\lambda}$, is at least in some region not single valued. In order to reverse the mechanism he must *know* from which sheet of the map the trajectory came on each iteration. This amounts to changing the direction of the arrow marked “information” in Fig. 8; *to reverse the mechanism the information flow in time must also be reversed*.

Often the question is raised, in the discussion of solutions to one-dimensional maps, and particularly to digital or analog solutions to systems of differential equations: “But how do you know the solution you have computed is the *exact* solution?” Under our interpretation, the response is clear, an “exact” solution implies an *infinite* amount of information content. But the action of these systems is specifically to *reduce* the information available. While the notion of an “exact” solution may have some use in the description of information preserving conservative flows, the concept is completely inappropriate in the discussion of chaotic generators.

VI. Relation of One-Dimensional Maps to Dynamic Flows

We have argued that the maps discussed in Sect. IV *create* information, in the sense that any physical realization of them systematically brings the uncertainties, the bath of microscopic randomness in which anything physical is immersed, up to

* Closed form solutions may exist in a formal sense, but have no predictive value for long times.

macroscopic expression. A measurement at some instant will yield, by definition, a certain constant amount of information. Information is defined in terms of the set of possible outcomes of the measurement, a set which is fixed in advance. In order for this “pool” of information to remain of constant size, one bit of old information must be “destroyed”, or eliminated from having any effect on the state of the system, for each bit which is “created”, or brought up from the microscales to macroscopic influence. In this way the information contained in the initial conditions is systematically replaced by information *not* implicit in the initial conditions until, after a finite time or number of iterations, the state of the system is undetermined.

We have seen how this is accomplished in the map $y = 4x(1 - x)$, an average of 1 bit per iteration is added because the slope is greater than one on the average, causing nearby trajectories to diverge, and one bit of initial data is removed due to the two-onto-one nature of the map, we cannot tell from which of two points a given trajectory originated. These two processes are in some sense “orthogonal”, the information created in an iteration is *not* the same as that destroyed, unlike the example of Appendix 1. The object of this section is to demonstrate that dynamical flows in three dimensions can be treated in much the same way as iterated one-dimensional maps. We will illustrate this by computing the informational parameter $\bar{\lambda}$ for the Lorenz attractor.

Imagine a flow in three dimensions constructed in the following way. The flow occurs largely along one dimension, but trajectories tend to spread out in a second dimension, and tend to be crowded together in a third:

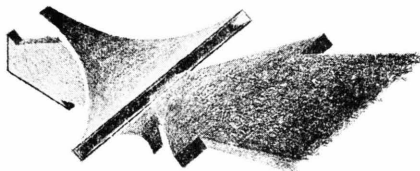


Fig. 23

Such a flow is called a *C-system*, a discussion is given in Arnold [20]. A physical realization of this flow will act on an initial volume of trajectories to *reduce* it in dimension to a quasi-two-dimensional sheet of thickness Δh , the uncertainty length.

This flow has one property we seek, nearby trajectories are separated as the sheet fans out, however we want to arrange that this separation can take place continually on a compact manifold. A general way of doing this is indicated in Figure 24.

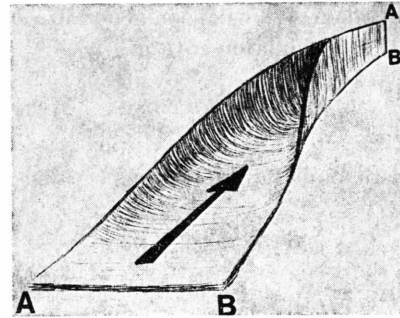


Fig. 24

A two-dimensional ribbon, after expanding to twice its width, is folded over, “sutured” together at AB, and mapped back into itself. If the two-dimensional object of Fig. 24 is given a half-twist to the right or left, the ends joined together and the resulting object imbedded in 3 space, we obtain Fig. 25 below:

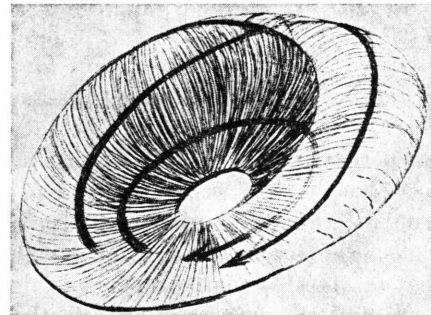


Fig. 25

What we have constructed in this manner bears a marked resemblance to a three-dimensional attractor constructed by Rössler [3]. An analog computer simulation of his equations is exhibited in Figure 26. If we take a cut across the attractor as indicated and construct a return map, we obtain the single humped map shown in Figure 27. The flow has performed a “horseshoe map”, similar to those discussed by Smale*, and the one-dimensional

* However, the horseshoe map described in Ref. [12] does not produce an attracting set.

reduction looks very much like one of the maps discussed by May.

The Rössler attractor is a three-dimensional object, and what we have constructed is quasi-two-dimensional, with a thickness Δh . The idealized solutions to Rössler's equations contain a "Cantor Set", which often appears in discussions of strange attractors; the information as to the origin of a particular trajectory appears as which one of an infinite number of infinitesimally spaced sheets it lies on.

However, the comoving volume contraction rate, or Lie derivative

$$\frac{1}{v} \frac{dv}{dt} = \frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z} \quad (30)$$

is so large for this and most of the other examples which have appeared that only the outer leaf or outer few leaves of the Cantor set are visible in any simulation. The uncertainty present in *any* physical system *truncates the Cantor set*, whole families of trajectories become identified, and the object can be treated as the single-sheeted quasi two-dimensional object described.

The exact point where this truncation takes place or where the "suture" in Fig. 25 is made is to some extent arbitrary, depending on Δh , the uncertainty length. But the *fact* of the *eventual* joining of trajectories is a logical consequence of the Uncertainty Principle. The asymptotical approach of trajectories was remarked upon in Lorenz's original paper, and has been referred to in the modern mathematics literature as "locally eventually onto" [21].

This procedure of reducing a three dimensional flow to a one dimensional map, although not rigorously justified here or elsewhere, has become common. It appears in the original paper by Lorenz, as

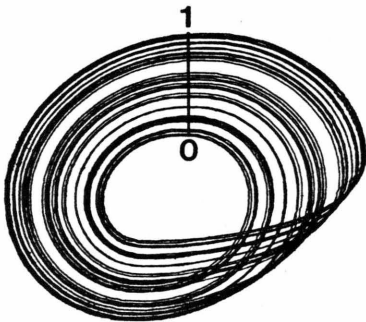


Fig. 26

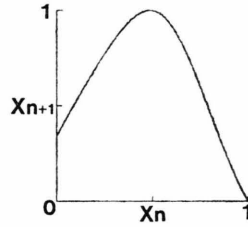


Fig. 27

well as more recent work by Guckenheimer, Rössler, and others. The one-dimensional map preserves much of the behavior of the full flow, including, as we shall demonstrate, the same information production characteristic.

The discussion of attractors as flows along two-dimensional sheets was begun by Lorenz [1], and developed further by Williams [21]. He calls these structures "branched manifolds", and a slight re-drawing of his picture of the structure of the Lorenz attractor (see Fig. 29) appears below:

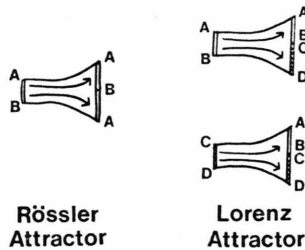


Fig. 28

There are now *two* sheets, sutured together as shown. Trajectories are either mapped back onto the same sheet, or onto the opposite sheet, depending on where the trajectory hits the suture.

VII. Calculation of $\bar{\lambda}$ for the Lorenz Attractor

The principal conclusion of Lorenz's paper "Deterministic Nonperiodic Flow" is that there exist deterministic systems in which very long range forecasting is impossible. The following quote comes from the last paragraph of that pioneering work: "There remains the very important question as to how long is 'very long range' ... the answer may be obtained by comparing pairs of numerical solutions having nearly identical initial conditions."

We now answer this question for the case of the Lorenz equations (see also Ref. [44])

$$\begin{aligned}\dot{x} &= 10y - 10x, \\ \dot{y} &= -y - xz + Rx, \quad R = 28, \\ \dot{z} &= xy - (8/3)z,\end{aligned}\quad (31)$$

by computing the informational parameter $\bar{\lambda}$ for this system. With this parameter in hand, a prediction can be made as to how long the information in a given initial condition will persist. We then verify this prediction in just the manner suggested by Lorenz.

First we construct a return map in the fashion done in his paper, we compute a trajectory, and record successive passes through a maximum in the z direction. The initial point used was the same as that taken by Lorenz, but the transient associated with this initial condition was allowed to die away before points on the return map were accumulated. Taking advantage of the computing power unavailable in 1962, we assemble the map with 975 points, using a standard 4 point Runge-Kutta algorithm with a dimensionless time step of $h = 0.001$.

The return map obtained is given in Figure 30. The time between successive passes through the return map is given in Fig. 31, note the singularity associated with the fixed point at the origin in both figures.

The use of the maximum in z corresponds to taking the return map on a cut in the flow marked "A" in Figure 29. As the attractor is symmetric (with respect to changes in coordinates $x \rightarrow -x$, $y \rightarrow -y$, $z \rightarrow z$, not mirror symmetric), the cut on one leaf will yield the same return map as a symmetrically placed cut on the other, it suffices to give only a single map, if the trajectory falls on the right half of the return map it signals a change of the trajectory from one leaf to the other.

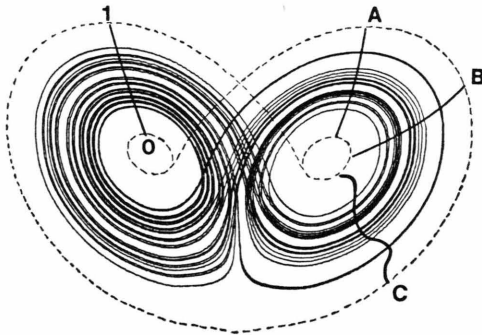


Fig. 29

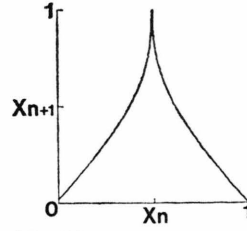


Fig. 30

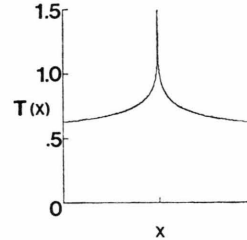


Fig. 31

It might be asked: Why the cut marked "A"? Why not assemble a return map from a cut such as "B", or even "C"? Oseledec's theorem gives the answer, *any* cut which is transversal to the flow may be taken, the $\bar{\lambda}$ computed from any such cut will be the same. Taking a different cut amounts to a coordinate transformation on the return map, and $\bar{\lambda}$ is an *invariant* quantity. This justifies our ad hoc procedure of using the maximum in z .

Having the return map in hand, the computation of $\bar{\lambda}$ proceeds just as for the earlier one-dimensional maps. First, the probability density $P(x)$ is computed using Eq. (11). After 13 iterations, $P(x)$ had converged to within a percent, the resulting distribution is shown in Figure 32. Then, the integral of Eq. (13) is evaluated numerically. We determine the information production for the return map of Fig. 30 to be:

$$\bar{\lambda} = 0.98 \text{ bits per iteration.} \quad (32)$$

This number is about what we would expect from our earlier examples. The return map is not quite two-onto-one, and is nearly symmetrical, yielding an information production of slightly less than one bit per iteration.

In order to describe the informational character of the Lorenz and other attractors in *time*, it is necessary to fold in the time taken per iteration $T(x)$ of Fig. 31

$$\bar{\lambda}_t \equiv \left| \frac{dH}{dt} \right|_{\text{average}} = \int_0^1 \frac{P(x)}{T(x)} \log \left| \frac{dy}{dx} \right| dx. \quad (33)$$

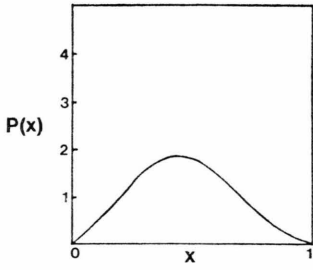


Fig. 32

The equivalent of Eq. (12) now is:

$$\bar{\lambda}_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\log |dy/dx|}{T_n}, \quad (34)$$

where T_n is the length of time taken by the n th iterate. The information production rate in time for the Lorenz attractor was found to be:

$$\bar{\lambda}_t = 1.19 \text{ bits/sec} \quad (35)$$

if the dimensionless time units are taken to be seconds. Some pains, but not great pains, were taken with the numerical computer work. These numbers are probably good to a percent or so.

VIII. Experimental Verification

We now demonstrate that the $\bar{\lambda}_t$ calculated in the preceding section has predictive value. A given observation of an initial condition has a certain value in bits, and when a strange flow has generated this amount of information, the initial data is lost. We prepare a system implementing the Lorenz equations, in this case an analog computer, measure the information content of an initial condition, and then allow the system to run for varying lengths of time, noting the dispersion of trajectories emanating from the "same" initial condition as time passes. When the initial data is gone, we expect the trajectories to be "completely dispersed".

The analog computer used, a Systron Donner 10/20, is a highly stable device, capable of a voltage resolution at the computation speeds used of about 1–2 mv. This was measured by the method described in Appendix 1. The return map of Fig. 30 spanned a range of 15 volts on the computer with the scaling used, thus the informational value of a given initial condition is:

$$H_{\text{in}} = \log_2 \frac{15}{0.002} \approx 13 \text{ bits}. \quad (36)$$

On the average, after a time \hat{t} , where

$$\hat{t} = H_{\text{in}} / \bar{\lambda}_t \quad (37)$$

we expect the initial data to be lost. $\hat{t} = 13/1.19 \approx 11 \text{ sec.}$ for this system.

Figure 33 was constructed by initiating the analog computer off of the "same" initial condition fifty times, and then halting the computation after various multiples of a time $1/\bar{\lambda}_t$ had passed. This is the time required for the flow to generate *one new bit* of information. The initial condition used was obtained by allowing the computer to run for some time, to insure that the trajectory was indeed "on" the attractor, and then halting the calculation, assuming that the point obtained, one near the center of the flow, is a "typical" value. It is marked by the cross.

The first figure in the series shows the motion of the fifty trajectories after eight new bits have been added. The family of trajectories is still only slightly larger than the size of a recorder pen dot. As the next few bits are brought up, though, the trajectories are sheared, split, and scattered over the body of the attractor. At the predicted value of 13 bits, corresponding to the time $\hat{t} = 11 \text{ sec.}$, the trajectories are well dispersed over the width of the flow.

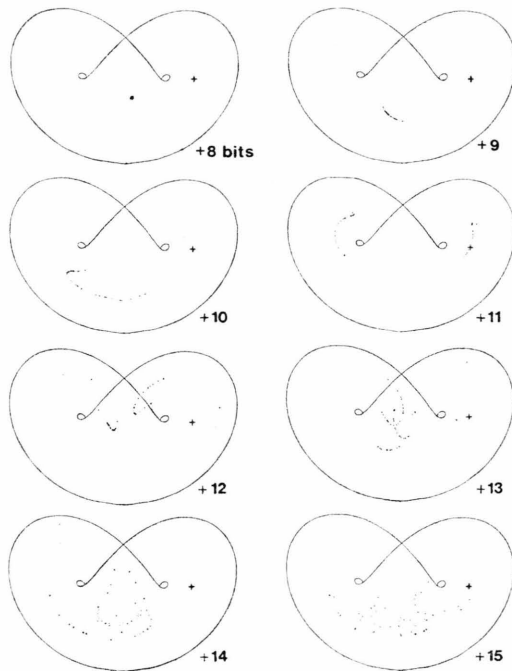


Fig. 33

There still remains some localization, as the mixing only occurs transversally to the flow. In fact, due to the varying lengths of time which orbits take depending on their position on the return map, the initial position data could be partially recovered from its “coded” form in the *ensemble* of final positions. A more sophisticated informational analysis is required for a description of this. However, after only a few more seconds, trajectories are distributed essentially uniformly over the body of the attractor.

A complete verification of the computed $\bar{\lambda}_t$ would require a statistical study of this type from many initial conditions. However, due to the exponential spreading of trajectories, our simple analysis suffices to show an approximate agreement. The first bits of initial data lost are hardly visible, and the spreading occurs very rapidly as the last few bits are lost. As the last bit of initial data information allows us only to distinguish the initial position as coming from one of two randomly distributed sets, this information may be of no use to us.

IX. Determinism and Reversibility in Phase Space

The description of dynamical systems in terms of orbits in phase space has proven to be an abstraction of great power. One problem with this approach, however, is that all the mathematical theorems and constructs pertaining to Hausdorff manifolds have been applied to dynamical systems theory even when they are not fully appropriate. Evidence for this lies in the recurrent appearance in the literature of such blatantly unphysical objects as Cantor sets, and such unphysical distinctions as analytic vs. non-analytic mappings, countable vs. uncountable infinities of orbits, etc., occurring in what are purportedly models for physical phenomena.

Implicit in any flow on a Hausdorff manifold are two properties:

- 1) *Determinism*: There exists a rule governing the motion at each point.
- 2) *Reversibility*: All trajectories are distinct and remain so under the flow.

In a given dynamical system the first property follows from the simple existence of a differential equation governing the flow, and the second follows from the Existence and Uniqueness Theorem on Hausdorff manifolds [22].

However, the Second Law of Thermodynamics, perhaps the single most profound observed fact of Nature, is in direct contradiction with property 2. A simple geometrical expression of the Second Law is that trajectories *do* emerge, information *is* lost, in the universal increase in entropy. The Second Law is particularly important in the *dissipative* systems which attractors model, it would thus behoove us to take this into account at the most basic level.

We therefore suggest the logical separation of the above two properties, and the abandonment of the second. Determinism need *not* imply reversibility. Figure 34 shows that by the simple expedient of having trajectories join one can have a flow with a perfectly well defined rule for getting from one point to the next, yet which is irreversible.

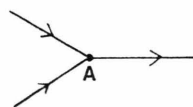


Fig. 34

At attempt to reverse the flow of Fig. 34 will lead one to point A, where the Axiom of Choice will have to be invoked to “decide” which path to take.

The identification of trajectories precludes the use of Hausdorff manifolds to describe this type of flow; the direction of the flow is not continuous in the neighborhood of point A. To render the joining of trajectories consistent with a smooth flow it is necessary to posit a “fuzzing distance” Δh , any trajectories which approach each other to within this distance are considered identified. The absence of an uncertainty length in what might be called “classical mathematics” is responsible for the inappropriate uniqueness theorem.

In flows of the strange attractor type, in which some trajectories approach each other arbitrarily closely, this quantization block may be taken to be arbitrarily small, but it is *logically required* to have some finite value to maintain consistency with the Second Law. The logical relationship between the Second Law and quantization in physics has been remarked upon by Fermi and Von Neumann among others during the conceptual development of quantum mechanics, and the resolution of many paradoxes of the “Maxwell’s Demon” variety requires the invocation of the quantum principle in some form. Strange attractor flows contain an expression of this logical relationship in a pure geometrical

form. We will argue next that these entropic flows are more appropriately modelled by a manifold which has the linked postulates of irreversibility and quantization *built in*.

X. Steps Toward a Classification of Strange Attractors

The study of the topology of strange attractors has been largely discussed in terms of the topology of the *solutions* to the differential equations governing the flow. This topology, studied at greatest length in the Lorenz attractor, has proven to be very complicated, with an uncountable infinity of solutions interweaving through each other, often described by a “kneading sequence”, a string of binary digits of infinite length [21]. The solutions are of course highly unstable, in the sense that any small perturbation will completely alter them in finite time, as described earlier in this paper.

There is, however, a highly stable object intrinsic to a strange attractor, if “stable” is taken to mean perseverance of form through time. While a strange attractor *solution*, whether specified by a kneading sequence or otherwise, is undetermined and without physical meaning past a time $t = H_{1n}/\lambda_t$, the *envelope* formed by the stochastic motions of solutions is a time independent object, if the governing equations are autonomous (i.e., if the physical mechanism producing the solution, the “specification”, isn’t changing with time).

It is this envelope which is represented in Figs. 25 and 28 of Sect. VI, by the qualitative descriptions of the flows in the papers by Lorenz and Rössler, and by the “branched manifolds” discussed by Williams. Each of the “uncountable infinity” of solutions to the Lorenz equations move on a *single* branched manifold. Inasmuch as a physical “solution” to a strange attractor flow is a transitory object, existing only when some physical mechanism is operating, and governed as we have seen by the chaotic motion of the heat bath, a classification scheme might more profitably be based on the branched manifolds on which solutions move.

The considerations of the preceding section, particularly the necessity for incorporating the Second Law of Thermodynamics as a *postulate* in any discussion of strange attractor forms, reduce the question of classification to the following:

What topological types of branched manifolds are possible, subject to the restriction that trajectories can join but not split?

Figures 24 and 36 illustrate how in three dimensions this seemingly paradoxical condition can be met. A compact, ribbon-like structure can be imbedded in 3-space, in which trajectories are continuously separated, yet which contains *no decision points*, places where some outside agency has to be invoked to determine which path a given orbit will take. The Lorenz attractor, which splits families of trajectories, accomplishes this with an *unstable fixed point*, the singular trajectory which would arbitrarily have to “decide” which way to turn simply halts:

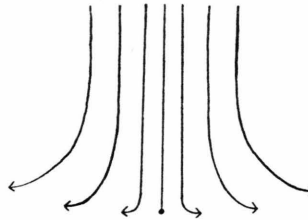


Fig. 35

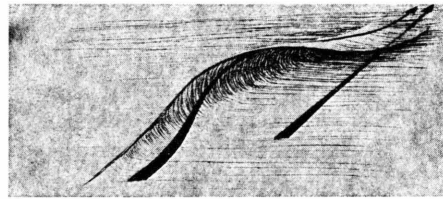


Fig. 36

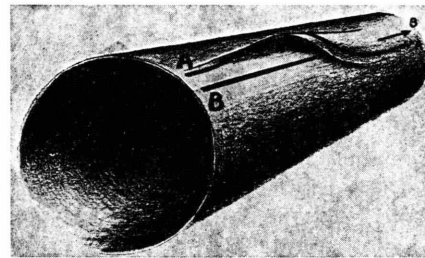


Fig. 37

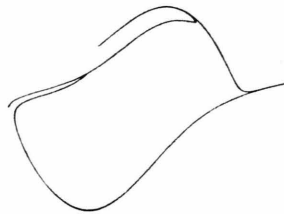


Fig. 38

Both the amount of information generated and the time taken for this singular orbit are infinite, see Figs. 30 and 31.

Although unable to present a complete classification of strange attractors in three dimensions, we can divide the forms so far observed into the following general categories:

1) *No fixed points, no boundaries.* Figure 36 exhibits a curl-like structure which has often been observed as the solution envelope in analog computer studies of non-linear oscillators with a periodic driving force. Examples are the driven Van der Pol in a restricted parameter range [23], a variant of the driven Van der Pol, Eqns. 38 below in a much larger parameter range, and a driven Duffing's equation studied by Holmes [24].

$$\dot{x} = ky + ux(a - y^2), \quad \dot{y} = -x + A \sin \omega t. \quad (38)$$

The typical situation is illustrated in Figure 37. If an oscillator which already has a limit cycle is driven, the phase plane is extended in a third dimension which is periodic in the driving frequency. If the driving amplitude is small, the form generated by the extended limit cycle remains topologically a torus, but under the appropriate circumstances part of this tube can be taken up and folded back down onto itself, as illustrated. This means that if we order the points along the original limit cycle and then examine the order 2π later, we will find a shuffling of trajectories. The de facto three-onto-one mapping leads as before to a creation of new information.

Taking a cross-section of the flow of Eq. (38) yields Figure 38. Three "ears" are seen to be in operation, each in a different stage of folding. The period for one complete fold in this example is 3π , not the period of the driving term [25]. This same behavior is in fact present in the ordinary driven Van der Pol equation, but because the narrow parameter range in which it occurs is far from the region usually studied (high μ) it has not been previously observed.

Note the close similarity of the curl of Fig. 36 to Thom's "cusp" catastrophe [26]. In fact, we can crib a chapter from catastrophe theory, and claim that Fig. 36 is the *only* boundary-free strange attractor in three dimensions. Catastrophe is a *static* theory, categorizing the ways in which a potential function can develop discontinuities in observed

equilibria as we move quasistatistically in some control space. The potential function forming the cusp goes from single- to multi-valued as a particular parameter is varied. If we make the identification between this parameter and time, the catastrophe surface models the *dynamic* flow, **except** for the subsequent return of the flow to a single sheet as trajectories merge along the suture. The number of boundary-free strange flows in the next few higher dimensions (up to dimension 5) are similarly restricted by catastrophe theory.

2) *No fixed points, but boundaries included.* The flow of Fig. 24 which produces Smale's horseshoe, does contain trajectories which are on a boundary. Although no easy correspondence can be made with existing theory, as in the preceding flow, the evidence, based on analog computer examinations of an increasing number of systems, is that this is the only such flow in three dimensions. Of course, the horseshoe need not involve a symmetric doubling, *any* degree of folding suffices to produce shuffling of trajectories and strange behavior.

The flow below contains *two* such folds in sequence [27].

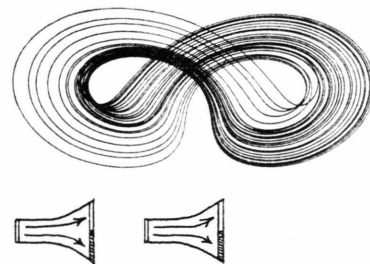


Fig. 39

3) *Fixed points and boundaries included.* If flows are split by the insertion of fixed points, and the resulting split ribbons woven around and given various numbers of twists before being sutured back onto the main flow, any number of topologically distinct branched manifolds can be constructed. The Lorenz attractor of Figs. 28 and 29 is an example of one such flow, no doubt many others will be forthcoming.

An interesting point is that both the structures 1) and 2), as well as the Lorenz attractor (as noted by Ruelle) [28], are *orientable*, that is, non-equivalent right- and left-handed versions exist. However,

a symmetric object (“boson”?) can be constructed by hooking together two similar forms of opposite parity.

The number of distinct branched manifolds which can be assembled in this way, although indefinitely large, is clearly *countable*. A classification scheme perhaps already exists in combinatorial theory or algebraic topology.

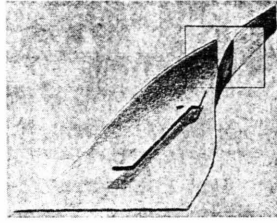


Fig. 42

XI. Relation Between 3D Flows, 2D Maps, 1D Maps

In this section we briefly discuss the different types of mathematical objects displaying turbulent behavior, and argue that many of them are just different aspects of the forms discussed in the preceding section.

The reduction of three-dimensional flows to one-dimensional maps has been mentioned earlier; below we quickly recapitulate this process. A particular parameter value of Eqn. 38 is chosen, and the projection of this flow onto the $x-y$ plane appears as below [29]:

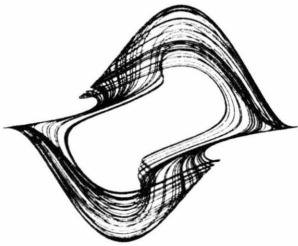


Fig. 40

This seems complicated, but the Poincaré map of this flow reveals it to be made up of three ribbons, each performing a horseshoe fold of the type illustrated in Figure 24:



Fig. 41

The relation between the flow and the return map is schematically illustrated below:

We now turn to iterated maps of the plane onto itself, in particular, the well-studied Hénon map [6].



Fig. 43

$$\begin{aligned} x_{n+1} &= y + 1 - ax^2, \\ y_{n+1} &= bx, \\ a &= 1.4, \quad b = 0.3. \end{aligned}$$

First we note the general similarity in form between this map and the flow cross-section of Figure 42. Next we see that although the form is similar, several of the outer leaves of the Cantor set are visible in this simulation. The volume contraction ratio between sections in the case of the flow is given by the integral of the Lie derivative:

$$\ln \left(\frac{V_{2\pi}}{V_0} \right) = \int_0^1 \left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{t}}{\partial t} \right) dt. \quad (39)$$

The average value of $V_{2\pi}/V_0$ is .004 for the case of Eq. (38) and Figure 41. The corresponding quantity for an iterated map is just the Jacobian, for the usual parameters taken in the Hénon map this is $-.3$. Comparing the absolute value of these two numbers makes understandable the difference in the number of visible leaves.

The flow of Fig. 40 can be reduced to a one-dimensional map simply by placing coordinates along the ribbon cross-section which, to machine accuracy, is one-dimensional. As in the case of the Rössler attractor, this leads to a one-hump difference equation of the type reviewed by May.

The complicated outer structure of the Hénon map precludes the simple reduction to a one-dimensional map. However this map, despite formally

being one-to-one, identifies trajectories in the same way that a one-dimensional map does explicitly. Figure 44 below shows the *lines of identification*, or “stable manifolds” for the Hénon map. These have also been computed by Carles Simó, of the University of Barcelona, Spain.

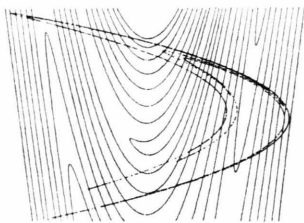


Fig. 44

All the points along any of these curves continually move closer to each other along succeeding images of the curve, until they all are within an uncertainty length, and become identified. They were computed by determining the direction of maximal contraction at points on the plane (maximal under many iterations), and then integrating this direction field. Members of this family of curves are taken into each other by the action of the map, and the boundaries of the basin of attraction are in this family. Motion along these contracting directions does not add any new information to the initial conditions (as per the discussion of Figure 6). Coordinates could in principle be assigned to the lines of identification, and the rule which takes them into each other would then be given by a one-dimensional map. Sidnie Feit recently computed the Liapunov exponent for the Hénon map, by keeping track of the average rate of divergence of trajectories [14]. This rate is the same as the growth rate of an uncertainty length along the maximally unstable directions.

The Hénon map is closely related to the simple parabolic map $y = rx(1 - x)$. It reduces to the parabolic map in the case $b = 0$, and can be written in terms of an expansion of the parabolic map when $b \neq 0$ [30]. The similarity is emphasized by the qualitative resemblance of the spectrum of the characteristic exponent as a function of the parameter a computed by Feit to the corresponding spectrum of the parabolic equation computed in this paper (Figure 19).

It is clear that any flow on which one can define a return map induces an iterated map. Holmes in fact is able to approximate his forced oscillator with an iterated polynomial map and reproduce the qualitative behavior [24]. However, inverting this process presents a somewhat stiffer conceptual challenge. The iterated polynomial maps of Hénon [6], of Stein and Ulam [7], and of Yorke [13] bear what is often an amazingly detailed resemblance to the cross-section of some flow. One might ask, where is this flow?

This question allows us to reiterate the claim that the fundamental objects in all of these examples are the dynamic forms discussed in the preceding section, which represent the *action* of the map or flow. The specification of some iterated map in the form of a set of equations sitting on a piece of paper is a static object, but when the map is *physically implemented* its corresponding dynamic form is invoked. As an example, imagine someone presented with the graph of a one-dimensional difference equation, perhaps the parabolic equation, and asked to iterate it using only compass and straight-edge. He could do so, and numbers would never enter into the proceedings. Now imagine that he performs one iteration per second, and that a glowing light is attached to some active part of this apparatus. The graph now defines a rule for getting from one point to another in a periodic phase space, and the light will trace out the appropriate form, in the case of the parabolic equation, it would be that of Figure 24. Even if a digital computer did the iterations, a topologically equivalent form would be manifested in the appropriate state space, with the irreversible joining of trajectories occurring where the machine ran out of significant figures.

Iterated maps of the plane onto itself are often described in terms of a series of actions to be performed, e.g., stretch it, bend it double, etc., this is a representation of the *physical action* which is the basic form. Smale's original paper discussed strange attractors in these terms [12]. These prescriptions allow us to infer certain facts about the underlying dynamic forms, for instance, if the series of distortions of a plane into itself can be carried out within a two-dimensional subspace, then the dynamic form can be represented in three-space, as in Figs. 24 and 36. But if a folding out into a third dimension is required, then the underlying form cannot be so imbedded.

The Hénon map of Fig. 43 has a negative Jacobian, and thus its action cannot be represented in three-space as a single iterated fold of Figure 24. It can, however, be represented as two folds in series, resembling Fig. 39, if the folds are taken with *opposite handedness*. The resulting object is thus topologically mirror symmetric. If the parameter b is taken to be negative, the Jacobian becomes positive, and the resulting map (which can be turbulent) [14] can be represented as a single fold in three dimensions, with a *definite parity*.

In conclusion, we have argued that forms representing the *physical action* of a map are the natural basis to characterize the various turbulent maps and flows. The Rössler attractor, the Hénon map, Fig. 41 and the single-humped difference equations are all examples of Fig. 24, the basic horseshoe. The cross-section of some flow may appear very complicated, with many of the outer layers of the Cantor set visible [31] but in *each case* the observation is that any three-dimensional flow can be reduced to some combination of the dynamic forms of Section XIII.

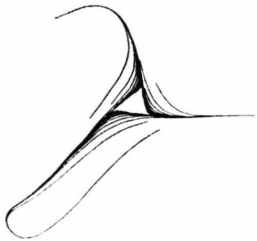


Fig. 45

XII. Strange Attractor Phenomenology

A fact that has emerged in the recent study of nonlinear equations is that even very simple systems are capable of very complicated behavior. For example, Eqn. 38, a variant of the driven Van der Pol, has only three independent parameters, yet within this three dimensional parameter space there are literally hundreds of regions with qualitatively different behavior. It is clearly going to take time and the work of many people to completely describe even a few systems. There are, however, a few general features which have recurred in nearly all of the systems studied. The generalizations below are based on many hours of analog and

digital computer observations on dozens of nonlinear systems.

1) *Strange attractor solutions are commonplace.* More often than not, in the analog simulation of nonlinear systems with three or more degrees of freedom, chaotic regimes with no observed periodicity are observed in some range of parameter values. All that is required is a compact flow which repetitively deforms some volume of phase space into a Smale horseshoe or other form, as discussed in the preceding section.

It seems increasingly clear that this commonness of occurrence extends into the physical world as well. Many "open systems" which take "work" energy from the surroundings (usually expanding some volume of phase space) and then dissipate this work as heat at a later time (contracting the phase space volume), may have the capability of forming a horseshoe or other stochastic form. Take for example the system governed by the equations which appear below:

$$\begin{aligned}\ddot{x} + b\dot{x} + kx &= \varepsilon xy, \\ \ddot{y} - b^1\dot{y} + k^1y &= \varepsilon^1 x^2, \quad b, b^1, k, k^1 > 0.\end{aligned}\quad (40)$$

This system, suggested by Doug Keeley, consists of two nonlinearly coupled harmonic oscillators, one supplied with energy ("negatively damped"), and the other dissipating energy ("positively damped"). The system is general enough to be a reasonable model for a variety of physical systems. Chaotic behavior, as recognized by no observed periodicity in an analog simulation, is present in this system over a *wide* range of parameter values.

2) *Attractors are either "strange", or "not strange".* The equilibrium activity of a solution on an attractor is either chaotic, or it is periodic or stationary, and the divisions between these two types of behavior are usually sharp. This is understandable from the qualitative change we expect as the informational parameter $\bar{\lambda}_t$ changes from positive to negative: either new information is being produced, or it is not.

Chaotic behavior in an analog or digital simulation is recognized by the lack of any observed periodicity. It could be asked: "How do you know your chaotic solution is not a *periodic* solution with a very high period?" In so-called "axiom A" flows, in which neighboring trajectories are further separated by every portion of the flow, it can be

proved that there exist no stable periodic trajectories. Most of the examples before us, though, are not of this class, often containing both chaotic and periodic behavior, often interspersed as some parameter is varied. One way to meet the above objection is to compute $\bar{\lambda}_t$ operationally using (34). If the limit converges to a positive value, we can be sure no periodic behavior has occurred. The period clearly cannot be any higher than the total range of the flow divided by the uncertainty length, any periodic orbit whose basin of attraction is narrower at any point than the intrinsic noise in the physical mechanism realizing the simulation will not be observed.

Although the absence of indefinitely high period orbits has not been proven, the observational evidence for true chaotic behavior in many non-axiom A flows is strong. High period attracting orbits, when they do clearly exist, are associated with *narrow* parameter ranges (see discussion in May [4]), whereas the observed chaotic regimes are often associated with *wide* parameter windows. Even if an orbit will eventually be caught by a high period orbit (see section on “preturbulence” below), operational calculations of $\bar{\lambda}_t$ show the system to be performing its informational function in the interim.

3) “Preturbulent” and related behavior. In a recent paper Kaplan and Yorke [32] discuss a regime of the Lorenz equations in which the associated return map appears as below:

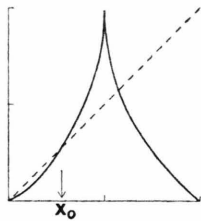


Fig. 46

A portion of this map has a slope less than one, and it has a stable fixed point at the origin. Any trajectory falling at points less than x_0 will be attracted to the origin. However, a trajectory initiated above x_0 may move chaotically for quite a while before becoming trapped in the region zero to x_0 . An observer would be unable to distinguish the operation of this object from a stable strange attractor, even

though it is technically a fixed point, all trajectories eventually falling to the origin. The parameter R in Eqs. (31) can be adjusted to make the trapping region from zero to x_0 as small as desired, resulting in arbitrarily long “half-lives”. The situation here is identical to that illustrated in Fig. 18, and an exponential decay time to the origin is likewise experimentally observed.

Kaplan and Yorke’s “preturbulence” could be described as a strange attractor with a certain probability of decaying into a fixed point. This probabilistic change of behavior is in fact a generally occurring phenomenon, also observed are a) decays of strange attractors into limit cycles, and even b) decays of strange attractors into other strange attractors. Both of these cases occur in analog simulations of equations, and the latter case is illustrated in Figure 47. By making very small adjustments in the parameters, the attractor of diagram A can be made the stable one, with orbits of attractor B eventually moving to it, or the reverse can be arranged. Also, as shown in diagram C, transitions may be possible in both directions, resulting in a single large attractor with two metastable states. One dimensional maps having these properties easily can be constructed [42].

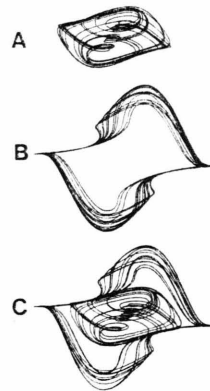


Fig. 47

4) An unbounded strange attractor. An example of a flow which stretches the definition of an “attractor” is given below:

$$\begin{aligned}\dot{x} &= -10x - 10y, \\ \dot{y} &= -10xz - y, \quad V = 13. \\ \dot{z} &= 10xy + V,\end{aligned}\tag{41}$$

Equations (41) above, found by Bill Burke and the writer, govern a flow containing an attracting set which is not bounded, and which has no fixed points. A solution $\dot{z} = \text{const}$ stretches to infinity along the z axis, and the action of the flow is to funnel solutions toward this singular solution. However this solution is unstable, although the *possibility* exists that a trajectory will pass arbitrarily far from the origin, the *probability* is that it will remain in the vicinity of the origin. The analog simulation (Fig. 48) thus resembles the Lorenz and Rössler examples in that it appears to be a two-dimensional sheet-like structure. The boundary, though, is not well defined, being set only by a probability density. Figure 48 shows a solution which has filled out a local region of this attractor which actually is non-compact, visible at the top center is a trajectory which will move farther from the origin than any preceding. Figure 49 shows a cross-section taken at the lower part of the attractor, displaying its funnel-like structure. This roll is squashed back into a sheet toward the top of the attractor.

The map $F(x) = |\log x|$, suggested by Bill Spence, is a simple example of a one-dimensional map which has this non-local character. Solutions are overwhelmingly likely to remain at small positive values, but the possibility does exist that a solution will reach an arbitrarily high number.

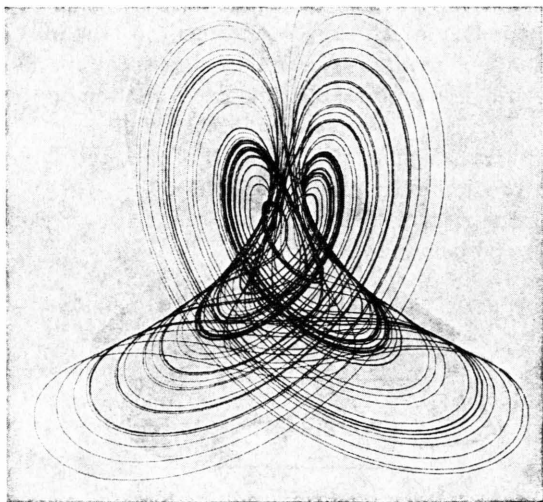


Fig. 48



Fig. 49

XIII. Informational Aspects of Turbulence

Hadamard [33] put forward three often-quoted criteria which solutions to differential equations should satisfy if they are to be physically meaningful:

- 1) The solutions should exist.
- 2) The solutions should be uniquely determined by the initial data.
- 3) The solutions should be stable, in the sense that small changes in the initial data produce only small changes in the solutions.

Fluid flow is a physical system governed by a set of partial differential equations, the Navier-Stokes equations, and for low flow speeds Hadamard's criteria seem well satisfied. For this "laminar" flow, the boundary conditions are time-independent, the energy supplied to the system is time-independent, and the solutions are correspondingly also time-independent. However, as the flow speed is increased, the solutions more or less abruptly become chaotic, with the fluid writhing in turbulent motion, even though the boundary conditions remain fixed. This is a *continuous* symmetry breaking which makes the "spontaneous symmetry breaking" of current interest in the study of phase transitions pale in comparison. This breakdown of Hadamard's second and third criteria has been discussed by Courant [34], and at some length, by Birkhoff [35]. Examining this situation, we are immediately led to ask:

What is the basic nature of the instability which causes turbulence?

Why does there seem to be a distinct change of state between laminar and turbulent flow?

Ruelle and Takens [8] have pointed out the similarity in the behavior of turbulent flow and strange attractors, suggesting that turbulence results from a strange attractor regime in the Navier-Stokes equations. Pursuing this analogy, we ask what light the considerations of this paper shed on the above questions. We are led to a *definition of turbulence in informational terms*.

An observer attempting to completely describe a given laminar flow would have a tedious but not intractable task. He could point-by-point examine the velocity field, and in principle record all this time independent data down to the resolution of his measuring instrument, and leave the room knowing that he had extracted all the information possible from the system.

However, if the flow speed were increased until turbulence set in, the same observer would be faced not only with the essentially impossible task of measuring the entire velocity field at each instant, but he could never leave the room, the flow would be a continuous source of information. Ream after ream of data could be taken, with his job of complete characterization no nearer completion.

We are faced with the same situation discussed earlier regarding strange attractors, *turbulent flows are an information source*. In fact, an estimate of the information production rate dH/dt of a turbulent flow can be made.

Simple dimensional analysis arguments yield length and time scales for the smallest eddies in a turbulent flow [36]. They are:

$$\text{length: } \eta = \left(\frac{\nu^3}{\varepsilon} \right)^{1/4} \quad \text{time: } \tau = \left(\frac{\nu}{\varepsilon} \right)^{1/2}, \quad (42)$$

where ν is the kinematic viscosity and ε is the energy dissipation rate per unit mass. Let us assume that the smallest eddies are strange attractors in some suitable function space, and further assume that turbulent eddies are all of roughly the same form, differing only in the direction of their axis of rotation. The high viscosity on the smallest length scales makes this somewhat plausible. Then a measurement of some particular eddy, say whether its vorticity is right- or left-handed around some chosen axis, will yield a random bit of information. Because the information increases only as the *log* of the number of states distinguishable in a measurement, our estimate won't be affected much by imagining a more precise measurement. The axis of

rotation of some eddy would be expected to change at a rate governed by the time scale τ . Thus the information production rate of a single smallest-scale turbulent eddy will be roughly:

$$1/\tau = (\varepsilon/\nu)^{1/2} \text{ bits/sec.} \quad (43)$$

When the strange attractor is found in the Navier-Stokes equations, we can feel sure that the associated $\bar{\lambda}_i$ will be of about this magnitude.

Using the approximate relation $\varepsilon \sim u^3/l$, where u and l are the velocity and length scales of the *largest* eddies in the flow, we can get an order of magnitude estimate for the information production rate of a turbulent fluid in a container of size l :

$$\frac{dH}{dt} \sim \left(\frac{l^7 u^{15}}{\nu^{11}} \right)^{1/4} \text{ bits/sec.} \quad (44)$$

A cup of coffee stirred in the morning generates on the order of 10^{12} bits per second of information. This is a huge number, but it should be noted that it is much smaller than the *microscopic* information loss associated with the entropy increase in the same cup of coffee. This is given by the relation of Szilard [37]:

$$\frac{dE}{dt} = T \frac{ds}{dt} = \frac{kT}{\ln 2} \frac{dH}{dt}. \quad (45)$$

This amounts to about 10^{18} bits per second.

The chief qualitative difference between laminar and turbulent flow is in the direction of information flow between the macroscopic and microscopic length scales. In laminar flow, motion is governed by boundary and initial conditions, no new information is generated by the flow, hence the motion is in principle predictable. Turbulent motion on the other hand is governed by information *generated continuously by the flow itself*, this fact precludes both predictability and reversibility.

At the risk of belaboring the point, we present Figure 50 below:

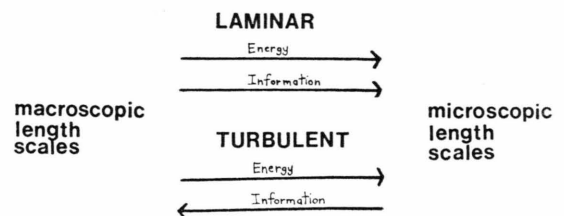


Fig. 50

Entropy increases in both laminar and turbulent systems, that is, energy in both cases moves from macroscopic to microscopic degrees of freedom. But *causality* in a dissipative system can act either way. The precise relationship between the entropy, which governs microscopic information, and the informational parameter λ_t , which governs information on the macroscopic scales, has yet to be made clear. This point is emphasized by the fact that even conservative nonlinear systems can have a positive dH/dt , though their entropy change is formally zero [11].

The concept of a change in the direction of information flow between the macro- and microscales makes clear both the nature of the transition to turbulence, and the reason why the transition is a sharp one. In fact, there has been observed what is seemingly an intermediate regime at Reynolds numbers near the transition, for instance, fluid flow past a point in a pipe may consist of alternating “slugs” of laminar and turbulent fluid. Some component of the velocity field will then be an example of an “intermittent variable”, schematically illustrated below [38]:

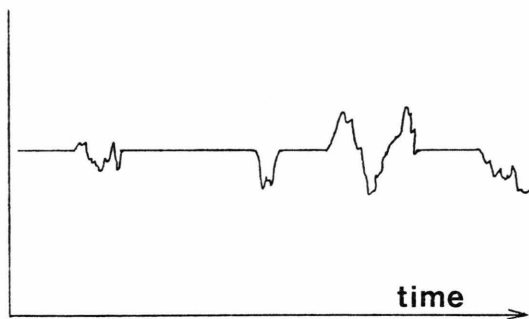


Fig. 51

The informational interpretation of this behavior is that the appropriate informational parameter is swinging from positive to negative values, the motion of the flow is being governed alternately by the fixed pipe walls and the chaotic heat bath.

The fact that small causes have large effects in turbulent flow is of course well known to students of fluid mechanics. Lorenz elegantly dubs this feature “the butterfly effect”: even if a complete solution of the equations of atmospheric motion were known at some time, the perturbation of a single butterfly would cause the solution to diverge in finite time.

The literature often refers to turbulence as being “extremely sensitive to initial conditions”.

We however feel that this way of viewing the chaotic nature of turbulence, as due somehow to sensitivity to initial conditions, obscures somewhat the essential character of turbulence, that of *continuous* information generation intrinsic to the flow itself.

XIV. $1/f$ Noise, Concluding Speculations

We have argued that chaotic behavior is not only a common feature of the solutions of nonlinear equations, but may be *completely ubiquitous* in the physical world. We have further claimed that the main physical function of these systems is to take information from microscopic length scales (“heat” degrees of freedom) and project it up to macroscopic expression. If this is the case, we might expect that the *noise spectrum* of a physical system containing one or more strange attractors might differ from the “white” or frequency independent noise power distribution which one would expect from a statistically equal distribution of the “heat” energy among all the degrees of freedom.

In fact, in many systems there is observed a mysterious noise with a characteristic $1/f$ power spectrum. This $1/f$ noise, also called “flicker noise” or “excess noise” (i.e., excess over white noise) has been observed in electronic components, quasi-stellar objects, tidal flows, and even in patterns of music and speech [39]. Any mechanism which would purport to explain such widespread phenomena must itself occur very generally.

We suggest that strange attractors provide such a mechanism. Although in the absence of a detailed demonstration that strange attractors or ensembles of strange attractors have a $1/f$ power spectrum this remains very much in the realm of speculation, we present below a few arguments in support of this view.

1) *$1/f$ noise is present only in driven systems.* A passive carbon resistor supplies only white Johnson noise, but a resistor with a *current* running through it produces $1/f$ noise in addition. It is clear from purely thermodynamic considerations that a source of “work” energy is required for the production of the additional information represented by $1/f$ noise. In our interpretation, the partial differential equations governing current and heat flow in the resistor

contain strange attractor solutions, and are the mechanism for this additional information production.

2) *Strange attractor solutions contain low frequency components.* Phillip Holmes [24] has recently measured the power spectrum of a nonlinear oscillator containing a strange attractor. The observed spectrum, although not exactly $1/f$, does show a steep rise toward the low end of the spectrum. Experiments in this laboratory on a variant of the Lorenz equations, done on an analog computer with the basic cycling frequency of the system scaled up to 10,000 to 20,000 Hz, show substantial low frequency components at well below 100 Hz. These low frequency components seem to be associated with fixed points and unstable periodic trajectories in the flow (see Figure 35). Trajectories passing close to such points can be delayed by an arbitrary time before rejoining the random motion of the flow, this may be the source of the large low frequency component.

The specific examples of strange attractors studied in this paper are almost entirely solutions of simple ordinary differential equations in three dimensions, the minimum dimensionality in which such behavior is possible. It is far from clear that exactly analogous behavior should occur in infinite dimensional partial differential equations, but we have used this assumption in our discussion of $1/f$ noise, and it is implicit in the Ruelle-Takens picture of turbulence. Let us embrace this assumption, and examine the general world picture to which it leads.

A partial differential equation governing the motion of some general flow does not usually describe the system at a molecular level. Only macroscopic variables enter into the equations. If we examine the causal links between events in the flow, we will find that the information flows in the spatial sense along definite trajectories, the "characteristics" of the equation, these paths often will be quite different from the paths of actual fluid elements. Now what are the consequences of the existence of attractors in the flow? If there is a stable fixed point or limit cycle in the flow, information impinging on it in the form of a "perturbation" emanating from some event will be absorbed as the attractor relaxes to its stable state. The fact that the "event" took place will have no effect on the system. If, on the other hand, there is an active strange attractor

present, its chaotic motion can be the source of new information which, travelling out along the characteristics, *initiates* events. Thus, areas in the flow containing a positive or negative dH/dt will be marked by positive or negative *divergences of characteristics*. This general situation is diagrammed below:

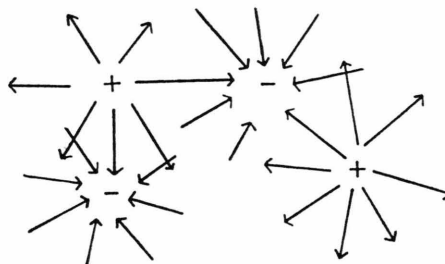


Fig. 52

Under this view, if "the world" can be regarded as a flow governed by some immense partial differential equation, then *information* moves across the face of the world along the characteristics of the system, from *sources* in regions of the flow where $dH/dt > 0$ to *sinks* where $dH/dt < 0$. Very generally, then, we should expect to find "excess noise" emanating from the sources. "Events" initiated by this "noise" are inherently *unpredictable*, caused as they are by random motions of the heat bath.

A philosophical implication of this point of view is that the nineteenth century view of the world as a machine which, if all positions and momenta could be measured, is completely determined is not only wrong in the small, where quantum mechanical uncertainty limits our knowledge of the exact state of a system, but it is wrong in the large. New information is continuously being injected into the *macroscopic* degrees of freedom of the world by every puff of wind and swirl of water.

Thus the ceaseless and tumultuous flow of events in the world reflects in a very direct way the chaotic motion of the heat bath. The constant injection of new information into the macroscales may place severe limits on our predictive ability, but it as well insures the constant variety and richness of our experience.

"All things are water."

Thales

Acknowledgements

I would like to thank Bill Burke, Jim Crutchfield, Doyne Farmer, Norman Packard, Bruce Rosenblum, Chris Shaw and Share Tringali for valuable discussions, suggestions, and support. Thanks are also due Ralph Abraham, John Guckenheimer, and Dan Sunday for advice on mathematical topics. The manuscript was prepared by Patty Burns, and Figs. 8, 24, 25, 36, 37 and 42 were drawn by Chris Shaw.

This work is respectfully dedicated to the memory of Arthur L. Shaw (1887–1977), past president, New England Water Works Association.

Appendix I

Dynamical System as a Simple Noise Amplifier

The equation appearing below governs a harmonic oscillator which is damped during half its cycle, and driven during the other half:

$$\ddot{x} + b|\dot{x}| + kx = 0. \quad (\text{A1})$$

This has a general solution:

$$x = A e^{\pm b/2t} \cos \omega_1 t$$

$$\text{where } \omega_1 = \left(k - \frac{b^2}{4}\right)^{1/2}. \quad (\text{A2})$$

An analog simulation is displayed below.



Fig. A1

Two solutions initiated on the left side of the figure cannot be restricted to be less than Δx apart, where Δx is determined by the intrinsic noise of the system implementing the above equations. This uncertainty length is mapped into a larger distance x' on the right side. (A small nonlinearity was added to the simulation to stabilize the average trajectory position.) Random fluctuations are amplified up to a larger length scale, the ratio $\Delta x'/\Delta x$ is given by $e^{b\pi/2\omega_1}$. This ratio can be brought to an arbitrarily high value by adjusting b , and this allows an easy measurement of the intrinsic noise. This was the method used to determine the intrinsic noise in the

simulation of the Lorenz equations discussed in Sects. VII and VIII.

The amount of random information observable on the right hand side of the flow is given by the Lie derivative integrated over the time it takes the trajectory to get from the left to the right side. The Lie derivative here is just the constant b , so the new information observable per cycle is:

$$\Delta h = \frac{b\pi}{\omega_1 \ln 2} \text{ bits per cycle.} \quad (\text{A3})$$

This system does have an informational function, an observer on the right hand side of the flow could use the information coming up from the heat bath to assemble a table of random numbers. But the information so generated does *not* govern the succeeding path of the trajectory, as is the case with strange attractors.

Appendix II

Evolution of Probability Densities

We have shown that flows with a positive $\bar{\lambda}$ continuously add new information to a particular solution, rendering the concept of “exact solution” useless. Even in the simple classical systems described in this paper, the description of trajectory positions in terms of probability densities is forced upon us. The problem of loss of initial data is thus stated as follows: Given a known initial probability distribution, how does it relax to the equilibrium probability distribution characteristic of the average long-time behavior of the system?

For our “system” we once again imagine a mechanism which implements a given one-dimensional map $y = F(x)$. An initial condition is now describable as a sharply peaked probability density $P_0(x)$. This probability density will be transformed by each iteration of the map, producing a series of distributions $P_0(x)$, $P_1(x)$, $P_2(x)$, ... In Fig. A2 (next page) we illustrate this process with the specific map $y = 4x(1 - x)$. As was discussed in Sect. IV, this map has a $\bar{\lambda}$ of 1 bit per iteration, and an equilibrium probability distribution, which we now write as \bar{P} , of

$$\bar{P}(x) = \frac{1}{\pi} [x(1 - x)]^{-1/2}. \quad (\text{A4})$$

After some number of iterations the initially sharply peaked distribution relaxes to this function.

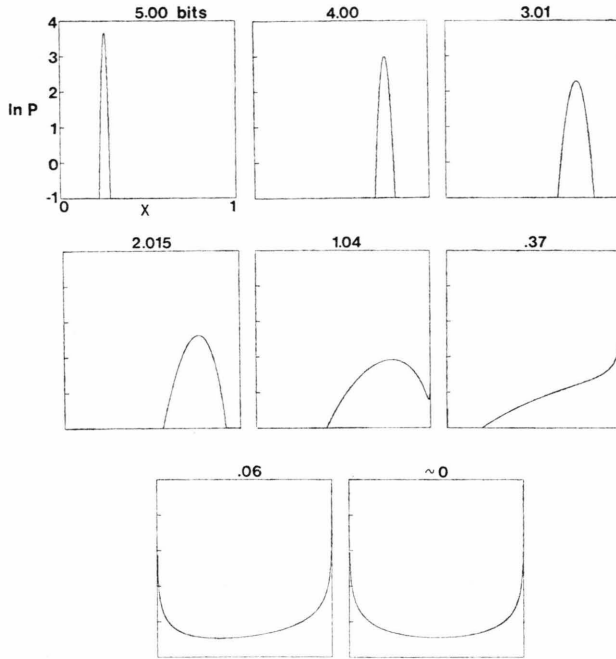


Fig. A2

To quantify this process, we need the continuum generalization of the definition of information:

$$H = - \sum_i P_i \log P_i. \quad (\text{A5})$$

P_i is the probability of finding the system in some state i . In the context of one-dimensional maps, this is given by the Boltzmann integral:

$$H = \int_0^1 P(x) \log_2 P(x) dx. \quad (\text{A6})$$

If we have obtained some knowledge of a trajectory position, for instance, by performing an observation, and this knowledge is expressed in the form of a probability density $P(x)$, then Eq. (A6) gives us the informational value of that knowledge in bits. This assumes that our a priori expectation in the absence of the knowledge $P(x)$ is that the trajectory could be anywhere in the interval with equal probability, and Eq. (A6) goes correctly to zero when our “knowledge” is only that $P(x) = 1$.

But what if we have some other expectations? The map $y = 4x(1-x)$, for instance, has an expected distribution $\bar{P}(x)$ given by (A4). The correct generalization of Eq. (A6) is:

$$H(P, \bar{P}) = \int_0^1 P(x) \log \frac{P(x)}{\bar{P}(x)} dx. \quad (\text{A7})$$

This is a well-known result in information theory [43]. $H(p, \bar{p})$ is often referred to as the “information gain”: what we gain by learning $P(x)$ when our a priori expectation is $\bar{P}(x)$. It can be justified by considering the coordinate transformation $x' = x'(x)$ which will take our a priori expectation $\bar{P}(x)$ into a constant, $\bar{P}'(x') = 1$. In these primed coordinates, Eq. (A6) will hold.

The map $y = 4x(1-x)$ is one of the rare cases where such a transformation can be found *globally* [7]. The transformation $x' = 2/\pi \sin^{-1} \sqrt{x}$ takes the parabolic map into the symmetric tent map of Figure 12. However, even for an arbitrary map, conservation of probability determines the transformation *locally* at each small interval. We have:

$$P'(x') dx' = P(x) dx. \quad (\text{A8})$$

Requiring $P'(x') = 1$ and transforming Eq. (A6) yields (A7).

This latter equation allows us to precisely compute the informational value of an initial condition:

$$H_{\text{in}} \equiv H(P_0, \bar{P}) = \int_0^1 P_0 \log \left[\frac{P_0}{\bar{P}} \right] dx. \quad (\text{A9})$$

In Fig. A2 we chose for $P_0(x)$ a gaussian with a width such that Eq. (A9) yielded 5 bits. This width is roughly equal to 2^{-5} . The informational value is computed for each iteration and displayed on the figure. $\bar{\lambda}$ correctly gives the rate of loss of the initial data. The characteristic number of iterations this data will persist is given as before by:

$$n = H_{\text{in}} / \bar{\lambda}. \quad (\text{A10})$$

Only at the very end of this process does the information loss rate deviate from $\bar{\lambda}$, when size effects become important, i.e., states can be occupied from the initial conditions by more than one route.

Note the conceptual parallel to quantum mechanics: we perturb the system by observing it, the system then relaxes to an equilibrium characterized by a minimum knowledge of the system state.

There are a few subtleties in this process of loss of initial data, information is only defined *with respect to an a priori expectation*, and this must be carefully agreed upon for calculations to make sense.

To demonstrate, we recall the two observers, and give them a copy of the specification $y = 4x(1-x)$ so that they both have complete knowledge of the mechanism. The early man starts off the system

with some initial probability distribution P_0 and subsequent iteration produces the series $P_0, P_1, P_2, \dots, \bar{P}$. The late man then examines the mechanism after some number of iterations and can observe only the most recent P_i (we might imagine the experiment repeated many times). If he knows *when* the mechanism was initiated, his a priori expectation is *not* necessarily \bar{P} . For instance, if he enters the room and checks the mechanism before it has performed *any* iterations, his expectation is just $P(x)=1$, equal probability over the interval. As time passes, *his expectations are also transformed* by the action of the map, generating a sequence $P_0', P_1', \dots, \bar{P}$ where $P_0' = 1$. In this instance, the information he gains by knowing the starting time is small, application of Eq. (A7) yields

$$H(1, \bar{P}) = \log_2(\pi/e) = 0.2088 \text{ bits}. \quad (\text{A11})$$

As illustrated in Fig. A3, the flat distribution quickly relaxes to $\bar{P}(x)$. In the interim, however, the information with respect to *his expectation* is given by:

$$H(P_i, P_i') = \int_0^1 P_i \log \frac{P_i}{P_i'} dx \quad (\text{A12})$$

and *not* $H(P_i, \bar{P})$. The numbers in the two series $H(P_i, P_i')$ and $H(P_i, \bar{P})$ differ somewhat in their first few terms.

The map $y = 4x(1-x)$ is known to be ergodic over the entire interval, that is, the image of any finite sub-interval will eventually shadow the whole interval. Thus, for this example, any initial probability distribution will converge to the \bar{P} of (A4). It is clear that this is not the general case, an extreme example being when an interval contains disjoint basins, as in Figure 22. The information as to where in the interval the trajectory was initiated is propagated indefinitely far into the future. Another example is a periodic orbit, or a distribution with disjoint sections which are visited in some sequence. Here *phase* information is propagated into the future. In this case though, we can define a single \bar{P} characteristic of minimum knowledge of the system state. The sequence of probability functions $P_i(x)$ will become periodic, and by averaging over this period we can remove the phase information.

The fact that the series $H(P_i, P_i')$ of Eq. (A12), obtained by acting with the map on *both* observed and expected probability densities, *must* monotoni-

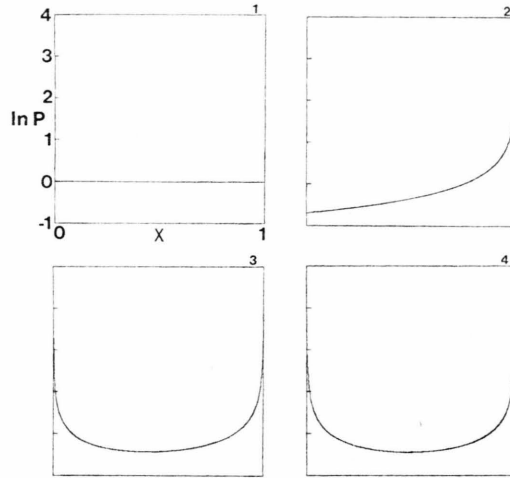


Fig. A3

cally decrease follows from first principles. Information cannot increase in the absence of observations. (This at least is the prejudice of the physicist.) The series will converge to a positive number in exactly those cases when the system is capable of carrying information into the indefinite future, and to zero otherwise.

The physical meaning of the information gain was recognized by Schlögl in a 1971 paper, in the context of master equations [40]. It can be proven that a master equation, where the laws of motion are explicitly probabilistic, converges to a single time-independent probability distribution \bar{P}_i , where the index i now ranges over the states described, as long as all states are connected by finite transition probabilities [41]. Schlögl showed that the information gain:

$$H(P_i(t), \bar{P}_i) = \sum P_i(t) \log \left[\frac{P_i(t)}{\bar{P}_i} \right] \quad (\text{A13})$$

is a *Liapunov function* for the master equation, guaranteeing the monotonic convergence of $P_i(t)$ to \bar{P}_i . In our case, the governing equations are deterministic, if viewed from point-to-point context, with the spreading of trajectories due to a *topology*, rather than explicit probability transitions, and the equilibrium probability distribution can oscillate.

We now respond to a criticism of the nature of the function $\bar{P}(x)$. If a given one-dimensional map has a point where the slope is zero, i.e., has a smooth hump, then a denominator of Eq. (11) will have a zero. Thus any continuous probability distribution

will get mapped into one with a singularity in it. A system with a positive $\bar{\lambda}$ will usually not subsequently map this singularity back into itself, hence many spikes in $\bar{P}(x)$ of infinite height will be produced by repeated iterations. How do we reconcile the picture of a continuous probability distribution with all these spikes?

Once again, the uncertainty length Δh resolves the difficulties. To any finite resolution, there is no difference between delta function spread infinitesimal distances apart, and a continuous distribution. $\bar{P}(x)$ may not be a continuous function in the sense of present-day analysis, but it has a clear operational definition with respect to the procedure described in Section IV. The result of that procedure for negative $\bar{\lambda}$ is at least one isolated spike of minimum width Δh . For positive $\bar{\lambda}$, it is a series of num-

bers spaced at distances Δh along some part of the interval, which can be fit by a curve $\bar{P}(x)$. This curve may be quite complicated in appearance (see Fig. A4 below), but there is no problem in its computation.

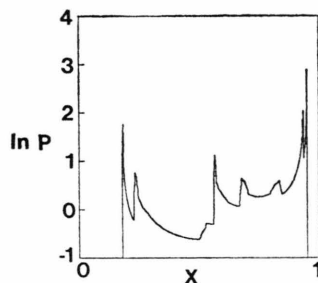


Fig. A4. Equilibrium probability distribution for the map $y = 3.8x(1-x)$.

- [1] E. N. Lorenz, *J. Atmos. Sci.* **20**, 130 (1963).
- [2] E. N. Lorenz, *N. Y. Acad. Sci. Trans.* **P. 409** (1963).
- [3] O. E. Rössler, *Phys. Lett.* **57A**, 397 (1976).
- [4] R. M. May, *Nature London* **261**, 459 (1976).
- [5] J. Guckenheimer, G. Oster, and A. Ipaktchi, *J. Math. Biol.* **4**, 101 (1977).
- [6] M. Hénon, *Comm. Math. Phys.* **50**, 69 (1976).
- [7] P. R. Stein and S. M. Ulam, *Rosprawy Matematyczne* **39**, 401 (1964).
- [8] D. Ruelle and F. Takens, *Comm. Math. Phys.* **20**, 167 (1971).
- [9] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication", University of Illinois Press, 1962.
- [10] D. Gabor, *J. Inst. Electr. Eng.* **93**, 429 (1946).
- [11] G. Benettin, L. Galgani, and J. Strelcyn, *Phys. Rev. A* **14**, 2338 (1976).
- [12] S. Smale, *Bull. Amer. Math. Soc.* **13**, 747 (1967).
- [13] J. L. Yorke, *Int. Conf. on Bifurcation Theory*, Nov. 1977, New York.
- [14] S. D. Feit, *Characteristic Exponents and Strange Attractors*, preprint (1978).
- [15] F. C. Hoppensteadt and J. M. Hyman, *Siam J. Appl. Math.* **32**, 73 (1977).
- [16] V. I. Oseledec, *Trans. Moscow Math. Soc.* **19**, 197 (1968).
- [17] D. Ruelle, *Int. Conf. on Bifurcation Theory*, Nov. 1977, New York.
- [18] D. Ruelle, *Proc. Int. Math. Physics Conf. of Rome*, 1977.
- [19] H. R. Lewis and C. H. Papadimitriou, "The Efficiency of Algorithms", *Sci. Amer. Jan.* 1978.
- [20] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, Benjamin 1968.
- [21] R. F. Williams, *The Structure of Lorenz Attractors*, Berkeley Turbulence Seminar, 1976.
- [22] V. I. Arnold, "Ordinary Differential Equations", MIT Press 1973.
- [23] Typical parameter values: $\dot{x} = -x(x^2 - 1) + y + A \sin \omega t$; $A = .63$; $\dot{y} = -.1x$; $\omega = 1.2$.
- [24] P. Holmes, "A Nonlinear Oscillator with a Strange Attractor" preprint, Cornell University (1977).
- [25] Parameters of Eqn. 38 for Figure 38: $k = .7$, $\mu = 10$, $a = .1$, $A = .25$, $\omega = 1.57$.
- [26] R. Thom, "Structural Stability and Morphogenesis", Benjamin, 1975.
- [27] Flow of Figure 39 given by Eqn. 41 with $V = 48$. The schematic figure is meant only to indicate the type of folding and, like Figure 28, describing the Rössler attractor, does not accurately reflect the fact that the folding is partial.
- [28] D. Ruelle, "The Lorenz Attractor and the Problem of Turbulence" Conference Report, Bielefeld, 1975.
- [29] Parameters of Eqn. 38 for Figure 40: $k = .63$, $\mu = 10$, $a = .07$, $A = .4$, $\omega = 2$.
- [30] R. Bridges and G. Rowland, *Phys. Lett.* **63A**, 189 (1977).
- [31] Figure 45 is given by Eqn. 38 with parameters: $k = .37$, $\mu = 7.3$, $a = .07$, $A = .5$, $\omega = 1.7$.
- [32] J. L. Kaplan and J. A. Yorke, "Preturbulence: A Regime Observed in a Fluid Flow Model of Lorenz", preprint, University of Maryland (1977).
- [33] J. Hadamard, "Lectures on Cauchy's Problem", Yale University Press, 1923.
- [34] R. Courant, *Proc. Int. Conf. of Math.* 1950 Vol. II, P. 277.
- [35] G. Birkhoff, "Hydrodynamics", Princeton University Press, 1960.
- [36] H. Tennekes and J. L. Lumley, "A First Course in Turbulence", MIT Press, 1972.
- [37] L. Szilard, *Z. Physik* **53**, 840 (1925), English translation in: *Behav. Sci.* **9**, 301 (1964); L. Brillouin, "Science and Information Theory", Academic Press, 1962; D. Gabor, *Progress in Optics*, Vol. I, P. 109 (1961).
- [38] D. C. Leslie, "Developments in the Theory of Turbulence", Clarendon Press, 1973.
- [39] R. F. Voss and J. Clarke, *Nature* **258**, 317 (1975); W. H. Press, "Flicker Noises in Astronomy and Elsewhere", preprint, Harvard College Observatory, (1977).
- [40] F. Schlögl, *Z. Physik* **243**, 303 (1971).
- [41] J. Schnakenberg, *Rev. Mod. Phys.* **48**, 571 (1976).
- [42] The approximate parameters of Eqn. 38 which yield the double attractor of Figure 47 are: $k = .657$, $\mu = 10$, $a = .062$, $A = .39$, $\omega = 1.73$.
- [43] S. Kullback, "Information Theory and Statistics", Wiley, 1951; S. Goldman, "Information Theory", Dover, 1953.