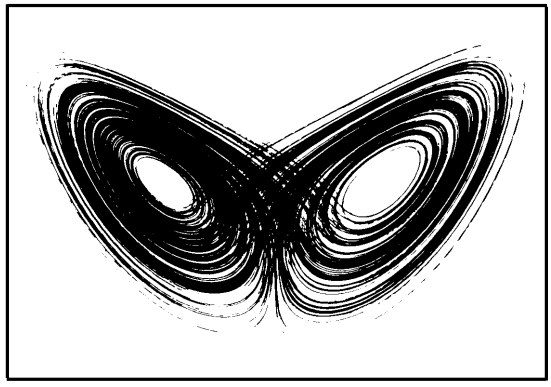
NONLINEAR SCIENCE

from Paradigms to Practicalities



by David K. Campbell

No tribute to the legacy of Stan Ulam would be complete without a discussion of "nonlinear science," a growing collection of interdisciplinary studies that in the past two decades has excited and challenged researchers from nearly every discipline of the natural sciences, engineering, and mathematics. Through his own research Stan played a major role in founding what we now call nonlinear science, and through his encouragement of the work of others, he guided its development. In this survey article I will try to weave the thread of Stan's contributions into the pattern of recent successes and current challenges of nonlinear science. At the same time I hope to capture some of the excitement of research in this area.

Introduction

Let me start from a very simple, albeit circular, definition: nonlinear science is the study of those mathematical systems and natural phenomena that are *not* linear. Ever attuned to the possibility of *bons mots*, Stan once remarked that this was "like defining the bulk of zoology by calling it the study of 'non-elephant animals'." His point, clearly, was that the vast majority of mathematical equations and natural phenomena are nonlinear, with linearity being the exceptional, but important, case.

Linear versus Nonlinear. Mathematically, the essential difference between linear

and nonlinear equations is clear. Any two solutions of a linear equation can be added together to form a new solution; this is the superposition *principle.* In fact, a moment of serious thought allows one to recognize that superposition is responsible for the systematic methods used to solve, independent of other complexities, essentially any linear problem. Fourier and Laplace transform methods, for example, depend on being able to superpose solutions. Putting it naively, one breaks the problem into many small pieces, then adds the separate solutions to get the solution to the whole problem.

In contrast, two solutions of a nonlinear equation *cannot* be added together to form another solution. Superposition fails. Thus, one must consider a nonlinear problem *in toto;* one cannot—at least not obviously—break the problem into small subproblems and add their solutions. It is therefore perhaps not surprising that no general analytic approach exists for solving typical nonlinear equations. In fact, as we shall discuss, certain nonlinear equations describing chaotic physical motions have *no* useful analytic solutions.

Physically, the distinction between linear and nonlinear behavior is best abstracted from examples. For instance, when water flows through a pipe at low velocity, its motion is *laminar* and is characteristic of linear behavior: regular, predictable, and describable in simple analytic mathematical terms. However, when the velocity exceeds a critical value, the motion becomes *turbulent,* with localized eddies moving in a complicated, irregular, and erratic way that typifies nonlinear behavior. By reflecting on this and other examples, we can isolate at least three characteristics that distinguish linear and nonlinear physical phenomena.

First, the motion itself is qualitatively different. Linear systems typically show smooth, regular motion in space and time that can be described in terms of well-behaved functions. Nonlinear systems, however, often show transitions from smooth motion to chaotic, erratic, or, as we will see later, even apparently random behavior. The quantitative description of *chaos* is one of the triumphs of nonlinear science.

Second, the response of a linear system to small changes in its parameters or to external stimulation is usually smooth and in direct proportion to the stimulation. But for nonlinear systems, a small change in the parameters can produce an enormous qualitative difference in the motion. Further, the response to an external stimulation can be different from the stimulation itself: for example, a periodically driven nonlinear system may exhibit oscillations at, say, one-half, one-quarter, or twice the period of the stimulation.

Third, a localized "lump," or pulse, in a linear system will normally decay by spreading out as time progresses. This phenomenon, known as dispersion, causes waves in linear systems to lose their identity and die out, such as when low-amplitude water waves disappear as they move away from the original disturbance. In contrast, nonlinear systems can have highly coherent, stable localized structures—such as the eddies in turbulent flow-that persist either for long times or, in some idealized mathematical models, for all time. The remarkable order reflected by these persistent coherent structures stands in sharp contrast to the irregular, erratic motion that they themselves can undergo.

Togo beyond these qualitative distinctions, let me start with a very simple physical system—the plane pendulum—that is a classic example in at least two senses. First, it is a problem that all beginning students solve; second, it is a classic illustration of how we mislead our students about the prevalence and importance of nonlinearity.

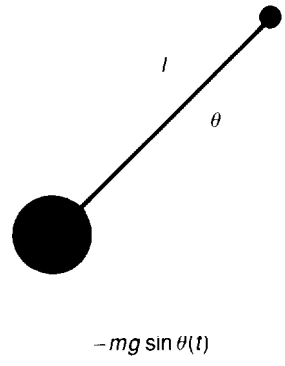
THE SIMPLE PENDULUM

Applying Newton's law of motion to the plane pendulum shown in Fig. 1 yields an ordinary second-order differential equation describing the time evolution:

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{l}\sin\theta(t) = 0, \qquad (1)$$

where θ is the angular displacement of the pendulum from the vertical, l is the length of the arm, and g is the acceleration due to gravity. Equation 1 is obviously nonlinear

Fig. 1. It can be seen that a nonlinear equation describes the motion of the simple, plane pendulum when, in accordance with Newton's force law, the component of the gravitational force in the angular direction, $-mg\sin\theta(t)$, is set equal to the rate of change of the momentum, $ml\,d^2\theta(t)/dt^2$, in that direction.



because $\sin(\theta_1 + \theta_2) \neq \sin\theta_1 + \sin\theta_2$.

What happens, however, if we go to the regime of small displacements? The Taylor expansion of $\sin\theta$ ($\approx \theta - \frac{\theta^3}{6} + \cdots$) tells us that for small θ the equation is approximately linear:

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{l}\theta(t) \approx 0. \tag{2}$$

The general solution to the linear equation is the superposition of two terms,

$$\theta(t) = \frac{1}{\omega}\left(\frac{d\theta}{dt}\right)_0 \sin\omega t + \theta_0 \cos\omega t, \tag{3}$$

where θ_0 and $(d\theta/dt)_0$ are the angle and angular velocity at the initial time and the frequency ω is a constant given by $\omega = \sqrt{g/l}$.

Equation 3 is the mathematical embodiment of Galileo's famous observation that the frequency of a pendulum is independent of its amplitude. But in fact the result is a consequence of the linear approximation, valid only for small oscillations. If the pendulum undergoes very large displacements from the vertical, its motion enters the nonlinear regime, and one finds that the frequency depends on amplitude, larger excursions having longer periods (see "The Simple But Nonlinear Pendulum"). Of course, grandfather clocks would keep terrible time if the linear equation were not a good approximation; nonetheless, it remains an approximation, valid only for small-amplitude motion.

The distinction between the full nonlinear model of the pendulum and its linear approximation becomes substantially more striking when one studies the pendulum's response to an external stimulus. With both effects of friction and a periodic driving force added, the pendulum equation (Eq. 1) becomes

$$\frac{d^2\theta}{dt^2} + \alpha\frac{d\theta}{dt} + \frac{g}{l}\sin\theta = \Gamma\cos\Omega t, \tag{4}$$

where α is a measure of the frictional force and Γ and Ω are the amplitude and frequency, respectively, of the driving force. In the regime of small displacements, this reduces to the linear equation

$$\frac{d^2\theta}{dt^2} + \alpha\frac{d\theta}{dt} + \frac{g}{l}\theta \approx \Gamma\cos\Omega t. \tag{5}$$

A closed-form solution to the linear equation can still be obtained, and the motion can be described analytically for all time. For certain values of α, Γ, and Ω, the solution to even the nonlinear equation is periodic and quite similar to that of the linear model. For other values, however, the solution behaves in a complex, seemingly random, unpredictable manner. In this chaotic regime, as we shall later see, the motion of this very simple nonlinear system defies analytic description and can indeed be as random as a coin toss.

Dynamical Systems: From Simple to Complex. Both the free pendulum and its damped, driven counterpart are particular examples of *dynamical systems. The* free pendulum is a *conservative* dynamical system-energy is constant in time—whereas the damped, driven pendulum is a *dissipative* system-energy is not conserved. Loosely speaking, a dynamical system can be thought of as anything that evolves in time according to a well-defined rule. More specifically, the variables in a dynamical system, such as *q* and *p,* the canonical position and momentum, respectively, have a rate of change at a given time that is a function of the values of the variables themselves at that time: $\dot{q}(t) = f\left(q(t), p(t)\right)$ and $\dot{p}(t) = g\left(q(t), p(t)\right)$ (where a dot signifies differentiation with respect to time). The abstract "space" defined by these variables is called the *phase*

The Simple but NONLINEAR PENDULUM

Elementary physics texts typically treat the simple plane pendulum by solving the equation of motion only in the linear approximation and then presenting the general solution as a superposition of sines and cosines (as in Eq. 3 of the main text). However, the full nonlinear equation can also be solved analytically in closed form, and a brief discussion of this solution allows us to illustrate explicitly several aspects of nonlinear systems.

It is most instructive to start our analysis using the Hamiltonian for the simple pendulum, which, in terms of the angle θ (a generalized coordinate) and the corresponding (generalized) momentum $p_\theta \equiv ml^2 \frac{d\theta}{dt}$, has the form

$$H(p_\theta, \theta) = \frac{p_\theta^2}{2ml^2} - mgl\cos\theta. \qquad (1)$$

Using the Hamiltonian equations

$$\frac{\partial H}{\partial p_\theta} = \frac{d\theta}{dt},$$

$$\text{and} \quad \frac{\partial H}{\partial \theta} = -\frac{dp_\theta}{dt},$$

we obtain (after substituting for p_θ) an equation solely in terms of the angle θ and its derivative:

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0. \qquad (2)$$

Recognizing that $\frac{d\theta}{dt}\frac{d^2\theta}{dt^2} = \frac{d}{dt}(\frac{1}{2}(\frac{d\theta}{dt})^2)$ and $\frac{d\theta}{dt}\sin\theta = \frac{d}{dt}(-\cos\theta)$, we see that Eq. 2 can be converted to a perfect differential by multiplying by $d\theta/dt$:

$$\frac{d\theta}{dt}\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta\frac{d\theta}{dt} = \frac{d}{dt}\left(\frac{1}{2}\left(\frac{d\theta}{dt}\right)^2 - \frac{g}{l}\cos\theta\right) = 0. \qquad (3)$$

Hence, we can integrate Eq. 3 immediately to obtain

$$\frac{1}{2}\left(\frac{d\theta}{dt}\right)^2 = \frac{g}{l}\cos\theta + C. \qquad (4)$$

By comparing Eqs. 1 and 3 and recalling the definition of p_θ, we see that

$$C = \frac{H(p_\theta, \theta)}{ml^2}.$$

That the constant C is proportional to the *value* of the Hamiltonian, of course, is just an expression of the familiar conservation of energy and shows that the value of the conserved energy determines the nature of the pendulum's motion.

Restricting our considerations to librations—that is, motions in which the pendulum oscillates back and forth without swinging over the top of its pivot point—we can evaluate C in terms of θ by using the condition $d\theta/dt = 0$ when $\theta = \theta_{max}$, which yields

$$C = -\frac{g}{l}\cos\theta_{max}.$$

This, in turn, means that

$$\frac{d\theta}{dt} = \sqrt{\frac{2g}{l}(\cos\theta - \cos\theta_{max})}. \qquad (5)$$

The full period of the motion T is then the definite integral

$$T = 4\int_0^\theta \frac{d\theta}{\sqrt{\frac{2g}{l}(\cos\theta - \cos\theta_{max})}}. \qquad (6)$$

This last integral can be converted, via trigonometric identities and redefinitions of variables, to an elliptic integral of the first kind. Although not as familiar as the sines and cosines that arise in the linear approximation, the elliptic integral is tabulated and can be readily evaluated. Thus, the full equation of motion for the nonlinear pendulum can be solved in closed form for arbitrary initial conditions.

An elegant method for depicting the solutions for the one-degree-of-freedom system is the "phase plane." If we examine such a plot (see Fig. 2 in the main text), we see that the origin ($\theta = 0$, $d\theta/dt = 0$)—and, of course, its periodic equivalents at $\theta = \pm 2n\pi$, $d\theta/dt = 0$—represent *stable fixed points* with the pendulum at rest and the bob pointing down. The point at $\theta = \pi$, $d\theta/dt = 0$—and, again, its periodic equivalents at $\theta = \pm(2n+1)\pi$, $d\theta/dt = 0$—represent *unstable fixed points* with the pendulum at rest but the bob inverted; the slightest perturbation causes the pendulum to move away from these points. The closed curves near the horizontal axis ($d\theta/dt = 0$) represent librations, or *periodic oscillations. The* open, "wavy" lines away from the horizontal axis (large $|d\theta/dt|$) correspond to *unbounded motions* in the sense that 6' increases or decreases forever as the pendulum rotates around its pivot point in either a clockwise ($d\theta/dt < 0$) or a counterclockwise ($d\theta/dt > 0$) sense.

What about other systems? A dynamical system that can be described by *2N* generalized position and momentum coordinates is said to have N degrees of freedom. Hamiltonian systems that, like the pendulum, have only one degree of freedom can *always* be integrated completely with the techniques used for Eqs. 2-6. More generally, however, systems with N degrees of freedom are *not* completely integrable; Hamiltonian systems with N degrees of freedom that *are* completely integrable form a very restricted but extremely important subset of all N-degree-of-freedom systems.

As suggested by the one-degree-of-freedom case, complete integrability of a system with N degrees of freedom requires that the system have N constants of motion—that is, N integrals analogous to Eq. *4-and* that these constants be consistent with each other. Technically, this last condition is equivalent to saying that when the constants, or integrals of motion, are expressed in terms of the dynamical variables (as C is in Eq. 4), the expressions must be "in involution," meaning that the Poisson brackets must vanish identically for all possible pairs of integrals of motion. Remarkably, one can find nontrivial examples of completely integrable systems, not only for N-degree-of-freedom systems but also for the "infinite' '-degree-of-freedom systems described by partial differential equations. The sine-Gordon equation, discussed extensively in the main text, is a famous example.

In spite of any nonlinearities, systems that are completely integrable possess remarkable regularity, exhibiting smooth motion in all regions of phase space. This fact is in stark contrast to nonintegrable systems. With as few as one-and-a-half degrees of freedom (such as the damped, driven system with three generalized coordinates represented by Eq. 4 in the main text), a nonintegrable system can exhibit deterministic chaos and motion as random as a coin toss. ■

space, and its dimension is clearly related to the number of variables in the dynamical system.

In the case of the free pendulum, the angular position and velocity at any instant determine the subsequent motion. Hence, as discussed in "The Simple But Nonlinear Pendulum," the pendulum's behavior can be described by the motion of a point in the two-dimensional phase space with coordinates θ and $\dot{\theta}$ (Fig. 2). In the traditional parlance of mechanics, the free pendulum is a Hamiltonian system having "one degree

HARMONIC-OSCILLATOR PHASE SPACE

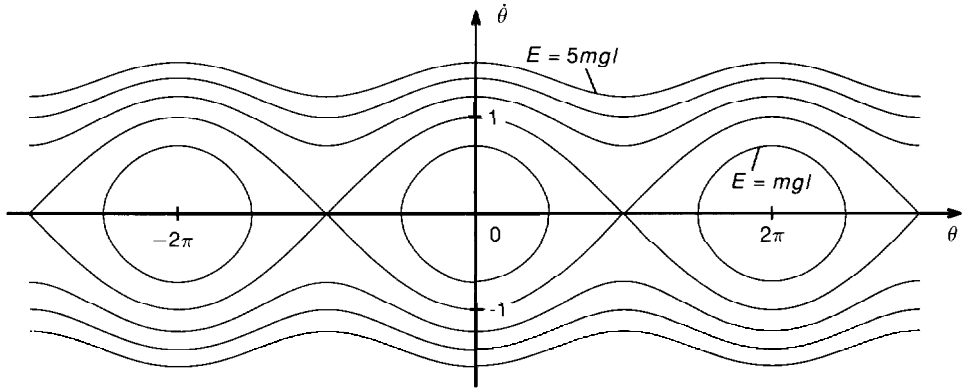


Fig. 2. The behavior of the simple pendulum is here represented by constant-energy contours in θ-$\dot{\theta}$ (roughly, position-momentum) phase space. The closed curves around the origin ($E < 2mgl$) represent librations, or periodic oscillations, whereas the open, "wavy" lines for large magnitudes of $\dot{\theta}$ ($E > 2mgl$) correspond to motions in which the pendulum moves completely around its pivot in either a clockwise ($\dot{\theta} < 0$) or counterclockwise ($\dot{\theta} > 0$) sense, causing θ to increase in magnitude beyond 2π. (Figure courtesy of Roger Eckhardt, Los Alamos National Laboratory.)

of freedom," since it has only one spatial variable (θ) and one generalized momentum (roughly, $\dot{\theta}$). Further, as discussed in the sidebar, this system is *completely integrable*, which in effect means that its motion for all time can be solved for analytically in terms of the initial values of the variables.

More typically, dynamical systems involve many degrees of freedom and thus have high-dimensional phase spaces. Further, they are in general *not* completely integrable. An example of a many-degree-of-freedom system particularly pertinent to our current discussion is the one first studied by Enrico Fermi, John Pasta, and Stan Ulam in the mid-fifties: a group of particles coupled together by nonlinear springs and constrained to move only in one dimension. Now celebrated as the "FPU problem," the model for the system consists of a large set of coupled, ordinary differential equations for the time evolution of the particles (see "The Fermi, Pasta, and Ulam Problem: Excerpts from 'Studies of Nonlinear Problems' "). Specifically, one particular version of the FPU problem has 64 particles obeying the equations

$$\ddot{x}_i = (x_{i+1} + x_{i-1} - 2x_i) + \alpha\left((x_{i+1} - x_i)^2 - (x_i - x_{i-1})^2\right) \quad \text{for } i = 1, 2, \ldots, 64, \qquad (6)$$

where α is the measure of the strength of the nonlinear interaction between neighboring particles. Thus there are 64 degrees of freedom and, consequently, a 128-dimensional phase space.

Still more complicated, at least *a priori,* are continuous nonlinear dynamical systems, such as **fluids.** Here one must define dynamical variables—such as the density *p(x, t)*-- at every point in space. Hence the number of degrees of freedom, and accordingly the phase-space dimension, becomes infinite; further, the resulting equations of motion become nonlinear *partial* differential equations. Note that one can view these continuous dynamical systems as the limits of large discrete systems and understand their partial differential equations as the limits of many coupled ordinary differential equations.

We can illustrate this approach using a continuous nonlinear dynamical system that will be important in our later discussion. Hopefully, this example will pique the reader's interest, for it also indicates how elegantly perverse nonlinearity can be. The system is represented by the so-called sine-Gordon equation

$$\frac{\partial^2\theta}{\partial t^2} - \frac{\partial^2\theta}{\partial x^2} + \sin\theta = 0, \qquad (7)$$

where the dependent variable $\theta = \theta(x,t)$ is a measure of the response of the system at position x and time t.

Computationally, one natural way to deal with this system is to introduce a discrete spatial grid with spacing Ax such that the position at the *nth* point in the grid is given by $x_n = n\,\Delta x$ and define $\theta_n(t) \equiv \theta(x_n,t)$ for $n = 1,2,\ldots N$. Using a finite difference approximation for the second derivative,

$$\frac{\partial^2\theta}{\partial x^2} \simeq \frac{1}{(\Delta x)^2}\left(\theta_{n+1}(t) - 2\theta_n(t) + \theta_{n-1}(t)\right) + \mathcal{O}(\Delta x), \tag{8}$$

leads to a set of N coupled ordinary differential equations

$$\frac{d^2\theta_n}{dt^2} = \frac{1}{(\Delta x)^2}\left(\theta_{n+1}(t) - 2\theta_n(t) + \theta_{n-1}(t)\right) + \sin\theta_n(t) \quad n = 1,2,\ldots N. \tag{9}$$

This is a *finite* degree-of-freedom dynamical system, like the FPU problem. In particular, it is just a set of simple plane pendula, coupled together by the discretized spatial derivative. Of course, the continuous sine-Gordon equation is recovered in the limit that $N \to \infty$ (and thus $\Delta x \to 0$). The perverseness of nonlinearity is that whereas the Hamiltonian dynamical system described by a finite number N of coupled ordinary differential equations is *not* completely integrable, the infinite-dimensional Hamiltonian system described by the continuum sine-Gordon equation *is!* Further, as we shall later demonstrate, the latter system possesses localized "lump" solutions—the famed *solitons*— that persist for all time.

Hopefully, this digression on dynamical systems has made the subtlety of nonlinear phenomena quite apparent: very simple nonlinear systems—such as the damped, driven pendulum-can exhibit chaos involving extremely complex, apparently random motions, while very complicated systems—such as the one described by the sine-Gordon equation-can exhibit remarkable manifestations of order. The challenge to researchers in this field is to determine which to expect and when.

Paradigms of Nonlinearity. Before examining in some detail how this challenge is being confronted, we need to respond to some obvious but important questions. First, why study nonlinear *science,* rather than nonlinear *chemistry,* or nonlinear *physics,* or nonlinear *biology?* Nonlinear science sounds impossibly broad, too interdisciplinary, or "the study of every thing." However, the absence of a systematic mathematical framework and the complexity of natural nonlinear phenomena suggest that nonlinear behavior is best comprehended by classifying its various manifestations in many different systems and by identifying and studying their common features. Indeed, both the interest and the power of nonlinear science arise precisely because common concepts are being discovered about systems in very different areas of mathematics and natural sciences. These common concepts, or *paradigms, give* insight into nonlinear problems in a large number of disciplines at once. By understanding these paradigms, one can hope to understand the essence of nonlinearity as well as its consequences in many fields.

Second, since it has long been known that most systems are inherently nonlinear, why has there been a sudden blossoming of interest in this field in the past twenty years or so? Why weren't many of these fundamental problems solved a century ago? On reflection, one can identify three recent developments whose synergistic blending has made possible revolutionary progress.

The first, and perhaps most crucial, development has been that of high-speed electronic computers, which permit quantitative numerical simulations of nonlinear systems. Indeed, the term *experimental mathematics* has been coined to describe computer-based investigations into problems inaccessible to analytic methods. Rather than simply confirming quantitatively results already anticipated by qualitative analysis,

experimental mathematics uses the computer to generate qualitative insight where none has existed before. As the visionary of this development, John von Neumann, wrote (in a 1946 article called "On the principles of large scale computing machines"):

> "Our present analytic methods seem unsuitable for the solution of the important problems rising in connection with nonlinear partial differential equations and, in fact, with virtually all types of problems in pure mathematics. . . . really efficient high-speed computing devices may, in the field of nonlinear partial differential equations as well as in many other fields which are now difficult or entirely denied of access, provide us with those heuristic hints which are needed in all parts of mathematics for genuine progress."

Stan Ulam, together with many of his Los Alamos colleagues, was one of the very first to make this vision a reality. Among Stan's pioneering experimental mathematical investigations was the seminal study of the FPU problem mentioned above. Another example was his early numerical work on nonlinear mappings, carried out in collaboration with Paul Stein (see "Iteration of Maps, Strange Attractors, and Number Theory-An Ulamian Potpourri"). Both of these studies will figure in our later discussion.

The second crucial development has been the experimental observation of "universal" nonlinear characteristics in natural systems that range from chicken hearts and chemical reactors to fluids and plasmas. In the past decade these experiments have reached previously inaccessible levels of precision, so that one can measure *quantitative* similarities in, for example, the route to chaotic behavior among an enormous variety of nonlinear systems.

The third and final development has been in the area of novel analytical mathematical methods. For instance, the invention of the *inverse spectral transform* has led to a systematic method for the explicit solution of a large number of nonlinear partial differential equations. Similarly, new methods based on the theory of Hamiltonian systems allow the analysis of nonlinear stability of a wide range of physically relevant mathematical models.

As we shall shortly see, the methodology based on these three developments has been remarkably successful in solving many nonlinear problems long considered intractable. Moreover, the common characteristics of nonlinear phenomena in very distinct fields has allowed progress in one discipline to transfer rapidly to others and confirms the inherently interdisciplinary nature of nonlinear science. Despite this progress, however, we do not have an entirely systematic approach to nonlinear problems. For the *general* nonlinear equation there is simply no analog of a Fourier transform. We do, however, have an increasing number of well-defined *paradigms* that both reflect typical qualitative features and permit quantitative analysis of a wide range of nonlinear systems. In the ensuing three sections I will focus on three such paradigms: *coherent structures and solitons, deterministic chaos and fractals,* and *complex configurations and patterns.* Of these the first two are well developed and amply exemplified, whereas the third is still emerging. Appropriately, these paradigms reflect different aspects of nonlinearity: coherent structures reveal a surprising orderliness, deterministic chaos illustrates an exquisite disorder, and complex configurations represent the titanic struggle between opposing aspects of order and chaos.

If we were to follow the biblical sequence we would start with chaos, but because it is frankly a rather counterintuitive concept, we shall start with solitons or, more generally and accurately, coherent structures.

Coherent Structures and Solitons

From the Red Spot of Jupiter through clumps of electromagnetic radiation in turbulent plasmas to microscopic charge-density waves on the atomic scale, spatially localized, long-lived, wave-like excitations abound in nonlinear systems. These nonlinear

waves and structures reflect a surprising orderliness in the midst of complex behavior. Their ubiquitous role in both natural nonlinear phenomena and the corresponding mathematical models has caused coherent structures and solitons to emerge as one of the central paradigms of nonlinear science. Coherent structures typically represent the natural "modes" for understanding the time-evolution of the nonlinear system and often dominate the long-time behavior of the motion.

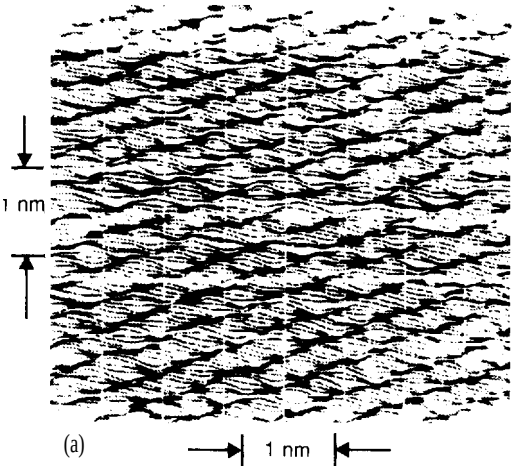
To illustrate this, let me begin with one of the most familiar (and beautiful!) examples in nature, namely, the giant Red Spot (Fig. 3a). This feature, first observed

COHERENT STRUCTURES IN NATURE

Fig. 3. (a) A closeup of the Red Spot of Jupiter, taken from the Voyager spacecraft. False color is used to enhance features of the image. In addition to the celebrated Red Spot, there are many other "coherent structures" on smaller scales on Jupiter. (Photo courtesy of NASA). (b) Nonlinear surface waves in the Andaman Sea off the coast of Thailand as photographed from an Apollo-Soyuz spacecraft. (Photo courtesy of NASA.)

from earth in the late seventeenth century, has remained remarkably stable in the turbulent cauldron of Jupiter's atmosphere. It represents a coherent structure on a scale of about 4 x 108 meters, or roughly the distance from the earth to the moon.

To give an example at the terrestrial level, certain classes of nonlinear ocean waves form coherent structures that propagate essentially unchanged for thousands of miles. Figure 3b is a photograph taken from an Apollo-Soyuz spacecraft of a region of open ocean in the Andaman Sea near northern Sumatra. One sees clearly a packet of five nearly straight surface waves; each is approximately 150 kilometers wide, so the scale of this phenomenon is roughly 10^5 meters. Individual waves within the



COHERENT STRUCTURE IN CRYSTALS

Fig. 4. (a) An image, made by using tunneling-electron microscopy, of a cleaved surface of tantalum diselenide that shows the expected graininess around atomic sites in the crystal lattice. (b) A similar image of tantalum disulfide, showing coherent structures called charge-density waves that are *not* simply a reflection of the crystal lattice but arise from nonlinear interaction effects. (Photos courtesy of C. G. Slough, W. W. McNairy, R. V. Coleman, B. Drake, and P. K. Hansma, University of Virginia.)

packet are separated from each other by about 10 kilometers. The waves, which are generated by tidal forces, move in the direction perpendicular to their crests at a speed of about 2 meters per second. Although the surface deflection of these waves is small—about 1.8 meters—they can here be seen from orbit because the sun is directly behind the spacecraft, causing the specular reflection to be very sensitive to variations of the surface. These visible surface waves are actually a manifestation of much larger amplitude—perhaps ten times larger—internal waves. The internal waves exist because thermal or salinity gradients lead to a stratification of the subsurface into layers. *A prim-i* such large internal waves could pose a threat to submarines and to off-shore structures. Indeed, the research on these waves was initiated by Exxon Corporation to assess the actual risks to the oil rigs they planned to construct in the area. Fortunately, in this context the phenomenon turned out to be more beautiful than threatening.

Our final physical illustration is drawn from solid-state physics, where the phenomenon of charge-density waves exemplifies coherent structures on the atomic scale. If one studies a crystal of tantalum diselenide using an imaging process called tunneling-electron microscopy (Fig. 4a), one finds an image that is slightly denser around the atomic sites but otherwise is uniform. Given that the experimental technique focuses on specfic electronic levels, this graininess is precisely what one would expect at the atomic level; there are no nonlinear coherent structures, no charge-density waves. In contrast, tantalum disulfide, which has nearly identical lattice parameters, exhibits much larger structures in the corresponding image (Fig. 4b); in fact, the image shows a hexagonal array of coherent structures. These charge-density waves are separated by about 3.5 normal lattice spacings, so their occurence is not simply a reflection of the natural atomic graininess. Rather, these coherent structures arise because of a nonlinear coupling between the electrons and the atomic nuclei in the lattice. Notice that now the scale is 10^{-9} meter.

Solitons. We have thus identified nonlinear coherent structures in nature on scales ranging from 10^{8} meters to 10^{-9} meter—seventeen orders of magnitude! Clearly this paradigm is an essential part of nonlinear science. It is therefore very gratifying that during the past twenty years we have seen a veritable revolution in the understanding of coherent structures. The crucial event that brought on this revolution was the discovery, by Norman Zabusky and Martin Kruskal in 1965, of the remarkable soliton. In a sense, solitons represent the purest form of the coherent-structure paradigm and thus are a natural place to begin our detailed analysis. Further, the history of this discovery shows the intricate interweaving of the various threads of Stan Ulam's legacy to nonlinear science.

To define a soliton precisely, we consider the motion of a wave described by an equation that, in general, will be nonlinear. A *traveling wave* solution to such an equation is one that depends on the space x and time t variables only through the combination $\xi = x - vt$, where v is the constant velocity of the wave. The traveling wave moves through space without changing its shape and in particular without spreading out or dispersing. If the traveling wave is a localized single pulse, it is called a solitary wave. A soliton is a solitary wave with the crucial additional property that it preserves its form exactly when it interacts with other solitary waves.

The study that led Kruskal and Zabusky to the soliton had its origin in the famous FPU problem, indeed in precisely the form shown in Eq. 6. Experimental mathematical studies of those equations showed, instead of the equipartition of energy expected on general grounds from statistical mechanics, a puzzling series of recurrences of the initial state (see "The Ergodic Hypothesis: A Complicated Problem of Mathematics and Physics"). Through a series of asymptotic approximations, Kruskal and Zabusky related the recurrence question for the system of oscillators in the FPU problem to the nonlinear partial differential equation

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \qquad (10)$$

Equation 10, called the Korteweg-deVries or KdV equation, had first been derived in 1895 as an approximate description of water waves moving in a shallow, narrow channel. Indeed, the surface waves in the Andaman Sea, which move essentially in one direction and therefore can be modeled by an equation having only one spatial variable, are described quite accurately by Eq. 10. That this same equation should also appear as a limiting case in the study of a discrete lattice of nonlinear oscillators is an illustration of the generic nature of nonlinear phenomena.

To look analytically for a coherent structure in Eq. 10, one seeks a localized solution $u_s(\xi)$ that depends only on $\xi = x - vt$, thereby reducing the partial differential equation to an ordinary differential equation in ξ. The result can be integrated explicitly and, for solutions that vanish at infinity, yields

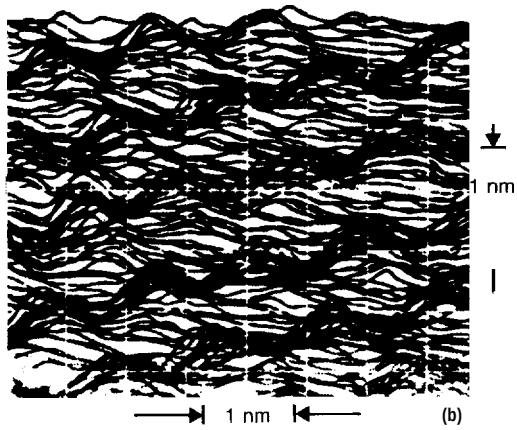
$$u_s(x,t) = 3v \operatorname{sech}^2 \frac{\sqrt{v}}{2}(x - vt) \tag{11}$$

This solution describes a solitary wave moving with constant velocity v. Moreover, the amplitude of the wave is proportional to v, and its width is inversely proportional to $\sqrt{v}$. The faster the wave goes, the narrower it gets. This relation between the shape and velocity of the wave reflects the nonlinearity of the KdV equation.

Intuitively, we can understand the existence of this solitary wave as a result of a delicate balance in the KdV equation between the linear dispersive term $\frac{\partial^3 u}{\partial x^3}$, which tends to cause an initially localized pulse to spread out and change shape as it moves, and the nonlinear *convective derivative* term $u\frac{\partial u}{\partial x}$, which tends to increase the pulse where it is already large and hence to bunch up the disturbance. (For a more precise technical analysis of these competing effects in another important nonlinear equation, see "Solitons in the Sine-Gordon Equation.")

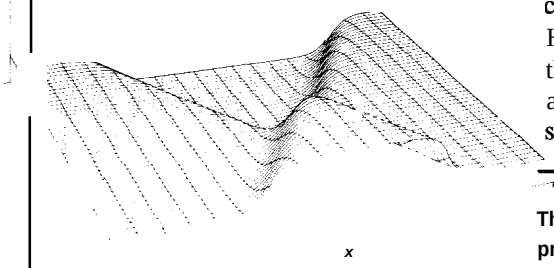
Although the solution represented by Eq. 11 is, by inspection, a coherent structure, is it a soliton? In other words, does it preserve its form when it collides with another solitary wave? Since the analytic methods of the 1960s could not answer this question, Zabusky and Kruskal followed another of Ulam's leads and adopted an experimental mathematics approach by performing computer simulations of the collision of two solitary waves with different velocities. Their expectation was that the nonlinear nature of the interaction would break up the waves, causing them to change their properties dramatically and perhaps to disappear entirely. When the computer gave the startling result that the coherent structures emerged from the interaction unaffected in shape, amplitude, and velocity, Zabusky and Kruskal coined the term "soliton," a name reflecting the particle-like attributes of this nonlinear wave and patterned after the names physicists traditionally give to atomic and subatomic particles.

In the years since 1965 research has revealed the existence of solitons in a host of other nonlinear equations, primarily but not exclusively in one spatial dimension. Significantly, the insights gained from the early experimental mathematical studies have had profound impact on many areas of more conventional mathematics, including infinite-dimensional analysis, algebraic geometry, partial differential equations, and dynamical systems theory. To be more specific, the results of Kruskal and Zabusky led directly to the invention of a novel analytic method, now known as the "inverse spectral transform," that permits the explicit and systematic solution of soliton-bearing equations by a series of effectively linear operations. Further, viewed as nonlinear dynamical systems, the soliton equations have been shown to correspond to infinite-degree-of-freedom Hamiltonian systems that possess an infinite number of independent conservation laws and are thus completely integrable. Indeed, the invariance of solitons under interactions can be understood as a consequence of these conservation laws.



Applied Solitons. From all perspectives nonlinear partial differential equations containing solitons are quite special. Nonetheless, as our examples suggest, there is a surprising mathematical diversity to these equations. This diversity is reflected in the

SOLITONS in the SINE-GORDON Equation



INTERACTION BETWEEN TWO SOLITONS

The robustness of solltons is indicated in this projected space-time plot of the Interaction of a kink and an antikink in the sine-Gordon equation (Eq. 5). As time develops (toward the reader), the two steplike solitary waves approach each other, interact nonlinearly, and then emerge unchanged in shape, amplitude, and velocity. The collision process is described analytically by Eq. 9. (The figure was made at the Los Alamos National Laboratory by Michel Peyrard, University of Bourgogne, France.)

To understand quantitatively how solitons can result from a delicate balance of dispersion and nonlinearity, let us begin with the *linear, dispersionless,* bidirectional wave equation

$$\frac{\partial^2 u}{\partial t^2} - c_0^2 \frac{\partial^2 u}{\partial x^2} = 0. \qquad (1)$$

By direct substitution into Eq. 1, it is easy to check that, with $\eta \equiv x - c_0 t$ and $\xi \equiv x + c_0 t$,

$$u(x,t) = f(\eta) + g(\xi)$$

is a solution for *any* functions f and g. Thus if we take, for example, $f(\eta) = e^{-\eta^2}$ and $g(\xi) = e^{-\xi^2}$, we will have two "solitary waves," one moving to the left and one to the right. At $t \to -\infty$, the lumps are isolated, at $t = O$ they collide, and at $t \to +\infty$ they re-emerge. Hence, by our definition these solutions to the linear, dispersionless wave equation are solitons, although trivial examples of such.

Now consider an equation, still linear, of the form

$$\frac{\partial^2 u}{\partial t^2} - c_0^2 \frac{\partial^2 u}{\partial x^2} + m^2 u = 0. \qquad (2)$$

Such equations arise naturally in descriptions of optically-active phonons in solid state physics and in relativistic field theories. An elementary (plane wave) solution of this equation has the form

$$u(x,t) = Ae^{i(\omega t + kx)}, \qquad (3)$$

where A is a constant, w is the frequency, and k is the wave number. Substituting into Eq. 2 shows that this plane wave can be a solution of Eq. 2 *only if*

$$-\omega^2 + c_0^2 k^2 + m^2 = 0, \qquad (4a)$$

so that

$$\omega(k) = \pm\sqrt{m^2 + c_0^2 k^2}. \qquad (4b)$$

This relation between w and k is known technically as a *dispersion relation* and contains essential information about how individual plane waves with different k's (and hence different w's) propagate. In particular, the group velocity,

$$v_g(k_0) \equiv \frac{\partial \omega}{\partial k}\Big|_{k_0},$$

measures how fast a small group of waves with values of k peaked around a particular value k_0 moves. Notice that for the dispersion relation Eq. 4b,

$$v_g(k_0) = \pm \frac{c_0^2 k_0}{\sqrt{c_0^2 k_0^2 + m^2}},$$

corresponding variety of real-world applications to problems in the natural sciences and engineering. In fiber optics, conducting polymers and other quasi-one-dimensional systems, Josephson transmission lines, and plasma cavitons—as well as the surface waves in the Andaman Sea!—the prevailing mathematical models are slight modifications of soliton equations. There now exist several numerical and analytic perturbation techniques for studying these "nearly" soliton equations, and one can use these to describe quite accurately the behavior of real physical systems.

One specific, decidedly practical illustration of the application of solitons concerns effective long-distance communication by means of optical fibers. Low-intensity light pulses in optical fibers propagate linearly but dispersively (as described in "Solitons in the Sine-Gordon Equation"). This dispersion tends to degrade the signal, and, as a consequence, expensive "repeaters" must be added to the fiber at regular intervals to reconstruct the pulse.

However, if the intensity of the light transmitted through the fiber is substantially increased, the propagation becomes nonlinear and solitary wave pulses are formed. In fact, these solitary waves are very well described by the solitons of the "nonlinear Schrodinger equation," another of the celebrated completely integrable nonlinear partial differential equations. In terms of the (complex) electric field amplitude *E(x, t),* this equation can be written

so that (groups of) waves with different values of k_0 will have different group velocities. Now consider a general solution to Eq. 2, which, by the principle of superposition, can be formed by adding together many plane waves (each with a different constant). Since the elementary components with different wave numbers will propagate at different group velocities, the general solution will change its form, or disperse, as it moves. Hence, the general solution to Eq. 2 cannot be a soliton.

Next consider adding a nonlinear term to Eq. 2. With considerable malice aforethought, we change notation for the dependent variable and choose the nonlinearity so that the full equation becomes

$$\frac{\partial^2\theta}{\partial t^2} - c_0^2\frac{\partial^2\theta}{\partial x^2} + m^2\sin\theta = 0, \quad (5)$$

the "sine-Gordon" equation. We can compare Eq. 5 to our previous Eq. 2 by noting that in the limit of small θ, Eq. 5 reduces to

$$\frac{\partial^2\theta}{\partial t^2} - c_0^2\frac{\partial^2\theta}{\partial x^2} + m^2\theta - \frac{1}{6}\theta^3 + \cdots = 0, \quad (6)$$

where the remaining terms are $\mathcal{O}(\theta^5)$ and higher.

Based on remarks made in the introductory section of the main text, we see that Eq. 5 looks like a bunch of simple, plane pendulums coupled together by the spatial derivative term $\partial^2\theta/\partial x^2$. In fact, the sine-Gordon equation has many physical applications, including descriptions of chain-like magnetic compounds and transmission lines made out of arrays of Josephson junctions of superconductors. Also, the equation is one of the celebrated completely integrable, infinite-degree-of-freedom Hamiltonian systems, and the initial-value problem for the equation can be solved exactly by the analytic technique of the "inverse spectral transform." Since the details of this method are well beyond the scope of a general overview, we shall only quote the solutions relevant to our discussion. First, just as for the KdV equation (Eq. 10 in the main text), one can find directly a single solitary-wave solution:

$$\theta_k(x,t) = 4\tan^{-1}e^{\gamma(\zeta - v\tau)}, \quad (7)$$

with $\gamma = 1/\sqrt{1-v^2}$, $\zeta = mx/c_0$, and $\tau = mt$.

Since this solution approaches O as $x \to -\infty$ and 2π as $x \to +\infty$, it describes a glitch in the field θ localized around $\zeta = v\tau$, that is, around $x = c_0 vt$. As a consequence, it is known as a "kink." Importantly, it does represent a physically truly localized excitation, because all the energy and momentum associated with this wave are exponentially centered around the kink's location. Similarly, the so-called anti-kink solution

$$\theta_{\bar{k}}(x,t) = 4\tan^{-1}e^{-\gamma(\zeta - v\tau)}, \quad (8)$$

interpolates between 2π as $x \to -\infty$ and 0 as $x \to +\infty$.

Are the kinks and anti-kinks solitons? Here we can avail ourselves of the miracle of integrability and simply write down an analytic solution that describes the scattering of a kink and an antikink. The result is

$$\theta_{k\bar{k}}(x,t) = 4\tan^{-1}\left(\frac{\sinh\frac{v\tau}{\sqrt{1-v^2}}}{v\cosh\frac{\zeta}{\sqrt{1-v^2}}}\right). \quad (9)$$

The dedicated reader can verify that as $t \to -\infty$, $\theta_{k\bar{k}}$ looks like a widely separated kink and anti-kink approaching each other at velocity v. For t near O they interact nonlinearly, but as $t \to +\infty$, the kink and anti-kink emerge with their forms intact. Readers with less dedication can simply refer to the figure, in which the entire collision process is presented in a space-time plot. Note that since the equation is invariant under $\theta \to \theta + 2n\pi$, a kink that interpolates between 2π and 4π is physically equivalent to one that interpolates between 0 and 2π.

In the interest of historical accuracy, we should add one final point. The analytic solution, Eq. 9, showing that the kink and anti-kink are in fact solitons, was actually known, albeit not widely, *before* the discovery of the KdV soliton. It had remained an isolated and arcane curiosity, independently rediscovered several times but without widespread impact. That such solutions could be constructed analytically in a wide range of theories was not appreciated. It took the experimental mathematics of Zabusky and Kruskal to lead to the soliton revolution. ■

$$i\frac{\partial E}{\partial t} + \frac{\partial^2 E}{\partial x^2} + |E|^2 E = 0. \quad (12)$$

The soliton corresponding to the nonlinear pulse moving with velocity v through the optical fiber has the form

$$E(x,t) = \left(2\omega + \frac{v^2}{2}\right)^{\frac{1}{2}} e^{i\omega t + \frac{vx}{2}}\,\mathrm{sech}\left(\left(\omega + \frac{v^2}{4}\right)^{\frac{1}{2}}(x - vt)\right). \quad (13)$$

In the idealized limit of no dissipative energy loss, these solitons propagate without degradation of shape; they are indeed the natural stable, localized modes for propagation in the fiber. An intrinsically nonlinear characteristic of this soliton, shown explicitly in Eq. 13, is the relation between its amplitude (hence its energy) and its width. In real fibers, where dissipative mechanisms cause solitons to lose energy, the individual soliton pulses therefore broaden (but do not disperse). Thus, to maintain the separation between solitons necessary for the integrity of the signal, one must add optical amplifiers. based on stimulated Raman amplification, to compensate for the loss.

Theoretical numerical studies suggest that the amplification can be done very effectively. An all-optical system with amplifier spacings of 30 to 50 kilometers and

Raman pump power levels less than 100 milliwatts can use solitons of 20 picosecond duration to send information at a bit rate of over 10 gigahertz. This is more than an order of magnitude greater than the rate anticipated for conventional (linear) systems. Although laboratory experiments have confirmed some of these results, full engineering studies have yet to be carried out. In addition, a critical and still unresolved issue is the relative cost of the repeaters supporting the linear system versus that of the amplifiers in the soliton-based approach. Nonetheless, the prospects for using optical solitons in long-distance communication are exciting and real.

Coherent Structures. Thus far our discussion of the coherent-structure paradigm has focused almost exclusively on solitons. Although this emphasis correctly indicates both the tremendous interest and the substantial progress to which this aspect of the paradigm has led, it obscures the much broader role that *nonsoliton* coherent structures play in nonlinear phenomena. Vortices in fluids, chemical-reaction waves and nonlinear diffusion fronts, shock waves, dislocations in metals, and bubbles and droplets can all usefully be viewed as instances of coherent structures. As in the. case of the solitons, the existence of these structures results from a delicate balance of nonlinear and dispersive forces.

In contrast to solitons, however, these more general coherent structures typically interact strongly and do not necessarily maintain their form or even their separate identities for all times. Fluid vortices may merge to form a single coherent structure equivalent to a single larger vortex. Interactions among shock waves lead to diffraction patterns of incident, reflected, and transmitted shocks. Droplets and bubbles can interact through merging or splitting. Despite these nontrivial interactions, the coherent structures can be the nonlinear *modes* in which the dynamics is naturally described, and they may dominate the long-time behavior of the system. To exemplify more concretely the essential role of these general coherent structures in nonlinear systems, let me focus on two broad classes of such structures: vortices and fronts.

The importance of vortices in complicated fluid flows and turbulence has been appreciated since ancient times. The giant Red Spot (Fig. 3a) is a well-known example of a fluid vortex, as are tornados in the earth's atmosphere, large ocean circulation patterns called "modons" in the Gulf Stream current, and "rotons" in liquid helium. In terms of practical applications, the vortex pattern formed by a moving airfoil is immensely important. Not only does this pattern of vortices affect the fuel efficiency and performance of the aircraft, but it also governs the allowed spacing between planes at takeoff and landing. More generally, vortices are the coherent structures that make up the turbulent boundary layer on the surfaces of wings or other objects moving through fluids. Further, methods based on idealized point vortices provide an important approach to the numerical simulation of certain fluid flows.

The existence of fronts as coherent structures provides yet another illustration of the essential role of nonlinearity in the physical world. Linear diffusion equations cannot support wave-like solutions. In the presence of nonlinearity, however, diffusion equations can have traveling wave solutions, with the propagating wave front representing a transition from one state of the system to another. Thus, for example, chemical reaction-diffusion systems can have traveling wave fronts separating reacted and unreacted species. Often, as in flame fronts or in internal combustion engines, these traveling chemical waves are coupled with fluid modes as well. Concentration fronts arise in the leaching of minerals from ore beds. Moving fronts between infected and non-infected individuals can be identified in the epidemiology of diseases such as rabies. In advanced oil recovery processes, (unstable) fronts between the injected water and the oil trapped in the reservoir control the effectiveness of the recovery process.

Given their ubiquity and obvious importance in nonlinear phenomena, it is gratifying that recent years have witnessed remarkable progress in understanding and modeling these general coherent structures. Significantly, this progress has been achieved by pre-

cisely the synergy among computation, theory, and experiment that we have argued characterizes nonlinear science. Further, as a consequence of this progress, coherent structures and solitons have emerged as an essential paradigm of nonlinear science, providing a unifying concept and an associated methodology at the theoretical, computational, and experimental levels. The importance of this paradigm for technological applications, as well as its inherent interest for fundamental science, will guarantee its central role in all future research in this subject.

Deterministic Chaos and Fractals

Deterministic chaos is the term applied to the aperiodic, irregular, unpredictable, random behavior that in the past two decades has been observed in an incredible variety of nonlinear systems, both mathematical and natural. Although the processes are strictly deterministic and all forces are known, the long-time behavior defies prediction and is as random as a coin toss,

That a system governed by deterministic laws can exhibit effectively random behavior runs directly counter to our normal intuition. Perhaps it is because intuition is inherently "linear;" indeed, deterministic chaos *cannot* occur in linear systems. More likely, it is because of our deeply ingrained view of a clockwork universe, a view that in the West was forcefully stated by the great French mathematician and natural philosopher Laplace. If one could know the positions and velocities of all the particles in the universe and the nature of all the forces among them, then one could chart the course of the universe for all time. In short, from exact knowledge of the initial state (and the forces) comes an exact knowledge of the final state. In Newtonian mechanics this belief is true, and to avoid any possible confusion, I stress that we are considering only dynamical systems obeying classical, Newtonian mechanics. Subsequent remarks have nothing to do with "uncertainties" caused by quantum mechanics.

However, in the real world exact knowledge of the initial state is not achievable. No matter how accurately the velocity of a particular particle is measured, one can demand that it be measured more accurately. Although we may, in general, recognize our inability to have such exact knowledge, we typically assume that if the initial conditions of two separate experiments are *almost* the same, the final conditions will be *almost* the same. For most smoothly behaved, "normal" systems this assumption is correct. But for certain nonlinear systems it is false, and deterministic chaos is the result.

At the turn of this century, Henri Poincare, another great French mathematician and natural philosopher, understood this possibility very precisely and wrote (as translated in *Science and Method):*

> "A very small cause which escapes our notice determines a considerable effect that we cannot fail to see, and then we say that that effect is due to chance. If we knew exactly the laws of nature and the situation of the universe at the initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still only know the initial situation *approximately.* If that enabled us to predict the succeeding situation *with the same approximation,* that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the later. Prediction becomes impossible, and we have the fortuitous phenomenon."

Despite Poincare's remarkable insight, deterministic chaos remained virtually unexplored and unknown until the early 1960s. As the ensuing discussion will reveal, the reason for this long hiatus is that chaos defies direct analytic treatment. The seeds

planted by Poincare could only germinate when the advances in interactive computation made experimental mathematics a reality.

The Logistic Map. One remarkable instance of a successful experimental mathematical study occurred in a nonlinear equation simple enough to explain to an elementary school child or to analyze on a pocket calculator yet subtle enough to capture the essence of a whole class of real world phenomena. It is arguably the simplest model of a system displaying deterministic chaos, and as such has been studied by a host of distinguished researchers, including Ulam, von Neumann, Kac, Metropolis, Stein, May, and Feigenbaum (see "Iteration of Maps, Strange Attractors, and Number theory—An Ulamian Potpourri"). As we shall see, this focus of talent has been fully justified, for the simple model provides remarkable insight into a wealth of nonlinear phenomena. Thus it is a natural place to begin our quantitative study of deterministic chaos.

The model, known as the *logistic map,* is a discrete-time, dissipative, nonlinear dynamical system. The value of a variable x_n at time n is mapped to a new value x_{n+1} at time $n + 1$ according to the nonlinear function

$$x_{n+1} = rx_n(1 - x_n), \tag{14}$$

where the *control parameter* r satisfies $0 < r \leq 4$ and the allowed values—loosely speaking, the phase space—of the x_n are $0 \leq x_n \leq 1$. The map is iterated as many times as desired, and one is particularly interested in the behavior as time—that is, n, the number of iterations—approaches infinity. Specifically, if an initial condition is picked at random in the interval (O, 1) and iterated many times, what is its motion after all transients have died out?

The behavior of this nonlinear map depends critically on the control parameter and exhibits in certain regions sudden and dramatic changes in response to small variations in r. These changes, technically called *bifurcations,* provide a concrete example of our earlier observation that small changes in the parameters of a nonlinear system can produce enormous qualitative differences in the motion.

For $0 < r < 1$, the value of x_n drops to 0 as n approaches infinity no matter what its inital value. In other words, after the transients disappear, all points in the interval (0, 1) are *attracted* to the *fixed point* x^* at $x =$ O. This fixed point is analogous to the fixed point in Fig. 2 at $(\theta = 0, d\theta/dt = 0)$ with the very important distinction that the fixed point in the logistic map is an *attractor:* the dissipative nature of the map causes the "volume" in phase space to collapse to a single point. Such attractors are impossible in Hamiltonian systems, since their motion preserves phase-space volumes (see "Hamiltonian Chaos and Statistical Mechanics"). The mathematical statement of this behavior then is

$$\lim_{n\to\infty} x_n = x^* = 0. \tag{15}$$

We can easily calculate the (linearized) stability of this fixed point by considering how small deviations from it behave under the map. In Eq. 14 we set $x_n = x^* + \epsilon_n$ and $x_{n+1} = x^* + \epsilon_{n+1}$ and consider only terms linear in ϵ_n and ϵ_{n+1}. The resulting equation is

$$\epsilon_{n+1} = r(1 - 2x^*)\epsilon_n + \mathcal{O}(\epsilon_n^2), \tag{16}$$

so that for $x^* = 0$, the ϵ_n's will remain small for all iterations—provided $r < 1$.

This last comment suggests that something interesting happens as r passes 1, and indeed for $1 < r < 3$ we find an attracting fixed point with a value that depends on r. This value is readily calculated, since at a fixed point $x_n = x_{n+1} = x^*$. Substituting this relation into Eq. 14, we find

$$\lim_{n\to\infty} x_n = x^*(r) = 1 - \frac{1}{r}. \tag{17}$$

Hence as the value of r moves from I toward 3, the value of the fixed point x* moves from O toward 2/3. Notice that the linear stability analysis given above shows that this r-dependent fixed point is stable for $1 < r < 3$. Notice also that while $x^* = O$ is *still* a fixed point in this region, the linear stability analysis shows that it is unstable. Hence the point $x = 0$ is now analogous to the unstable fixed point in Fig. 2 at $\theta = \pi$, $d\theta/dt = 0$; the slightest perturbation will drive the solution away from $x = 0$ to the stable fixed point at $x^*(r)$.

A more interesting bifurcation occurs at $r = 3$. Suddenly, instead of approaching a single fixed point, the long-time solution oscillates between two values: thus the model has an *attracting limit cycle* of period 2! This limit cycle is the discrete analogue of the closed periodic oscillations shown in the phase plane of the pendulum (Fig. 2), again, of course, subject to the distinction that the logistic-map limit cycle is an attractor.

Although one can still continue analytically at this stage, it is easier to refer to the results of an experimental mathematical simulation (Fig. 5) that depicts the attracting set in the logistic map as a function of r. Here we see clearly the bifurcation to the period-2 limit cycle at r = 3. But more striking, as r moves toward 3.5 and beyond, period-4 and then period-8 limit cycles occur, followed by a region in which the attracting set becomes incredibly complicated. A careful anlysis of the logistic map shows that the period-8 cycle is followed by cycles with periods 16, 32, 64, and so forth. The process continues through all values 2^n so that the period approaches infinity. Remarkably, all this activity occurs in the *finite* region of r below the value $r_c \sim 3.57$.

Above r_c the attracting set for many (but not all) values of r shows no periodicity whatsoever. In fact, the set consists of a sequence of points x_n that never repeats itself. For these values of r, the simple logistic map exhibits deterministic chaos, and the attracting set—far more complex than the fixed points and limit cycles seen below r_c—is called a *strange attractor*. Beyond the critical value r_c, the logistic map exhibits a *transition to chaos.*

Although this complicated, aperiodic behavior clearly motivates the name "chaos," does it also have the crucial feature of sensitive dependence on initial conditions that we argued was necessary for the long-time behavior to be as random as a coin toss? To study this question, one must observe how two initially nearby points separate as they are iterated by the map. Technically, this can be done by computing the *Lyapunov exponent* λ. A value of λ greater than 0 indicates that the nearby initial points separate exponentially (at a rate determined by λ). If we plot the Lyapunov exponent as a function of the control parameter (Fig. 6), we see that the chaotic regions do have $\lambda > 0$ and, moreover, the periodic windows in Fig. 5 that exist above r_c correspond to regions where $\lambda < 0$. That such a filigree of interwoven regions of periodic and chaotic motion can be produced by a simple quadratically nonlinear map is indeed remarkable.

In view of the complexity of the attracting sets above r_c, it is not at all surprising that this model, like the typical problem in chaotic dynamics, defied direct analytic approaches. There is, however, one elegant analytic result—made all the more relevant here by its having been discovered by Ulam and von Neumann—that further exemplifies the sensitive dependence that characterizes deterministic chaos.

For the particular value $r = 4$, if we let $x_n = \sin^2 \theta_n$, the logistic map can be rewritten

$$\sin^2 \theta_{n+1} = 4 \sin^2 \theta_n \cos^2 \theta_n = (2 \sin \theta_n \cos \theta_n)^2 = (\sin 2\theta_n)^2 . \qquad (18)$$

Hence the map is simply the square of the doubling formula for the sine function, and we see that the solution is $\theta_{n+1} = 2\theta_n$. In terms of the initial value, θ_0, this gives

$$\theta_n = 2^n \theta_0. \qquad (19)$$

This solution makes clear, first, that there is a very sensitive dependence to initial conditions and, second, that there is a very rapid exponential separation from adjacent

initial conditions. For example, by writing θ_n as a binary number with a finite number of digits—as one would in any digital computer—we see that the map amounts to a simple shift operation. When this process is carried out on a real computer, round-off errors replace the right-most bit with garbage after each operation, and each time the map is iterated one bit of information is lost. If the initial condition is known to 48 bits of precision, then after only 48 iterations of the map no information about the initial condition remains. Said another way, despite the completely deterministic nature of the logistic map, the exponential separation of nearby initial conditions means that all long-time information about the motion is encoded in the initial state, whereas none (except for very short times) is encoded in the dynamics.

There is still much more that we can learn from this simple example. One question of obvious interest in nonlinear systems is the mechanism by which such systems move from regular to chaotic motion. In the logistic map, we have seen that this occurs via a *period-doubling cascade* of bifurcations: that is, by a succession of limit cycles with periods increasing as 2^n. In a classic contribution to nonlinear science, Mitchell Feigenbaum analyzed the manner in which this cascade occurred. Among his first results was the observation that the values of the parameter r at which the bifurcations occurred converged geometrically: namely, with δ_n defined by

$$\delta_n \equiv \frac{r_n - r_{n-1}}{r_{n+1} - r_n}, \tag{20}$$

he found

$$\lim_{n\to\infty} \delta_n \equiv \delta = 4.669\ldots . \tag{21}$$

More important, Feigenbaum was able to show that δ did not depend on the details of the logistic map-the function need only have a "generic" maximum, that is, one with a nonvanishing second derivative—and hence δ should be *universal* for all generic maps. Even more, he was able to argue convincingly that whenever a period-doubling cascade of bifurcations is seen in a dissipative dynamical system, the universal number δ, as well as several other universal quantities, should be observed *independent of the system's phase-space dimension.*

This prediction received dramatic confirmation in an experiment carried out by Albert Libchaber and J. Maurer involving convection in liquid helium at low temperatures. Their observation of the period-doubling cascade and the subsequent extraction of δ and other universal parameters provided striking proof of universal behavior in nonlinear systems. More recently, similar confirmation has been found in experiments on nonlinear electrical circuits and semiconductor devices and in numerical simulations of the damped, driven pendulum. Further, it is now known rigorously for dissipative systems that the universal behavior of the period-doubling transition to chaos in the logistic map can occur even when the phase-space dimension becomes infinite.

It is important to emphasize that the period-doubling cascade is by no means the only way in which dissipative nonlinear systems move from regular motion to chaos (see, for example, the discussion of the indented trapezoid map on pp. 103–104). Many other routes—such as *quasiperiodic* and *intermittent—have* been identified and universality theories have been developed for some of them. But the conceptual progenitor of all these developments remains the simple logistic map.

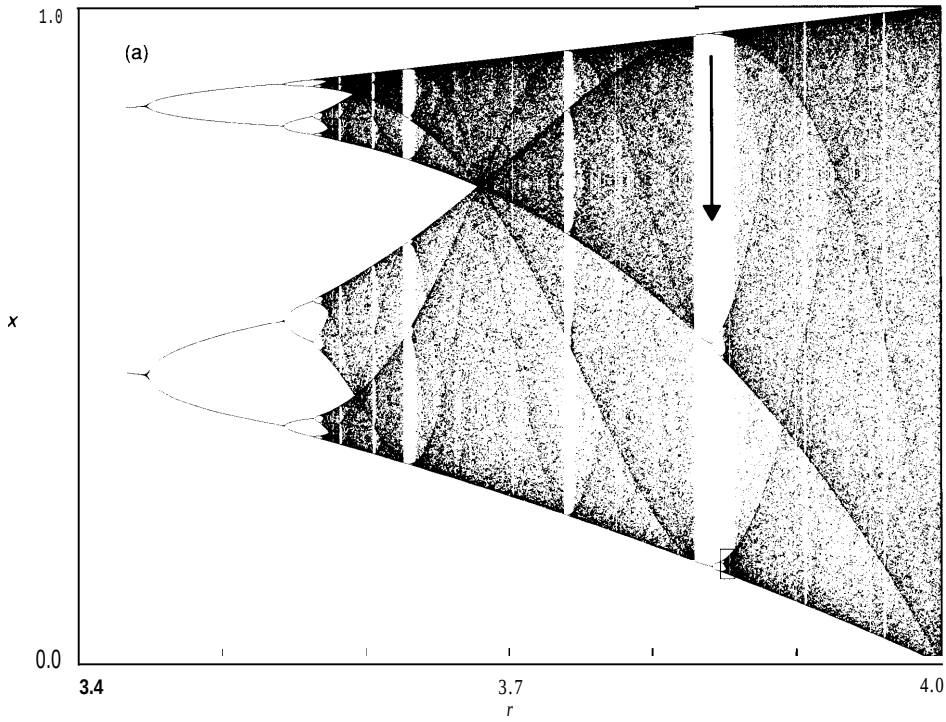
Finally, Fig. 5b illustrates one additional obvious feature of the attracting set of Eq. 14: namely, that it contains nontrivial—and, in fact, self-similar—structure under magnification. Indeed, in the mathematical model this self-similar structure occurs on *all* smaller scales; consequently, Fig. 5b is one example of a class of complex, infinitely ramified geometrical objects called *fractals.* We shall return to this point later.

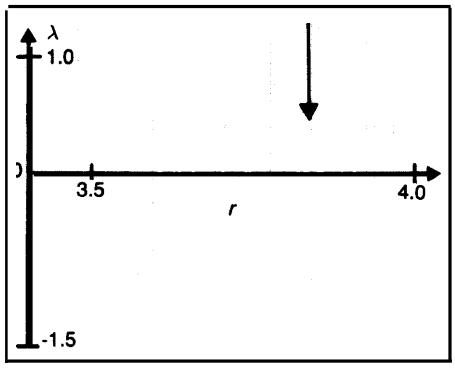
The Damped, Driven Pendulum. Armed with the quantitative insight gained from the logistic map, we can confront deterministic chaos in more conventional dynamical

systems. We start with a very familiar example indeed: namely, the plane pendulum subjected to driving and damping. We can now make precise our earlier assertion that this simple system can behave in a seemingly random, unpredictable, chaotic manner (see "The Simple but Nonlinear Pendulum").

The motion of the damped, driven pendulum is described by Eq. 4 above. Apart from its application to the pendulum, Eq. 4 describes an electronic device called a Josephson tunneling junction in which two superconducting materials are separated by a thin nonconducting oxide layer. Among the present practical applications of such junctions are high-precision magnetometers and voltage standards. The ability of these Josephson junctions to switch rapidly (tens of picosecond) and with very low dissipation (less than microwatt) from one current-carrying state to another may provide microcircuit technologies for, say, supercomputers that are more efficient than those based on conventional semiconductors. Hence the nature of the dynamic response of a Josephson junction to the external driving force—the $\Gamma \cos \Omega t$ term in Eq. 4—is a matter of technological, as well as fundamental, interest.

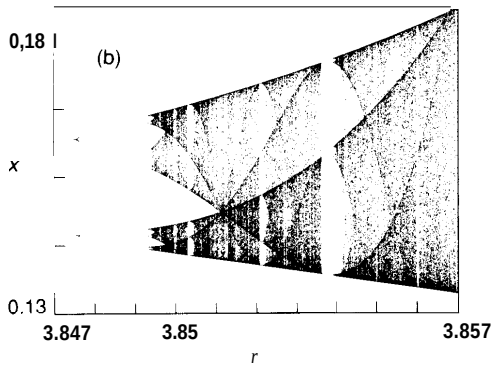
Since analytic techniques are of limited use in the chaotic regime, we demonstrate the existence of chaos in Eq. 4 by relying on graphical results from numerical simulations. Figure 7 illustrates how the phase plane (Fig. 2) of the pendulum is modified when driving and damping forces are included and, in particular, shows how the simple structure involving fixed points and limit cycles is dramatically altered.





THE LYAPUNOV EXPONENT

Fig. 6. A positive value for the Lyapunov exponent ($\lambda > 0$) indicates that nearby initial points separate exponentially, whereas negative values ($\lambda < 0$) indicate periodic or quasiperiodic motion. Here the Lyapunov exponent is plotted as a function of the control parameter *r* for the logistic map (Fig. 5), and it can be seen that the periodic windows of Fig. 5 correspond to regions where $\lambda < 0$. (Figure courtesy of Gottfried Mayer-Kress and Hermann Haken, Universitat Stuttgart, FRG.)



THE LOGISTIC MAP

Fig. 5. (a) The attracting set for the logistic map (Eq. 14 in the main text) generated by plotting 300 values of the iterated function (after the transients have died out) for each of 1150 values of the control parameter *r*. The map has a cycle of period 2 when the control parameter *r is* at 3.4 (left edge). This cycle quickly "bifurcates" to cycles of periods 4, 8, 16, and so forth as *r* increases, generating a period-doubling cascade. Above $r_c \approx 3.57$ the map exhibits deterministic chaos interspersed with gaps where periodic motion has returned. For example, cycles of periods 6, 5, and 3 can be seen in the three larger gaps to the right. (b) A magnified region (shown as a small rectangle in (a)) illustrates the self-similar structure that occurs at smaller scales. (Figure courtesy of Roger Eckhardt, Los Alamos National Laboratory.)

We note first that since there is an external time dependence in Eq. 4, the system really involves *three* first-order differential equations. In a normal dynamical system each degree of freedom results in two first-order equations, so this system is said to correspond to one-and-a-half degrees of freedom. To see this explicitly, we introduce a variable $z = \Omega t$, recall that the angular momentum $p_\theta = ml^2 d\theta/dt$, and rewrite Eq. 4, resulting in

THE DAMPED, DRIVEN PENDULUM: A STRANGE ATTRACTOR

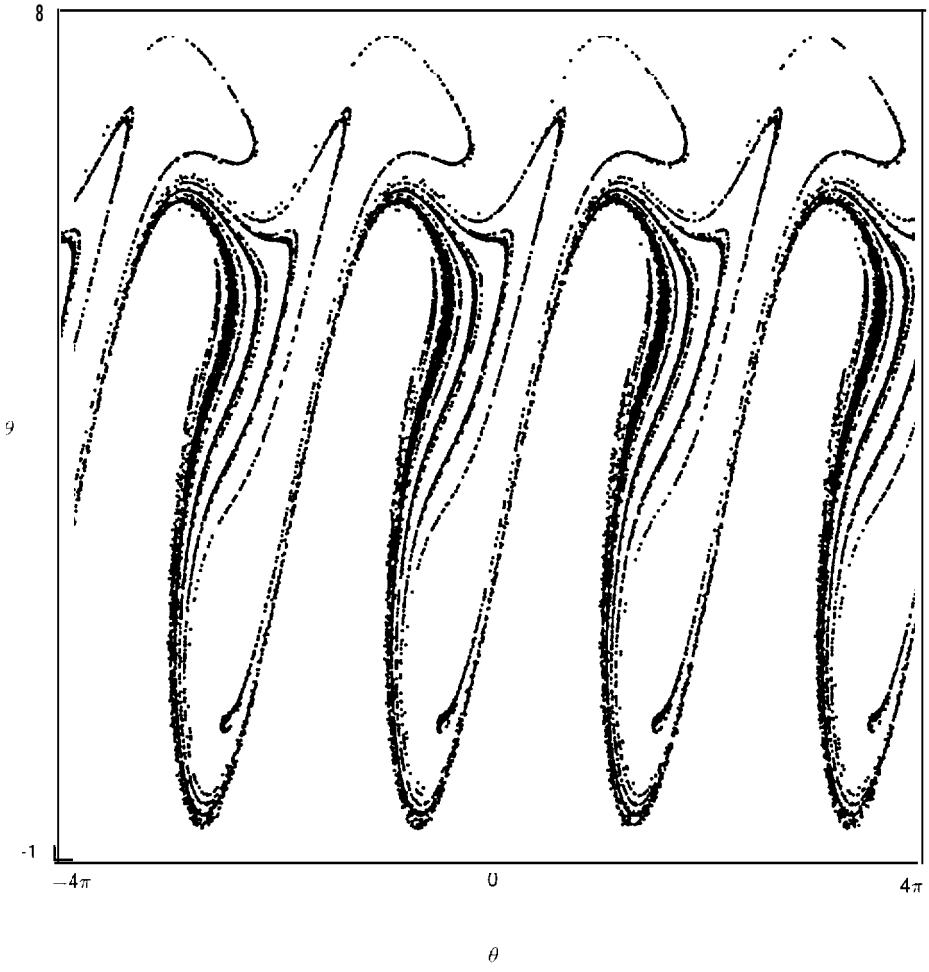


Fig. 7. The motion of a damped, periodically driven pendulum (Eq. 4 in the main text) for certain parameter values is chaotic with the attracting set being a "strange attractor." An impression of such motion can be obtained by plotting the position θ and velocity $\dot{\theta}$ of the pendulum once every cycle of the driving force (as shown here for $\alpha = 0.3$, $\Gamma = 4.5$, and $\Omega = 0.6$ in units with $g/l = 4$). The fact that the image is repeated at higher and lower values of 8 is a result of the pendulum swinging over the top of its pivot point. (Figure courtesy of James Crutchfield, University of California, Berkeley.)

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{1}{ml^2} p_\theta , \\ \frac{dp_\theta}{dt} &= -\alpha p_\theta - mgl \sin\theta + ml^2 \Gamma \cos z , \\ \frac{dz}{dt} &= \Omega , \end{aligned} \tag{22}$$

which shows how the system depends on the three generalized coordinates: θ, p_θ, and z. Note further that the presence of damping implies that the system is no longer Hamiltonian but rather is dissipative and hence can have attractors.

Analysis of the damped, driven pendulum neatly illustrates two separate but related aspects of chaos: first, the existence of a strange attractor, and second, the presence of several different attracting sets and the resulting extreme sensitivity of the asymptotic motion to the precise initial conditions.

Figure 7 shows one of the attractors of Eq. 22 for the parameter values $\alpha = 0.3$, $\Gamma = 4.5$, and $\Omega = 0.6$ (in units with $g/l = 4$). As in the case of the logistic map, only the attracting set is displayed; the transients are not indicated. To obtain Fig. 7, which

is a plot showing only the phase-plane variables θ and $\dot{\theta}$, one takes a "stroboscopic snapshot" of the motion once during every cycle of the driving force. The complicated attracting set shown in the figure is in fact a strange attractor and describes a never-repeating, nonperiodic motion in which the pendulum oscillates and flips over its pivot point (hence the repeated images at 2π-multiples of the angle) in an irregular, chaotic manner. The sensitive dependence on initial conditions implies that nearby points on the attractor will separate exponentially in time, following totally different paths asymptotically. Enlargements of small regions of Fig. 7 show a continuation of the intricate structure on all scales; like the attracting set of the logistic map, this strange attractor is a fractal.

To visualize the motion on this attractor, it may be helpful to recall the behavior of an amusing magnetic parlor toy that has recently been quite popular. This device, for which the mathematical model is closely related to the damped, driven pendulum equation, spins first one way and then the other. At first it may seem that one can guess its behavior. But just when one expects it to spin three times to the right and then go to the left, it instead goes four, five, or perhaps six times to the right. The sequence of right and left rotations is unpredictable because the system is undergoing the aperiodic motion characteristic of a strange attractor.

Figure 8 illustrates the important point that the strange attractor of Fig. 7 is not the only attractor that exists for Eq. 22. Specifically, for $\alpha = 0.1$, $\Gamma = 7/4$, and $\Omega = 1$ (now in units of $g/l = 1$), the system is attracted to *periodic* limit cycles of clockwise or counterclockwise motion. Figure 8 demonstrates this with another variant of our familiar phase-plane plot in which a color code is used to indicate the long-time behavior of all points in the plane. More precisely, this plot is a map of every initial state $(\theta, \dot{\theta})$ onto a "final state" corresponding to one of the attractors. A blue dot is plotted at a point in the plane if the solution that starts from that point at $t = 0$ is attracted asymptotically to the limit cycle corresponding to clockwise rotation of the pendulum. Similarly, a red dot is plotted for initial points for which the solution asymptotically approaches counterclockwise rotation.

In Fig. 8 we observe large regions in which all the points are colored red and, hence, whose initial conditions lead to counterclockwise rotations. Similarly, there are large blue regions leading to clockwise rotations, In between, however, are regions in which the tiniest change in initial conditions leads to alternations in the limit cycle eventually reached. In fact, if you were to magnify these regions even more, you would see further alternations of blue and red-even at the finest possible level. In other words, in these regions the final state of the pendulum-clockwise or counterclockwise motion—is an incredibly sensitive function of the exact initial point.

There is an important subtlety here that requires comment. For the red and blue regions the asymptotic state of the pendulum does *not* correspond to chaotic motion, and the two attracting sets are not strange attractors but are rather just the clockwise and counterclockwise rotations that exist as allowed motions even for the free pendulum (Fig. 2). The aspect of chaos that is reflected by the interwoven red and blue regions is the exquisite sensitivity of the final state to minute changes in the initial state. Thus, in regions speckled with intermingled red and blue dots, it is simply impossible to predict the final state because of an incomplete knowledge of initial conditions.

In addition to the dominant red and blue points and regions, Fig. 8 shows much smaller regions colored greenish white and black. These regions correspond to still other attracting limit sets, the greenish-white regions indicating oscillatory limit cycles (no rotation) and the black regions indicating points that eventually go to a strange attractor.

From the example of Fig. 8 we learn the important lesson that a nonlinear dissipative system may contain many different attractors, each with its own *basin of attraction,* or range of initial conditions asymptotically attracted to it. A subtle further consequence of deterministic chaos is that the boundaries between these basins can

themselves be extraordinarily complex and, in fact, fractal. A fractal basin boundary means that qualitatively different long-time behaviors can result from nearly identical initial configurations.

The Lorenz Attractor. In both cases of the logistic map and the damped, driven pendulum, we have indicated that strange attractors are intimately connected with the presence of dissipative deterministic chaos. These exotic attracting sets reflect motions

THE DAMPED, DRIVEN PENDULUM: PERIODIC LIMIT-CYCLE Attractors

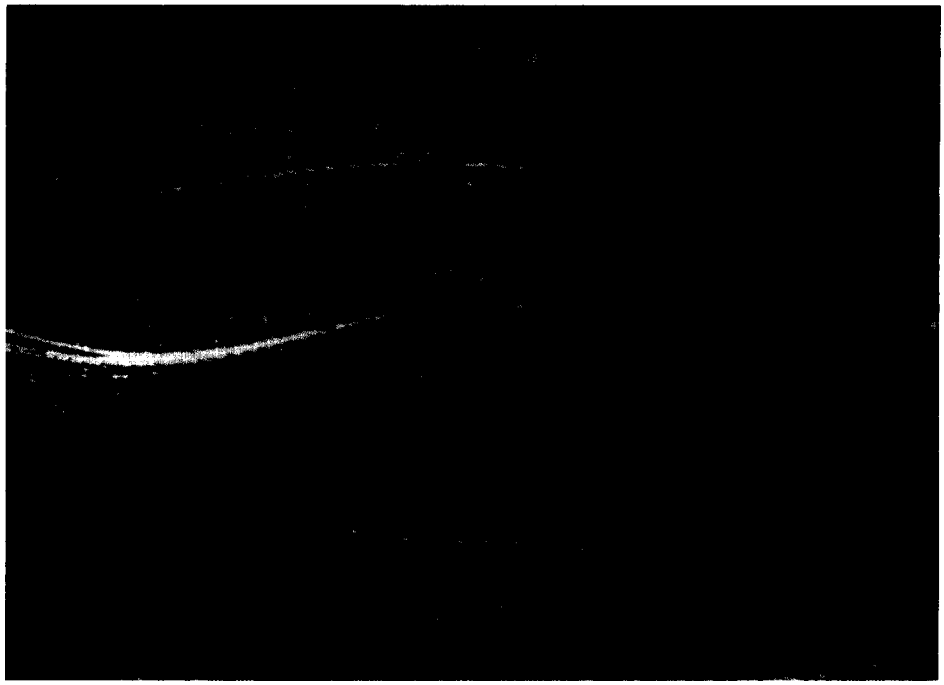


Fig. 8. In this variation of the phase plot for the damped, driven pendulum, a blue dot is plotted at a point in the plane if the solution that starts from that point at $t = 0$ is attracted to clockwise rotation, whereas a red dot represents an attraction to counterclockwise rotation, and a greenish-white dot represents an attraction to any oscillatory limit cycle without rotation. Only a portion of the phase plane is shown. The conditions used to show these limit-cycle attractors are $\alpha = 0.1$, $\Gamma = 7/4$, and $\Omega = 1$ (in units of $g/l = 1$). Despite the nonchaotic motion of the limit cycles, sensitive dependence on initial conditions is still quite evident from the presence of extensive regions of intermingled red and blue. Further, the black regions indicate initial conditions for which the limiting orbit is a strange attractor. (Figure courtesy of Celso Grebogi, Edward Ott, James Yorke, and Frank Varosi, University of Maryland,)

of the system that, even though they may occur in a bounded region of phase space, are *not* periodic (thus never repeating), and motions originating from nearby initial points on the attractor separate exponentially in time. Further, viewed as geometric objects these attractors have an infinitely foliated form and exhibit intricate structure on all scales.

To develop a clearer understanding of these admittedly bizarre objects and the dynamical motions they depict, we turn to another simple nonlinear dynamical model. Known as the *Lorenz equations,* this model was developed in the early 1960s by Edward Lorenz. a meteorologist who was convinced that the unpredictability of weather forecasting was not due to any external noise or randomness but was in fact compatible with a completely deterministic description. In this sense, Lorenz was attempting to make precise the qualitative insight of Poincare, who, in another prescient comment—all the more remarkable for its occurring in the paragraph immediately following our earlier quotation from *Science and Method*-- observed:

"Why have meteorologists such difficulty in predicting the weather with any certainty? . . . We see that great disturbances are generally produced in regions where the atmosphere is in unstable equilibrium. The meteorologists see very well that the equilibrium is unstable, that a cyclone will be formed somewhere, but exactly where they are not in a position to say; a tenth of a degree more or less at any given point, and the cyclone will burst here and not there, and extend its ravages over districts it would otherwise have spared. . . Here, again, we find the same contrast between a very trifling cause that is inappreciable to the observer, and considerable effects, that are sometimes terrible disasters. "

To demonstrate this sensitive dependence, Lorenz began with a simplified model approximating fluid convection in the atmosphere. By expanding this model in (spatial) Fourier modes and by *truncating* the expansion to the three lowest modes and explicitly ignoring couplings to higher modes, Lorenz obtained a closed system of three nonlinear ordinary first-order differential equations:

$$\begin{aligned} \dot{x} &= -\sigma x + \sigma y, \\ \dot{y} &= -xz + rx - y, \text{ and} \\ \dot{z} &= xy - bz. \end{aligned} \tag{23}$$

In the application to atmospheric convection, x measures the rate of convective overturning, y and z measure the horizontal and vertical temperature variations, respectively, σ is the Prandtl number, r is the Rayleigh number, and b ($\neq 1$) reflects the fact that the horizontal and vertical temperature structures do not generally damp at the same rate.

As in the case of the damped, driven pendulum, the model describes a system with one-and-a-half degrees of freedom because it consists of three first-order equations. One set of parameters ($\sigma = 10$, $r = 28$, and $b = 8/3$) yields the celebrated *Lorenz attractor*, for which perspective views (Fig. 9) of the attracting set in the *(x,* y, *z)* space reveal two "lobes" (Fig. 9a) and a thickness in the third direction (Fig. 9b) that shows the set is not planar.

Just as any initial point on a periodic orbit will eventually trace out the full orbit, so here any initial point on this strange attractor will follow a path in time that eventually traces out the full structure. Here, however, nearby initial points will diverge exponentially, reflecting the sensitive dependence on initial conditions. The two-lobed structure of the Lorenz attractor suggests a particularly useful analogy to emphasize this sensitivity. Choose two very nearby initial points and follow their evolution in time. Call each loop around the right lobe "heads" and around the left lobe "tails." Then the asymptotic sequences of heads and tails corresponding to the two points will be completely different and totally uncorrelated to each other. Of course, the nearer the initial points, the longer their motions will remain similar. But for any initial separation, there will be a finite time beyond which the motions appear totally different.

In his original study Lorenz observed this sensitive dependence in an unexpected manner, but one quite consistent with research in experimental mathematics. His own words (from p. 55 of his article in *Global Analysis)* provide a dramatic statement of the observation:

"During our computations we decided to examine one of the solutions in greater detail, and we chose some intermediate conditions which had been typed out by the computer and typed them in as new initial conditions. Upon returning to the computer an hour later, after it had simulated about two months of "weather," we found that it completely disagreed with the earlier solution. At first we expected machine trouble, which was not unusual, but we soon realized that the two solutions did not originate from identical conditions. The computations had been carried internally to about six decimal places, but the typed output contained only three, so that the new initial conditions consisted of old conditions plus small perturbations. These perturbations were amplifying quasi-exponentially, doubling in about four simulated days, so that after two months the solutions were going their separate ways. "

Notice that the doubling period of the small initial perturbation corresponds directly to the binary bit shift of the logistic map at $r = 4$. Again we see the exponential loss of information about the initial state leading to totally different long-time behavior.

Let me now focus on the geometric figure that represents the strange attractor of the Lorenz equations. Figure 9 is, in fact, generated by plotting the coordinates

x(t), y(t), and *z(t)* at 10,000 time steps (after transients have died out) and joining the successive points with a smooth curve. The first 5000 points are colored blue, the second 5000 green. The apparent white parts of the figure are actually blue and green lines so closely adjacent that the photographic device cannot distinguish them.

Notice how the blue and green lines interleave throughout the attractor and, in Fig. 9d, how this interleaving continues to occur on a finer scale. In fact, if the full attractor, generated by the infinite time series of points *(x(t),* y(?), *z* (t)), were plotted,

LORENZ ATTRACTOR

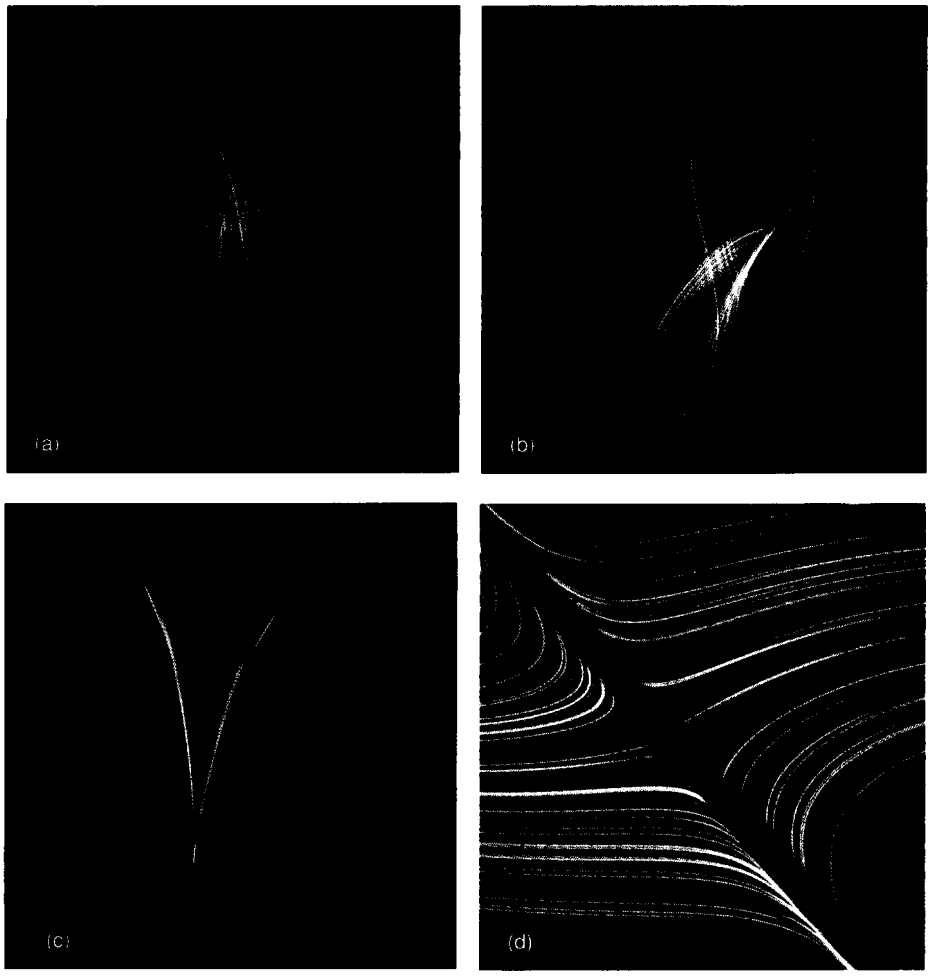


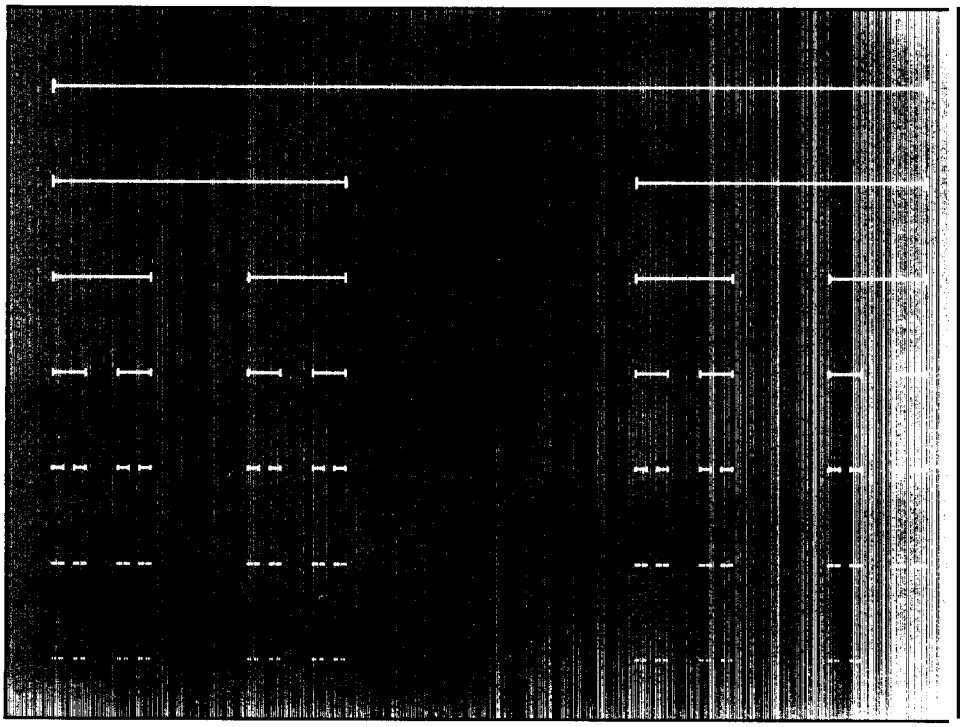
Fig. 9. The attracting set of the Lorenz attractor (Eq. 23 in the text with σ = 10, r = 28, and b = 8/3) formed by joining 10,000 time steps of a single orbit into a smooth curve with the first 5000 points plotted in blue and the second 5000 plotted in green. (a)–(c) These perspective views reveal the two-lobed, nonplanar shape and the thickness of the attractor. The red lines indicate the direction of the coordinate axes. (d) A closeup of the interleaving of the Lorenz orbit, which, even for an infinite time series of points, would never intersect and repeat itself. The attractor has a fractal dimension of 2.04, that is, between that of an area and a volume. (Figure courtesy of Gottfried Mayer-Kress, Los Alamos National Laboratory.)

we would see the trajectory looping around forever, never intersecting itself and hence never repeating. The exquisite filamentary structure would exist on all scales, and, even in the infinite time limit, the attractor would not form a solid volume in the *(x,* y, *z*) space. In a sense that we shall make precise shortly, the attractor is a fractal object with dimension between that of an area (dimension = 2) and that of a volume (dimension = 3). Indeed, the Lorenz attractor has *fractal dimension* of about 2.04.

Fractals. The term fractal was coined by Benoit Mandelbrot in 1975 to describe irregular, fragmented shapes with intricate structure on all scales. Fractals moved into the mainstream of scientific research when it became clear that these seemingly exotic geometric objects, which had previously been viewed as “a gallery of monsters,” were emerging commonly in many natural contexts and, in particular, as the attracting sets of chaotic dynamical systems. In fact, Mandelbrot traced many of the core concepts related to fractals back to a number of distinguished late 19th and early 20th century mathematicians, including Cantor, Hausdorff, and Julia. But, as in the case of deterministic chaos, the flowering of these concepts came only after experimental mathematics made precise visualization of the monsters possible.

The essential feature of fractals is the existence of similar, nontrivial structure on all scales, so that small details are reminiscent of the entire object. Technically, this property is known as *scaling* and leads to a theoretical approach that allows construction of fine details of the object from crude general features. The structure need not be exactly self-similar on all scales. Indeed, much current research focuses on ***self-affine*** fractals, in which the structures on different scales are related by linear transformations.

One consequence of this scale invariance is that fractal objects in general have



CANTOR SET

Fig. 10. The Cantor set is formed by starting with a line segment of unit length, removing its middle third, and, at each successive level, removing the middle third of the remaining segments. Although, the length of the remaining segments goes to zero as the number of iterations, or levels, goes to infinity, the set has a fractal dimension greater than zero, namely ln 2/ ln 3 ≈ 0.6309. (Figure courtesy of Roger Eckhardt, Los Alamos National Laboratory.)

fractional rather than integral dimension: that is, rather than being lines, areas, or volumes, fractals lie "somewhere in between." To understand this quantitatively, we recall the example of the recursively defined Cantor set (Fig. 10). At the zeroth level, the set consists of a continuous line segment from O to 1. At the first level, the middle third of the segment is eliminated. At the second level, the middle third of each of the two remaining continuous segments are eliminated. At the third level, the middle third of each of these four segments is eliminated, and so forth ***ad infinitum.*** At each level the Cantor set becomes progressively less dense and more tenuous, so that the end product is indeed something between a point and **a line.** It is easy to see in Fig. 10 that at the *nth* level, the Cantor set consists of 2" segments, each of length (1/3)". Thus, the "length" 1 of the set as n goes to infinity would be

$$\lim_{\iota\to\infty} l = \lim_{n\to\infty} 2^n \left(\frac{1}{3}\right)^n = 0. \tag{24}$$

In the 1920s the mathematician Hausdorff developed a theory that can be used to study the fractional dimension of fractals such as the Cantor set. In the present simple case, this theory can be paraphrased by asking how many small intervals, N(E), are required to "cover" the set at a length scale ϵ. As $\epsilon \to 0$, the fractal dimension d_f is defined by

$$d_f \equiv \lim_{\epsilon\to 0} \left(\frac{\ln N(\epsilon)}{\ln(\frac{1}{\epsilon})} \right). \tag{25}$$

Inverting Eq. 25, we see that

$$\lim_{\epsilon\to 0} N(\epsilon) = \left(\frac{1}{\epsilon}\right)^{d_f}. \tag{26}$$

continued on page 246

HAMILTONIAN CHAOS and STATISTICAL MECHANICS

The specific examples of chaotic systems discussed in the main text—the logistic map, the damped, driven pendulum, and the Lorenz equations—are all dissipative. It is important to recognize that nondissipative Hamiltonian systems can also exhibit chaos; indeed, Poincare made his prescient statement concerning sensitive dependence on initial conditions in the context of the few-body Hamiltonian problems he was studying. Here we examine briefly the many subtleties of Hamiltonian chaos and, as an illustration of its importance, discuss how it is closely tied to long-standing problems in the foundations of statistical mechanics.

We choose to introduce Hamiltonian chaos in one of its simplest incarnations, a two-dimensional discrete model called the *standard map.* Since this map preserves phase-space volume (actually *area* because there are only two dimensions) it indeed corresponds to a discrete version of a Hamiltonian system. Like the discrete logistic map for dissipative systems, this map represents an archetype for Hamiltonian chaos.

The equations defining the standard map are

$$p_{n+1} = p_n - \frac{k}{2\pi}\sin 2\pi q_n, \qquad (1)$$
$$q_{n+1} = p_{n+1} + q_n,$$

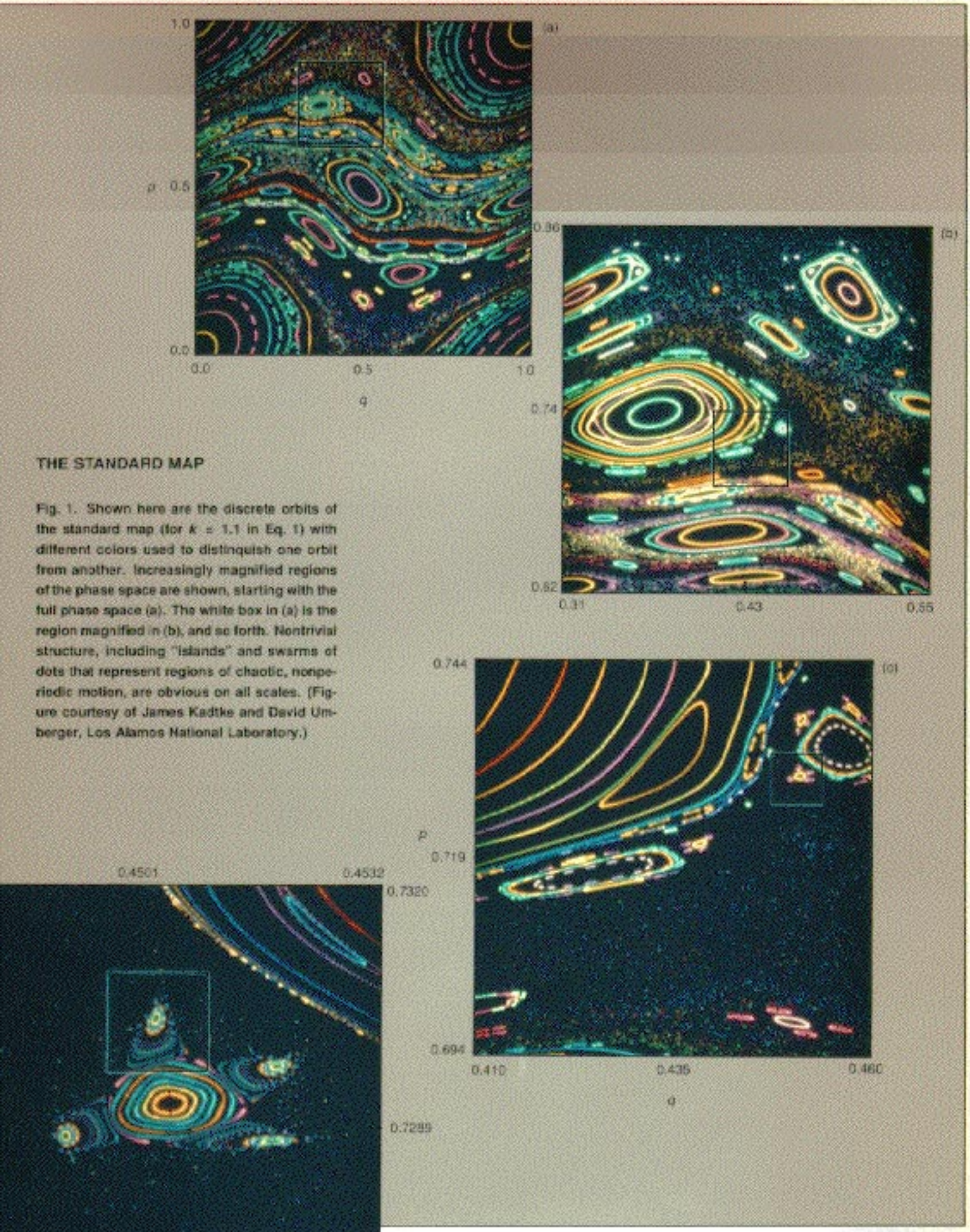
where, as the notation suggests, p_n is the discrete analogue of the momentum, q_n is the analogue of the coordinate, and the discrete index n plays the role of time. Only the fractional parts of p_n and q_n are kept; hence the motion is on a torus, periodic in both p and q. For any value of k, the map preserves the area in the (p, q) plane, since the Jacobian $\partial(p_{n+1}, q_{n+1})/\partial(p_n, q_n) = 1$.

The preservation of phase-space volume for Hamiltonian systems has the very important consequence that there can be no *attractors,* that is, no subregions of lower phase-space dimension to which the motion is confined asymptotically. Any initial point (p_o, q_o) will lie on some particular orbit, and the *image* of all possible initial points—that is, the unit square itself—is again the unit square. In contrast, *dissipative* systems have phase-space volumes that shrink. For example, the logistic map (Fig. 5 in the main text) at $\lambda = 3.1$ has all initial points in the interval (O, 1) attracted to just two points.

Clearly, for $k = O$ the standard map is trivially integrable, with $p_n = p_0$ being constant and q_n increasing linearly in time (n) as it should for free motion. The orbits are thus just straight lines wrapping around the torus in the q direction. For $k = 1.1$ the map produces the orbits shown in Figs. la-d. The most immediately striking feature of this set of figures is the existence of nontrivial structure on all scales. Thus, like dissipative systems, Hamiltonian chaos generates strange fractal sets (albeit "fat" fractals, as discussed below). On all scales one observes "islands," analogues in this discrete case of the periodic orbits in the phase plane of the simple pendulum (Fig. 2 in the main text). In addition, however, and again on all scales, there are swarms of dots coming from individual chaotic orbits that undergo nonperiodic motion and eventually fill a finite region in phase space. In these chaotic regions the motion is "sensitively dependent on initial conditions."

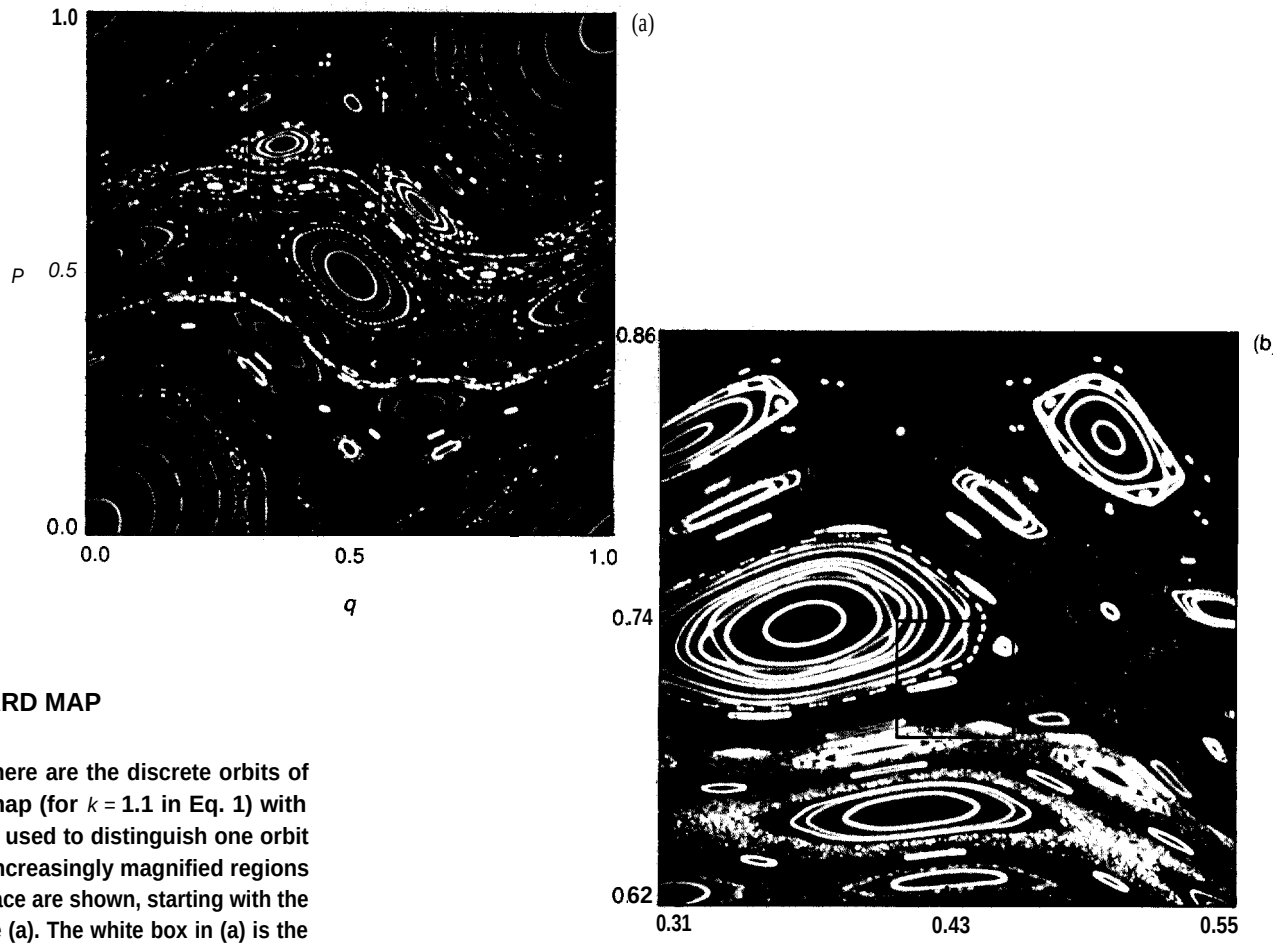
Figure 2 shows, in the full phase space, a plot of a *single* chaotic orbit followed through 100 million iterations (again, for $k = 1.1$). This object differs from the strange sets seen in dissipative systems in that it occupies a *finite* fraction of the full phase space: specifically, the orbit shown takes up 56 per cent of the unit area that represents the full phase space of the map. Hence the "dimension" of the orbit is the same as that of the full phase space, and calculating the fractal dimension by the standard method gives $d_f = 2$. However, the orbit differs from a conventional area in that it contains holes on all scales. As a consequence, the measured value of the area occupied by the orbit depends on the resolution with which this area is measured—for example, the size of the boxes in the box-counting method—and the approach to the finite value at infinitely fine resolution has definite scaling properties. This set is thus appropriately called a "fat fractal," For our later discussion it is important to note that the holes—representing periodic, nonchaotic motion—also occupy a finite fraction of the phase space.

To develop a more intuitive feel for fat fractals, note that a very simple example can be constructed by using a slight modification of the Cantor-set technique



THE STANDARD MAP

Fig. 1. Shown here are the discrete orbits of the standard map (for $k = 1.1$ in Eq. 1) with different colors used to distinguish one orbit from another. Increasingly magnified regions of the phase space are shown, starting with the full phase space (a). The white box in (a) is the region magnified in (b), and so forth. Nontrivial structure, including "islands" and swarms of dots that represent regions of chaotic, nonperiodic motion, are obvious on all scales. (Figure courtesy of James Kadtke and David Umberger, Los Alamos National Laboratory.)



THE STANDARD MAP

Fig. 1. Shown here are the discrete orbits of the standard map (for $k = 1.1$ in Eq. 1) with different colors used to distinguish one orbit from another. increasingly magnified regions of the phase space are shown, starting with the full phase space (a). The white box in (a) is the region magnified in (b), and so forth. Nontrivial structure, including "'islands" and swarms of dots that represent regions of chaotic, nonperiodic motion, are obvious on all scales. (Figure courtesy of James Kadtke and David Umberger, Los Alamos National Laboratory.)

described in the main text. Instead of deleting the middle one-third of each interval at every scale, one deletes the middle $(1/3)^n$ at level n. Although the resulting set is topologically the same as the original Cantor set, a calculation of its dimension yields $d_f = 1$; it has the same dimension as the full unit interval. Further, this fat Cantor set occupies a finite fraction-amusingly but accidentally also about 56 per cent-of the unit interval, with the remainder occupied by the "holes" in the set.

To what extent does chaos exist in the more conventional Hamiltonian systems described by differential equations? A full answer to this question would require a highly technical summary of more than eight decades of investigations by mathematical physicists. Thus we will have to be content with a superficial overview that captures, at best, the flavor of these investigations.

To begin, we note that completely integrable systems can never exhibit chaos, independent of the number of degrees of freedom N. In these systems all bounded motions are quasiperiodic and occur on hypertori, with the *N* frequencies (possibly all distinct) determined by the values of the conservation laws. Thus there cannot be any aperiodic motion. Further, since all Hamiltonian systems with N = 1 are completely integrable, chaos cannot occur for one-degree-of-freedom problems.

For N =2, ***non-integrable*** systems can exhibit chaos; however, it is not trivial to determine in which systems chaos can occur; that is, it is in general not obvious whether a given system is integrable or not. Consider, for example, two very similar N = 2 nonlinear Hamiltonian systems with equation of motion given by:

$$\frac{d^2x}{dt^2} = -x - 2xy, \qquad \frac{d^2y}{dt^2} = -y + y^2 - x^2, \tag{2}$$

and

$$\frac{d2x}{dt^2} = -x - 2xy, \qquad \frac{d^2y}{dt^2} = -y - y^2 - x^2. \tag{3}$$

Equation 2 describes the famous Henon-Heiles system, which is non-integrable and has become a classic example of a simple (astro-) physically relevant Hamiltonian system exhibiting chaos. On the other hand, Eq. 3 can be separated into two independent N = 1 systems (by a change of variables to $\zeta = x - y$ and $\eta = x + y$) and hence is completely integrable.

Although there exist explicit calculational methods for testing for integrability, these are highly technical and generally difficult to apply for large N. Fortunately, two theorems provide general guidance. First, Siegel's Theorem considers the space of Hamiltonians analytic in their variables: non-integrable Hamiltonians are ***dense*** in this space, whereas integrable Hamiltonians are not. Second, Nekhoroshev's Theorem leads to the fact that all non-integrable systems have a phase space that contains chaotic regions.

Out observations concerning the standard map immediately suggest an essential question: What is the extent of the chaotic regions and can they, under some circumstances, cover the whole phase space? The best way to answer this question is to search for *nonchaotic* regions. Consider, for example, a completely integrable N-degree-of-freedom Hamiltonian system disturbed by a generic ***non-integrable perturbation. The*** famous KAM (for Kolmogorov, Arnold, and Moser) theorem shows that, for this case, there are regions of finite measure in phase space that retain the smoothness associated with motion on the hypertori of the integrable system. These regions are the analogues of the "holes" in the standard map. Hence, for a typical Hamiltonian system with N degrees of freedom, the chaotic regions do ***not*** fill all of phase space: a finite fraction is occupied by "invariant KAM tori."

At a conceptual level, then, the KAM theorem explains the nonchaotic behavior and recurrences that so puzzled Fermi, Pasta, and Ulam (see "The Fermi, Pasta, and Ulam Problem: Excerpts from 'Studies of Nonlinear Problems' "). Although the FPU chain had many (64) nonlinearly coupled degrees of freedom, it was close enough (for the parameter ranges studied) to an integrable system that the invariant KAM tori and resulting pseudo-integrable properties dominated the behavior over the times of measurement.

There is yet another level of subtlety to chaos in Hamiltonian systems: namely, the structure of the phase space. For non-integrable systems, within every regular KAM region there are chaotic regions. Within these chaotic regions there are, in turn, regular regions, and so forth. For all non-integrable systems with N > 3, an orbit can move (albeit on very long time scales) among the various chaotic regions via a process known as "Arnold diffusion." Thus, in general, phase space is permeated by an Arnold web that links together the chaotic regions on all scales.

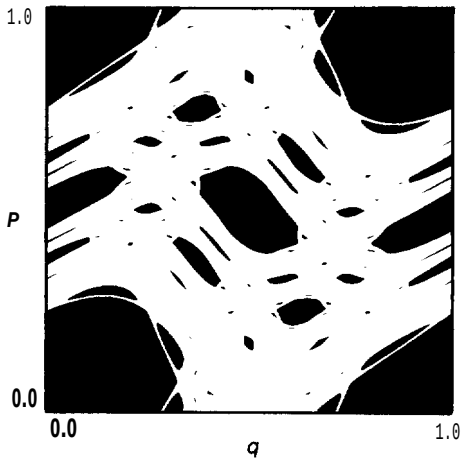
Intuitively, these observations concerning Hamiltonian chaos hint strongly at a connection to statistical mechanics. As Fig. 1 illustrates, the chaotic orbits in Hamiltonian systems form very complicated "Cantor dusts," which are nonperiodic, never-repeating motions that wander through volumes of the phase space, apparently constrained only by conservation of total energy. In addition, in these regions the sensitive dependence implies a rapid loss of information about the initial conditions and hence an effective irreversibility of the motion. Clearly, such wandering motion and effective irreversibility suggest a possible approach to the following fundamental question of statistical mechanics: How can one derive the irreversible, ergodic, thermal-

equilibrium motion assumed in statistical mechanics from a reversible, Hamiltonian microscopic dynamics?

Historically, the fundamental assumption that has linked dynamics and statistical mechanics is the *ergodic hypothesis,* which asserts that time averages over actual dynamical motions are equal to ensemble averages over many different but equivalent systems. Loosely speaking, this hypothesis assumes that all regions of phase space allowed by energy conservation are equally accessed by almost all dynamical motions.

What evidence do we have that the ergodic hypothesis actually holds for realistic Hamiltonian systems? For systems with finite degrees of freedom, the KAM theorem shows that, in addition to chaotic regions of phase space, there are nonchaotic regions of finite measure. These invariant tori imply that ergodicity does *not* hold for most finite-dimensional Hamiltonian sytems. Importantly, the few Hamiltonian systems for which the KAM theorem does not apply, and for which one can prove ergodicity and the approach to thermal equilibrium, involve "hard spheres" and consequently contain non-analytic interactions that are not realistic from a physicist's perspective.

For many years, most researchers believed that these subtleties become irrelevant in the thermodynamic limit, that is, the limit in which the number of degrees of freedom (N) and the energy (E) go to infinity in such a way that ***E/N*** remains a nonzero constant. For instance, the KAM regions of invariant tori may approach zero measure in this limit. However, recent evidence suggests that nontrivial counterexamples to this belief may exist. Given the increasing sophistication of our analytic understanding of Hamiltonian chaos and the growing ability to simulate systems with large N numerically, the time seems ripe for quantitative investigations that can establish (or disprove!) this belief. (For additional discussion of this topic, see "The Ergodic Hypothesis: A Complicated Problem of Mathematics and Physics.")



A "FAT" FRACTAL

Fig. 2. A singles chaotic orbit of the standard map for k = 1.1. The picture was made by dividing the energy surface Into a 512 by 512 grid and iterating the initial condition 10^8 times. The squares visited by this orbit are shown in black. Gaps in the phase space represent portions of the energy surface unavailable to the chaotic orbit because of various quasiperiodic orbits confined to tori, as seen In Fig. 1. (Figure courtesy of J. Doyne Farmer and David Umberger, Los Alamos National Laboratory.)

Among the specific issues that should be addressed in a variety of physically realistic models are the following.

- How does the measure of phase space occupied by KAM tori depend on N ? Is there a class of models with realistic interactions for which this measure goes to O? Are there non-integrable models for which a finite measure is retained by the KAM regions? If so, what are the characteristics that cause this behavior?
- How does the rate of Arnold diffusion depend on N in a broad class of models? What is the structure of important features—such as the Arnold web-in the phase space as N approaches infinity?
- If there is an approach to equilibrium, how does the time-scale for this approach depend on N? Is it less than the age of the universe?
- Is ergodicity necessary (or merely sufficient) for most of the features we associate with statistical mechanics? Can a less stringent requirement, consistent with the behaviour observed in analytic Hamiltonian systems, be formulated?

Clearly, these are some of the most challenging, and profound, questions currently confronting nonlinear scientists. ■

For the Cantor set, if we look at the nth level and use the small interval of length $\epsilon = (1/3)^n$, we know that $N(\epsilon) = 2^n$. Since as $n \to \infty$, $\epsilon \to 0$, we can use

$$d_f = \lim_{\epsilon \to 0} \left(\frac{\ln N(\epsilon)}{\ln \frac{1}{2}} \right) = \lim_{n \to \infty} \frac{\ln 2^n}{\ln 3^n} = \frac{\ln 2}{\ln 3}. \quad (27)$$

continued from page 241

The simple Cantor set has, in effect, only a single scale because the factor of 1/3 is always used in constructing successive levels of the set. In contrast, fractals that arise in chaotic dynamical systems have a range of scales and, typically, different scalings apply to different parts of the set; as a consequence, these more complex sets are sometimes termed multifractals. In his original work on the logistic map, Feigenbaum defined and discussed a *scaling function* that characterized these differences. Recently, a related analytic technique—called the $f(\alpha)$ approach—has been used to provide a detailed understanding of the many different scalings occurring in a variety of chaotic dynamical systems.

Although these constructions and techniques may seem to be just mathematical manipulations, nature abounds with structures that repeat themselves on many different scales and hence have approximate fractal structure. Familiar examples include clouds, lightning bolts, ferns, and, as shown in Fig. 11, snowflakes. Less familiar but technologically significant examples include the growth of dendritic crystals, dielectric breakdown in gas-filled cells, and "viscous fingering" in certain two-fluid flows.

A laboratory experiment illustrating this last phenomenon (Fig. 12) consists of a flat, effectively two-dimensional, cylindrical cell containing a high-viscosity fluid. An inlet in the center of the cell permits the injection under pressure of a second, less viscous fluid (in this case, water). Instead of smoothly and uniformly replacing the viscous fluid in the cell, the water splits into the highly branched, coral-like fractal object shown in Fig. 12. Using a box-counting technique similar to that used to measure the dimension of the Cantor set, one finds that the fractal dimension of the viscous finger is 1.70 ± 0.05. Hence, although it is composed of many thin but highly branched segments, the viscous finger possesses a fractional dimension closer to that of an *urea (d = 2)* than that of a line *(d* = 1).

To understand the processes that create such structures, one can use experimental mathematics to study specific physical models. One such study (Fig. 13) depicts the development of a fractal pattern on a triangular lattice. The model underlying the pattern depends primarily on the local pressure gradients driving the "fluid," but it also incorporates the effects of fluctuations (via a noise parameter) and of anisotropy. The study shows clearly that all the patterns grow primarily at the tips; almost no growth occurs in the "fjord" regions.

Figures 13a and 13b are examples of the fractal structures found when the noise parameter is held constant but the anisotropy *k is* considerably decreased. Notice the striking qualitative similarity between Figs. 11 and 13a. Interestingly, the fractal dimension of both Figs. 13a and 13b is about 1.5; it is *independent* of *k.* In Figs. 13c and 13d the anisotropy is held fixed but the noise is decreased. Here the fractal dimension of both is about 1.7.

Figure 13 makes clear that d_f alone is not sufficient to characterize a fractal, for although both Figs. 13a and 13b have $d_f = 1.5$, there are obvious visual differences. Mandelbrot has defined a number of higher order geometric properties-for example, *lacunarity,* a measure of the typical size of the holes in the fractal—that can be used to characterize fractals more precisely. Lacunarity and other higher-order features are, in effect, geometric restatements of our earlier remarks that multifractals generated by chaotic dynamical systems have a range of scalings and that $f(\alpha)$ and related analytic techniques can be used to study these scalings. A generally unsolved challenge in this area is the *fractal inverse problem*: given $f(\alpha)$ or related quantities, to what extent can one reconstruct the actual fractal set, including perhaps the order in which the points

of the set are generated dynamically?

Practicalities. The impacts of deterministic chaos and fractals are only now beginning to be felt throughout science. The concepts that even simple systems can exhibit incredibly complicated behavior, that simple rules can lead to enormously intricate geometric objects, and that this behavior and these objects can be quantified are now all widely appreciated and are being applied in many fields.

The fractal viscous-fingering phenomenon (Fig. 12) is of enormous technological interest, for it represents a major barrier to the development of efficient advanced oil-recovery techniques. Nearly half the oil deposited in limestone or other porous media **is** typically unrecovered because it remains stuck in the pores. To force out this oil, water is injected into a second well nearby. Viscous fingering limits the effectiveness of this technique, because when one of the thin fingers of water breaks through from the injector to the recovery well, only injected water rather than oil is thereafter recovered. Clearly a full understanding of this fractal phenomenon and ways to control it are of considerable economic importance.

Similarly, a direct application of fractals occurs in the design of the toughened ceramics used as engine parts. These special ceramics are designed to tolerate flaws, such as voids and cracks, without breaking into pieces. The flaws arise primarily from voids that develop during the sintering process and fractures that arise chiefly from the use of hard materials when machining the ceramics. By adding secondary constituents to the ceramics, propagating cracks can be forced to move through the ceramic along tortuous, convoluted routes, causing more energy to be expended than if the route were smooth and regular. Hence, for a given impulse, an irregular crack does not propagate as far through the ceramic and does less overall damage. Convoluted routes should lead to cracks in the form of complex fractal patterns. Indeed, microscopic studies of high performance ceramics have revealed such patterns and established that the higher the fractal dimension of the cracks, the tougher the ceramics.

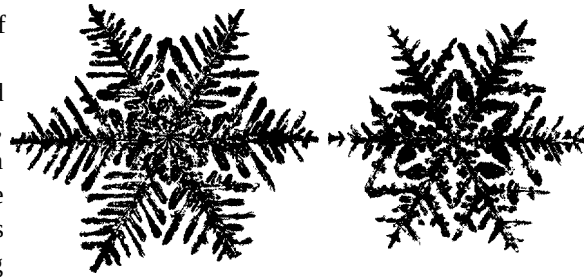
The results of deterministic chaos are also being applied across a broad range of disciplines. Experimentally, high-precision measurements of chaotic dynamics in many types of fluid flows, current and voltage responses of semiconductors and other solid-state electronic devices, and cardiac arhythymias have established the importance of dissipative chaos in fluid dynamics, condensed-matter physics, and medicine. Indeed, recent medical experiments have suggested that many physiological parameters vary chaotically in the *healthy* individual and that greater regularity can indicate a pathological condition; for example, normally chaotic oscillations of the densities of red and white blood cells become periodic in some anemias and leukemias. Hamiltonian chaos finds a direct application in accelerator design, where the potential loss of an apparently stable beam due to subtle long-time phenomena such as "Arnold diffusion" (see "Hamiltonian Chaos and Statistical Mechanics") is a vital issue of technology.

The central theoretical challenge in "applied chaos" is to develop deterministic chaotic models to explain these diverse phenomena. Rather than focusing on the details of specific applications, let me describe two broader problem areas of current research.

First, although we have stressed the randomness and unpredictability of the *long-time* behavior of chaotic systems, it nonetheless remains true that these systems are deterministic, following laws that involve no external randomness or uncertainty. Thus, it is possible to predict the behavior for *short* times, if the equations of motion are known. The analytic solution of the logistic map for r = 4 is a clear illustration; given two initial conditions known to, say, 10-bit accuracy, one can predict the relative positions—albeit with exponentially decreasing accuracy—for 10 iterations of the map. The subtler problem, currently under intense investigation, occurs when one observes that a system is deterministically chaotic but does not know the form of the underlying equations: can one nonetheless use the basic determinism to make *some* prediction? In view of the clear value of such predictive techniques-consider the stock market—

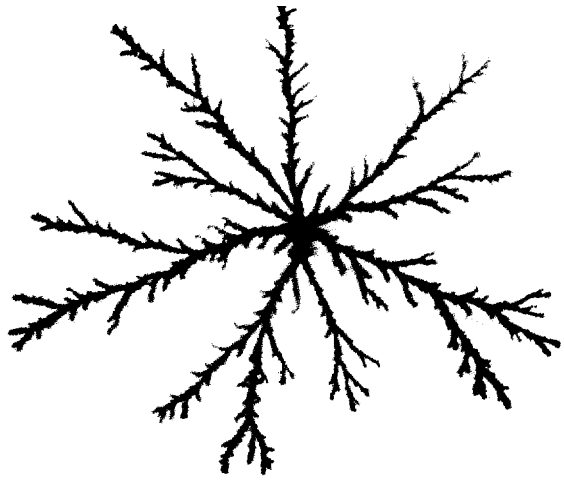
SNOWFLAKES

Fig. 11. The snowflake is an example of a fractal structure **in nature. (Photos reprinted from *Snow Crystals* by W. A. Bentley and W. J. Humpreys with permission of Dover Publications.)**



VISCOUS FINGERING

Fig. 12. A fractal structure formed by injecting water under pressure into a high-viscosity fluid. The fractal dimension of this object has been calculated to be $d_f = 1.70 \pm 0.05$. (Figure courtesy of Gerard Daccord and Johann Nittmann, Etudes et Fabrication Dowell Schlumberger, France, and H. Eugene Stanley, Boston University.)



substantial efforts are being focused on this question.

Second, and at a still broader level, are the related issues of universality and mode reduction in chaos, both of which we mentioned previously. Universality implies that for certain chaotic phenomena—such as the period-doubling cascade—the details of the system and the equations describing it are *irrelevant:* the observed complex behavior develops in a similar manner in every context, be it fluid dynamics, condensed-matter physics, or biology. Indeed, the term universality is borrowed from the statistical mechanics of phase transitions, where it has been shown that the details of the microscopic interactions are irrelevant for most of the important properties of the transitions. In the context of chaos, universality also lends tremendous power to analyses of certain phenomena; in essence, the simplest example—for instance, the logistic map for period doubling---contains the critical features of the entire effect.

A FRACTAL SIMULATION

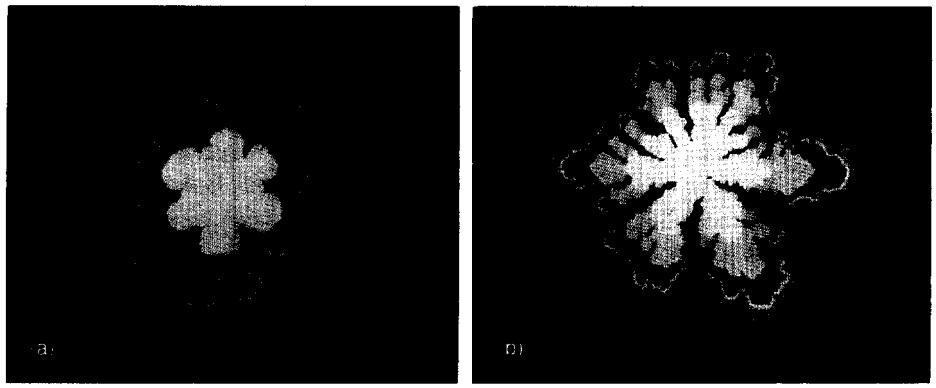


Fig. 13. The model used in the simulation to form these fractal patterns uses local pressure gradients to "drive the fluid" across a triangular lattice. The growth patterns of the objects are indicated by the color coding; the first one-sixth of the sites to be occupied are white, the next one-sixth blue, then magenta, yellow, green, and finally red, The model also incorporates the effects of fluctuations via the noise parameter σ and of anisotropy via the parameter *k.* The patterns in (a) and (b) have the same noise (σ = 0.02) but different anisotropy (*k* = 11 in (a) and *k* = 1.3 in (b)). The patterns in (c) and (d) have the same isotropy (*k* = 1) but the noise changes (from σ = 0.5 in (c) to σ = 0.005 in (d)). (Photos courtesy of Johann Nittmann, Etudes et Fabrication Dowell Schlumberger, France, and H. Eugene Stanley, Boston University.)

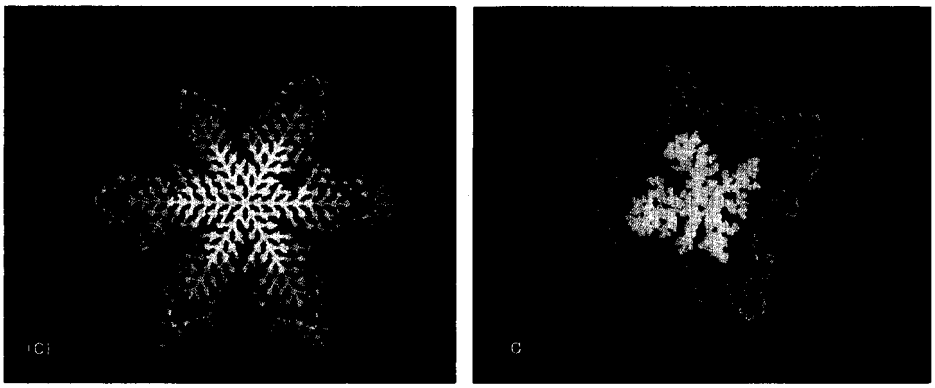
The central idea of mode reduction can most easily be visualized in fluid flows. In any given fluid motion not all the (infinitely!) many possible modes are "active," so the *effective* phase-space dimension is much smaller than the full dimension of the equations. The case of laminar flow in which fluid moves *en bloc is* a trivial illustration. A more interesting and much less obvious example is observed in experiments on Couette-Taylor flows, in which fluid is contained between two concentric rotating cylinders. As the speed of relative rotation is increased, the flow forms bands of Taylor vortices. Further increases in the relative rotation causes the bands to develop "wobbling" instabilities and finally to be replaced by fully developed turbulence. In these experiments clever techniques (sometimes referred to as "geometry from a time series" and related to rigorous mathematical embedding theorems) have been used to extract phase-space information directly from a time series of measurements on a single dynamical variable. Such techniques have revealed strange attractors with effective phase-space dimensions on the order of five. In such experiments there are, in one sense, on the order of only five active modes. Mode reduction reduces the number of degrees of freedom being modeled to the minimum necessary to capture the essence of the motion.

Several important aspects of the general problem of mode reduction should be clarified. First, techniques such as "geometry from a time series" offer no immediate information about the nature of the reduced modes nor about the effective equations governing their interactions. In general, obtaining such information remains an important open problem.

Second, mode reduction is distinct from mode *truncation.* Specifically, we noted that the Lorenz equations were obtained by simply truncating the Fourier expansion of the full equations, hence ignoring certain demonstrably nonzero couplings. Ideally, the process of mode reduction should be *deductive, controlled, and constructive;* that is, one should be able to derive the equations governing the reduced modes, to bound the error made in the reduction, and to "construct" the actual modes themselves. This, too,

remains an elusive goal, despite substantial recent progress.

Third, if one is able to obtain a true mode reduction, the benefits are substantial. For instance, the parameters of the mode-reduced equations can easily be forced in a time-dependent manner, and the reduced equations themselves can be damped and driven. In this manner it may be possible to predict the behavior of the full original system, where the effects of such forcing may be difficult to determine. A specific instance of this technique has been carried out recently by Rabinowitz in the Soviet



Union: using numerical experiments on mode-reduced equations as a guide, he was able to forestall the onset of turbulence in a nozzle flow by applying a periodic stress.

Fourth, rigorous mathematical results on mode reduction have been obtained for a class of nonlinear reaction-diffusion equations that describe unstable chemical reacting fronts, such as flames. One important example is the so-called Kuramoto-Sivashinsky equation, which can be written in the form

$$\frac{\partial\phi(x,t)}{\partial t} + \frac{\partial^2\phi(x,t)}{\partial x^2} + \frac{\partial^4\phi(x,t)}{\partial x^4} + \left(\frac{\partial\phi(x,t)}{\partial x}\right)^2 = 0, \tag{28}$$

where $\phi(x,t)$ is the amplitude, as a function of position and time, of the phenomenon being described. Although this equation represents, in dynamical-systems parlance, an infinite-degree-of-freedom system, it is nonetheless rigorously true that in a box of finite length L a finite number of modes proportional to L are sufficient to capture the long-time dynamics arising from essentially any initial condition. Although the link is not yet fully constructive, the nature of these modes can be determined, and they are related to coherent structures observed for this equation. This general connection between mode reduction in chaotic systems and coherent structures in spatially extended dynamical systems will be a central issue in our discussion of complex configurations and pattern selection.

Finally, the problem of mode reduction lies at the core of attempts to understand the relation between chaos and fully developed turbulence in fluids and plasmas. Chaos, as we have stressed, involves temporal disorder and unpredictability in dynamical systems with low effective phase-space dimension. Fully developed turbulence, in addition to the temporal disorder, involves disordered, random spatial structure on all scales (at least apparently). Further, different spatial regions of the turbulent system act independently, and spatial correlation functions are short-ranged. Thus the phase-space dimension of any attractor in fully developed turbulence appears, a priori, very high.

For example, a recent numerical simulation of turbulent Poiseuille flow at a Reynolds number of 2800 suggests that the turbulent solutions to the Navier-Stokes equations for the flow do lie on a strange attractor, but one that has fractal dimension of about 400! Although it is comforting to know that the turbulence observed in this case can be described qualitatively by deterministic chaos, it is obviously disconcerting

to contemplate trying either to analyze such flows experimentally or to model them theoretically in terms of a dynamical system with a 400-dimensional phase space. For higher Reynolds numbers this situation will become even worse. In the next section we will illustrate how mode reduction, coupled with a hierarchy of approximate equations, may make this situation more tractable.

In sum, the remarkable insights of the past twenty-five years have led to the emergence of deterministic chaos and fractals as a second central paradigm of nonlinear science. The impact of this paradigm on our basic view of complexity in the world, as well as on technologies affecting our daily lives, will continue to be profound for the foreseeable future.

Complex Configurations and Patterns

When a spatially extended nonlinear system is driven far from equilibrium, the many localized coherent structures that typically arise can organize themselves into a bewildering array of spatial *patterns,* regular or random. Perhaps the most familiar example is turbulent fluid flow, in which the temporal behavior is chaotic yet one frequently observes patterns of coherent structures: recall the complex configuration of vortices surrounding the Red Spot in Fig. 3a. The process of pattern formation and selection occurs throughout nature, in nonlinear phenomena ranging from electromagnetic waves in the ionosphere through mesocale textures in metallurgy to markings on seashells and stripes on tigers. Thus, complex configurations and patterns represents a third paradigm of nonlinear science. Although somewhat less developed than solitons or chaos, the paradigm already promises to provide the basis for a unified understanding of nonlinear phenomena in many fields.

Our previous discussion of dynamical systems provides a useful conceptual framework in which to approach the general problem of patterns. A typical extended nonlinear, nonequilibrium system will have many possible configurations or patterns; some of these will be stable, others unstable, and the vast majority metastable. Highly symmetric patterns may be accessible analytically, but general, anisotropic configurations must first be studied via experimental mathematics. In dissipative extended systems these patterns are loosely analogous to the attractors of simple dynamical systems—with the important proviso that they do not correspond to true asymptotic attractors because most are, in fact, merely *metastable.* Nonetheless, the multiple-attractor analogy correctly suggests that an extended nonlinear system has many basins of temporary attraction. In view of our results on the damped, driven pendulum, we expect the basin boundaries to be complicated, perhaps fractal, objects. As a result, the study of the *dynamics* of complex configurations and of the sequence of patterns explored, as well as of the pattern ultimately selected (if any), represents one of the most daunting challenges facing nonlinear science.

At present this challenge is still being confronted primarily at the experimental level, both in actual physical systems and via numerical simulations, rather than analytically. Hence we rely here chiefly on visual results from these experiments to indicate important aspects of the paradigm.

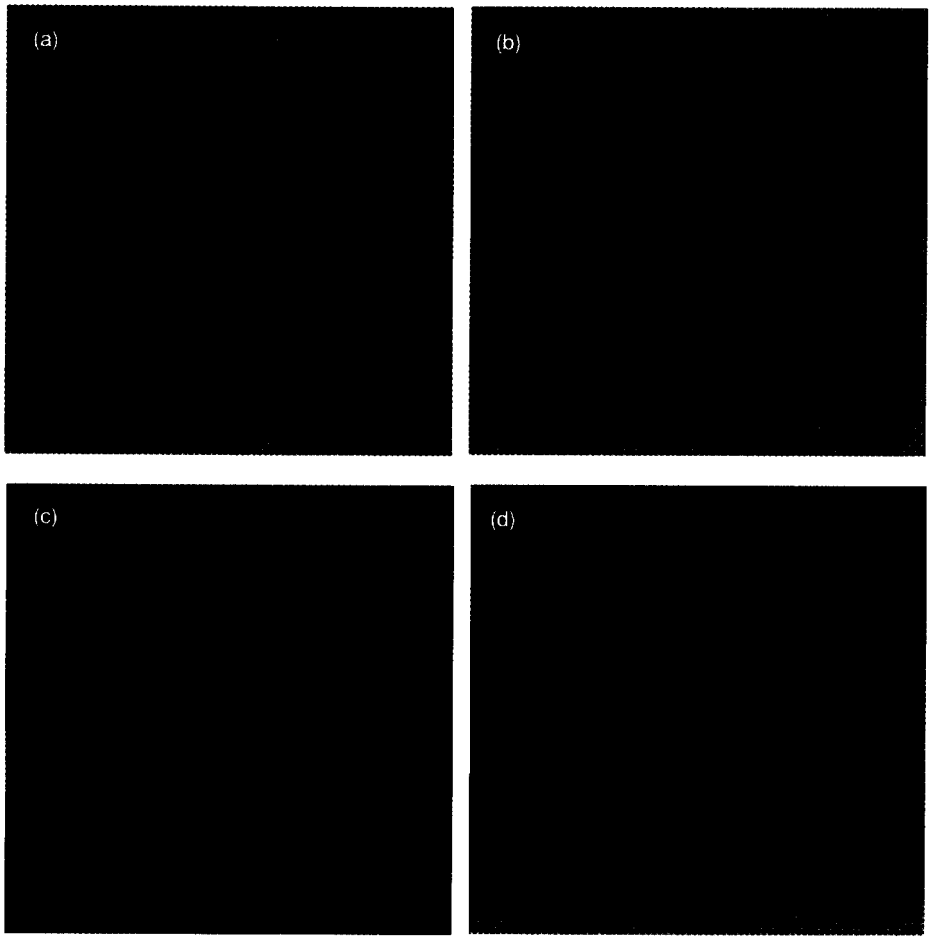
Experiments and Numerical Simulations. Consider, as a first illustration, a generalization of a familiar example: the sine-Gordon equation, only now damped, driven, and with two spatial dimensions. This equation, which models certain planar magnetic materials and large-area Josephson junction arrays, has the form

$$\frac{\partial^2\theta}{\partial t^2} - \frac{\partial^2\theta}{\partial y^2} - \frac{\partial^2\theta}{\partial x^2} + \alpha\frac{\partial\theta}{\partial t} + \sin\theta = \Gamma\cos\omega t\,. \tag{29}$$

We can anticipate from our earlier discussion that this model will contain coherent structures (although not solitons, because the two-dimensional sine-Gordon equation is

not completely integrable). We can also expect the model to contain chaos because of the driving and damping forces.

Four snapshots of the temporal development of the system are shown in **Fig. 14** for $\alpha = 0.1$, $\Gamma = 1.6$, and $\omega = 0.6$. Although it may seem obvious, the use of color coding as a means of enhancing the visual interpretation should be mentioned; color graphics, especially in a high-speed, interactive mode, are *not* a frivolous luxury but, in fact, are among the most powerful tools of experimental mathematics. Here, for example, the color indicates the approximate value of θ. The initial annular structure of this system (Fig. 14a) eventually forms other patterns that are, in fact, oscillatory in time. Because the boundary conditions are periodic in both *x* and y, the system retains a high degree of symmetry as it evolves into four structures in a two-by-two pattern (Fig. 14 b). Eventually, however, a "smearing" is detected parallel to the *x* axis (Fig. 14c) that leads to the striped configuration of Fig. 14d. No further qualitative change occurs after that.



SINE-GORDON EQUATION FOR TWO SPATIAL DIMENSIONS

Fig. 14. Four snapshots of the temporal behavior of the two-dimensional sine-Gordon equation. Red indicates values of θ near 2π, blue indicates values near O, and colors in the spectrum between red and blue indicate intermediate values. (a) The initial structure is annular. (b) After a time equal to approximately 100 units of the fundamental oscillation frequency of the system, the initial ring breaks into a symmetric, two-by-two pattern of four structures. (c) This last pattern is metastable and gradually slides off center, leading, at $t \sim$ 200, to a pattern that is clearly beginning to "smear" in the x-direction. (d) Eventually, for $t \sim$ 300, the smearing has led to a striped, stable configuration. The parameters used in Eq. 29 to generate these pictures are $\alpha = 0.1$, $\Gamma = 1.6$, and $\omega = 0.6$. (Figures courtesy of Peter Lomdahl, Los Alamos National Laboratory.)

Because of the original symmetry of the problem, the emergence of a final pattern striped in the *x* direction rather than the *y* direction must depend on a slight asymmetry external to the equations themselves, Possibilities are a slight difference in the initial conditions for *x* and *y* due to computer round-off or an asymmetry in the solution algorithm. Such asymmetries can be viewed as external noise that leads to a configuration that breaks the symmetry of the equations. The extreme sensitivity of certain pattern selection processes to external noise and to minor asymmetries has already been indicated in the fractal growth models of Fig. 13 and is observed experimentally in a wide variety of contexts, including the growth of dendrites such as the snowflakes of Fig. 11.

The emergence and evolution of configurations related to those seen in the numerical simulations has been the focus of many recent experiments involving Rayleigh-Benard convection. By using shadowgraph techniques that clearly distinguish ascending and descending streams of fluid, convection-roll structures are observed in silicone oil heated from below **(Figs. 15 and 16).** The asymmetric pattern of Fig. 15a is typical of configurations that last for only a few minutes. On the other hand, the more symmetric pattern of Fig. 15b is more stable, maintaining its form for ten minutes or more. Although one does not have a detailed understanding of the phenomenon, clearly boundary effects are causing the system to "pin" itself to these more stable configurations. Figure 16 demonstrates that a *mean structure,* or average pattern, can exist in such flows. The first two images (Figs. 16a and b) are snapshots of the flow, whereas the third (Fig. 16c) is a sum of ten such instantaneous pictures and clearly indicates the presence of a highly symmetrical average configuration.

RAYLEIGH-BERNARD CONVECTION PATTERNS

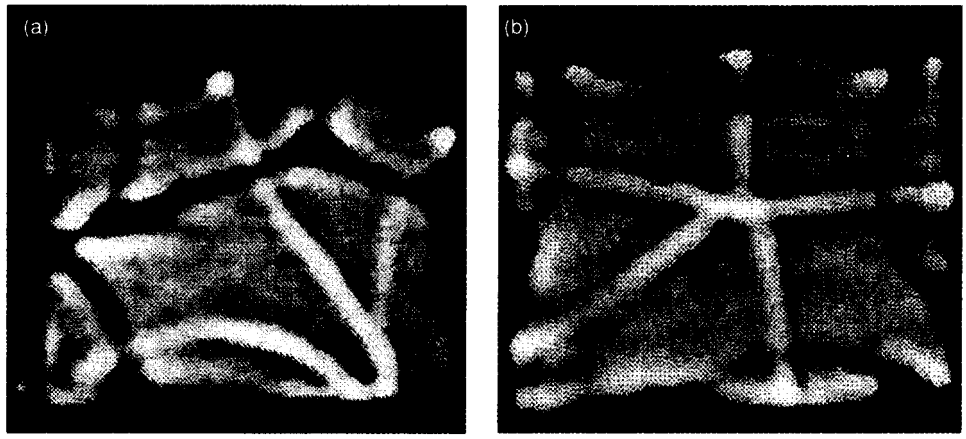


Fig. 15. Patterns of convection-roll streaming are created here using shadowgraph techniques in an experiment in which silicone oil is heated from below. The dark lines correspond to ascending streams of fluid, the bright lines to descending streams. (Photos courtesy of Pierre Berge, Commissariats a L'Energie Atomique, France.)

AN AVERAGE CONVECTION PATTERN

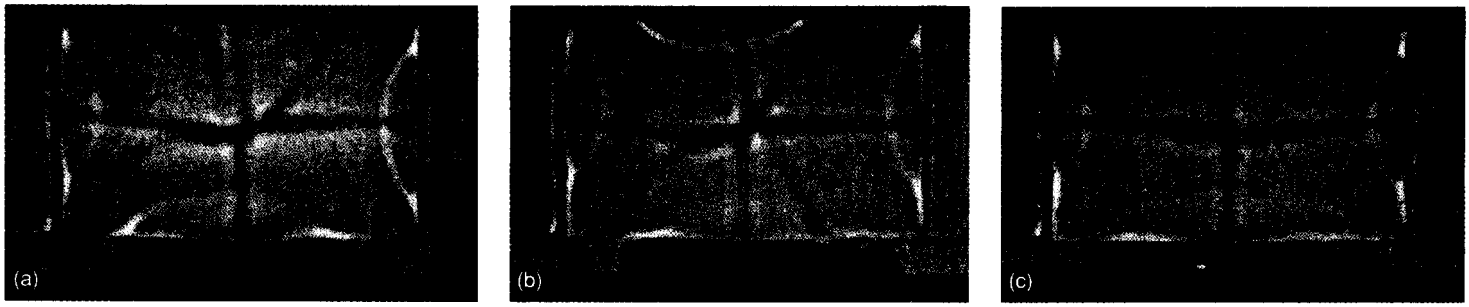
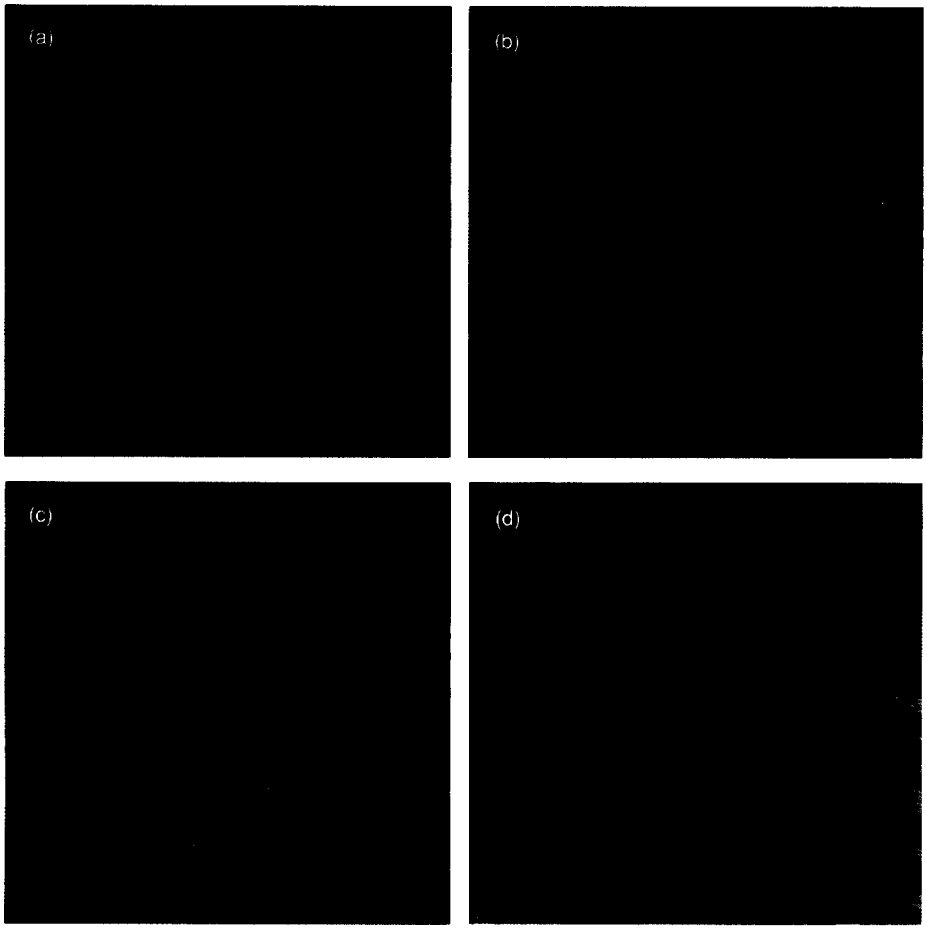


Fig. 16. The first two of these Rayleigh-Bernard convection patterns (a and b) are snapshots of the flow in the silicone oil experiment, whereas (c) is a sum of ten such instantaneous pictures. (Photos courtesy of Pierre Berge, Commissariats a L'Energie Atomique, France.)

Fluid dynamics abounds with other examples of complex configurations and pattern formation. Particularly relevant in technological applications is *shear instability,* which occurs when a fluid moves rapidly past a fixed boundary or when two fluids move past each other at different velocities. The performance and fuel-efficiency of aircraft, for example, are strongly affected by the turbulent boundary layer formed as a consequence of shear instabilities.

Figure 17 is a sequence of images of the "Kelvin-Helmholtz" shear instability simulated using the two-dimensional Euler equations that model compressible but inviscid fluid flow. (Strictly speaking, because the study does not resolve the thin turbulent boundary layer, it is technically a "slip-surface" instability.) The study reveals an incredible wealth of information, only some of which we will discuss here. Two streams of identical fluid flow past each other, both moving at the speed of sound. Initially, a small sinusoidal perturbation is given to the vertical velocity component of the flow at the boundary between the layers, and the resulting entrainment and roll-up phenomena that lead to the mixing of the two fluids is followed.

Shortly after the simulation starts, the roll-up of the boundary has already begun to generate coherent structures (Fig. 17a). These grow (actually, in a self-similar manner) until the periodic boundary conditions in the *x* direction cause the structures to interact (note the four vortex-like structures in Fig. 17b). In addition, sudden jumps in the intensity of the colors in the top and bottom regions reveal the presence of shock waves. The four vortices merge into two (Fig.17c) and thereafter entrain, forming a *bound* vortex pair (Fig. 17d). The roll-up phenomenon creates incredibly complex



SHEAR INSTABILITY

Fig. 17. Two streams of identical fluid flow past each other with the top stream (colored green to blue) moving to the right at Mach 1 (the speed of sound) and the bottom stream (colored red to purple) moving to the left also at Mach 1. The boundary between layers is a yellow line, and, initially, a small vertical sinusoidal velocity perturbation is applied at this boundary with the colors indicating the initial y value of a given bit of fluid. The series show the roll-up of the boundary (a) and the development of coherent structures in the form of vortices. By (d), a bound vortex pair has formed. (Figures made at Lawrence Livermore National Laboratory by Paul Woodward, University of Minnesota, and David Helder.)

(fractal) structure from the initially smooth boundary. Thus, we see in Fig. 17 precisely the interplay between large-scale coherent structures and chaotic, fractal dynamics that typifies complex configurations in extended nonlinear systems. Further, although different in detail, Fig. 17d clearly resembles in outline the more familiar shape of Fig. 3a; art-in this case, computer art-is indeed imitating Nature.

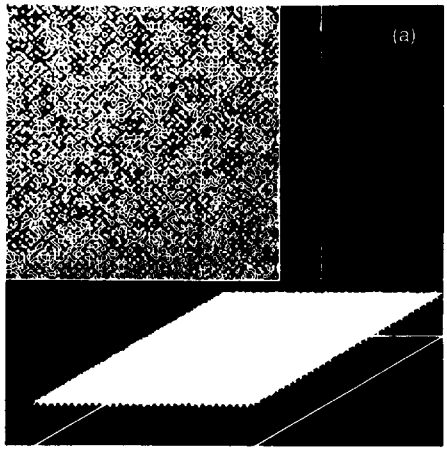
One feature common to all our previous examples is the limited number of coherent structures that participate in the observed patterns of the system. In each case, this limitation arises from the small size (relative to the scale of the coherent structures themselves) of the "box''—be it computational or physical—in which the pattern-forming system is contained.

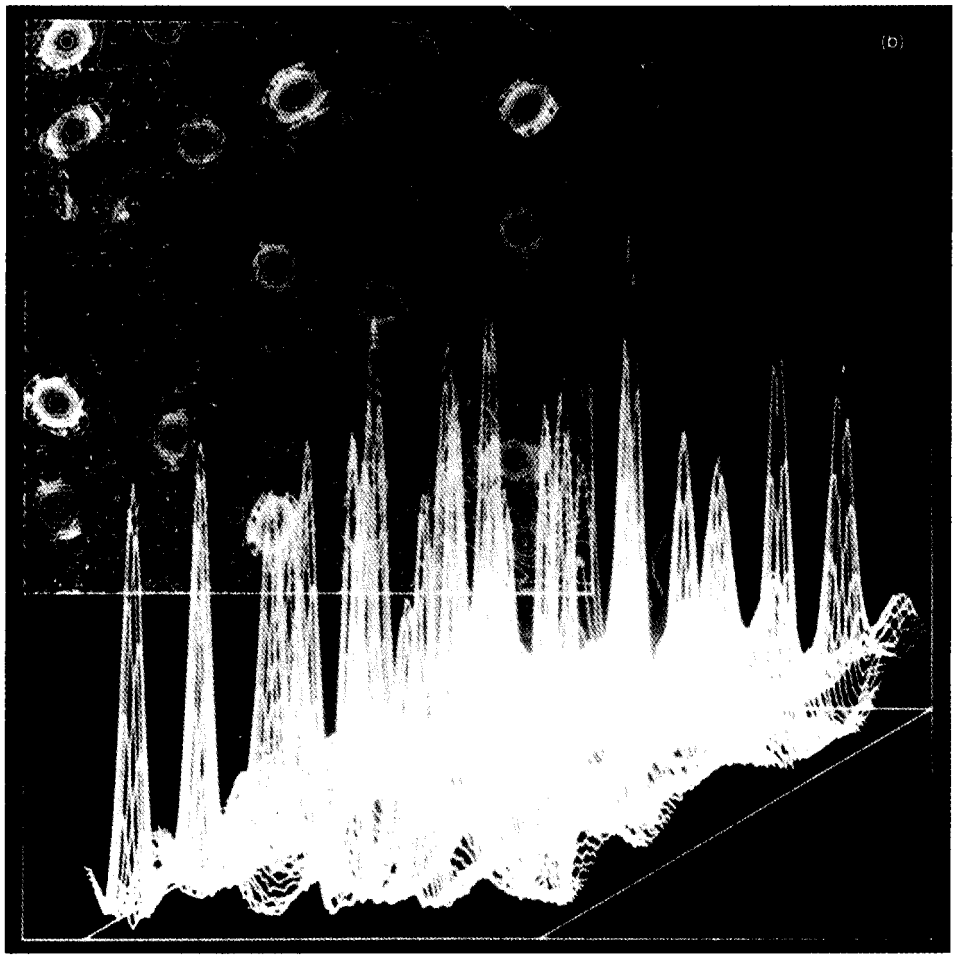
An example in which this constraint is relaxed is a numerical simulation, carried out at Los Alamos a decade ago by Fred Tappert, of the self-focusing instability that arises in the interaction of an intense laser beam with a plasma (Fig. 18). The instability is closely related conceptually to the mechanism by which solitons are formed in optical fibers and reflects an important difficulty in attempts to develop inertial confinement fusion. On a much different scale, this phenomenon leads to significant electromagnetic disturbances in the ionosphere.

The particular equation used in the simulation is a two-space-dimension variant of the nonlinear Schrodinger equation (Eq. 12). Here the equation has the specific form

$$i\frac{\partial E}{\partial t} + \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \left(1 - e^{-|E|^2}\right) E = 0, \tag{30}$$

where $E = E(x, y, t)$ is the electric field envelope function. For small $|E|^2$, the equation contains an effective cubic nonlinearity and thus becomes the direct two-dimensional generalization of Eq. 12. As E approaches infinity, the nonlinearity saturates, and **Eq. 30** becomes effectively linear.





CAVITONS: SELF-FOCUSING INSTABILITIES

Fig. 18. The development of a self-focusing instability in a laser beam passing through a plasma. These frames, taken from a computer-generated movie, show both a contour plot (upper left) and a projected plot of the laser intensity across the profile of the beam. (a) Initially, the beam is essentially uniform with a small amount of random spatial "noise," but as it moves into the plasma, the self-focusing instability generates filaments of high intensity that (b) grow dramatically as the beam progresses further. (Photos made at Los Alamos by Fred Tappert, University of Miami.)

From a random initial condition of spatial white noise (Fig. 18a), a complex configuration involving a large number of coherent structures develops (Fig. 18 b). Having observed these complex patterns involving many coherent structures, Tappert went on to isolate the individual coherent structures—now known as *cavitons*- and to study their interactions numerically. Since the dynamics can not be properly appreciated without showing the time evolution, I will not attempt to describe it here; however, this study is an excellent example of using experimental mathematics to unravel the role that analytically inaccessible coherent structures play in the formation of complex configurations.

Analytic Developments. Our brief pictorial survey of numerical and experimental studies of pattern formation should make clear the daunting nature of the general problem. Thus it is hardly surprising that current analytic approaches focus on special and isolated instances of pattern formation that reduce the problem to a more tractable

form. Although much remains to be done, this "divide and conquer" philosophy has recently demonstrated such substantial promise that, in the next decade, we should witness a revolution for patterns comparable to those for solitons and chaos.

One line of analytic attack is to start with a system that has exact solitons—the one-dimensional sine-Gordon or nonlinear Schrodinger equations, for example. One then perturbs the system with driving and damping forces and studies the patterns that emerge from the evolution of the analytically known coherent structures under the influence of the chaotic dynamics. This approach has been used extensively in the case of the damped, driven sine-Gordon equation, and a very rich phenomenology has developed. However, detailed quantitative understanding, even in the case of a perturbed integrable system, can only be produced at present if the purely analytic approach is guided and supplemented by numerical simulations.

To describe other semi-analytic approaches, let me focus on pattern formation in fluid flows. I shall discuss three related techniques that derive approximate or effective equations appropriate to specific situations called the *amplitude-*, *phase-*, and *protoype-equation* techniques.

As previously observed, a nonlinear system often exhibits bifurcations or sharp transitions in the qualitative behavior of its solutions as a function of one of its parameters. The Rayleigh-Benard instability in a fluid heated from below is one such case (Figs. 15 and 16). When the rate of heating is less than a certain threshold, the fluid simply conducts the heat from the hot bottom to the cooler top, in effect acting like a solid object. At a critical value of the heating rate, this conducting state becomes unstable and *convection-the* familiar rolling motion that can be seen in boiling water—sets in. This transition is the nonequilibrium analog of a phase transition. We can model the temporal and spatial structure of the transition with a phenomenological equation written in terms of a parameter that describes the amplitude of the convecting state. This convection *order parameter* will be zero below threshold and nonzero above. A variety of near-threshold phenomena have been treated successfully using such amplitude equations.

Now consider a nonlinear system already in a state with an overall, regular pattern: for example, a sequence of straight convection rolls in a large box or the rectangular arrangement of convection cells in Fig. 16c. Let this pattern be described by a dominant wave vector (or vectors) that we call k_o. Many patterns close to the initial one can be studied by considering slow spatial and temporal modulation of k_o. The resulting phase equations can be viewed as the nonequilibrium analogs of hydrodynamics since they apply to low-frequency, long-wavelength motions near a given state. Again, such phase equations have been used to analyze many specific fluid flows.

Prototype equations, although perhaps motivated by specific fluid motions, are not necessarily strictly derivable from the fundamental Navier-Stokes equations but rather are intended to capture the essence of certain nonlinear effects. More precisely, prototype equations often serve as a means of gaining insight into competing nonlinear effects and are thus extremely important in developing analytic understanding. The Korteweg-deVries equation (Eq. 10), which played a central role in the discovery of solitons, can be viewed as an example of a prototype equation. That it is also derivable for surface waves in shallow, narrow channels is an added bonus. Similarly, the Kuramoto-Sivashinsky (KS) equation (Eq. 28)—is another prototype equation.

Very recently, pattern formation in convecting fluid flows in large containers has been studied using the Kolmogorov-Spiegel-Sivashinsky, or KSS, equation—a generalization of the original KS equation. Because some very interesting results about the interplay of coherent structures and chaos have come from these studies, I will use the KSS equation to illustrate the prototype-equation technique.

The specific form of the equation is

$$\frac{\partial \phi}{\partial t} + \beta\phi + \gamma\left(\frac{\partial \phi}{\partial x}\right)^2 + \frac{\partial}{\partial x}\left[\left(\alpha - \delta\left(\frac{\partial \phi}{\partial x}\right)^2\right)\frac{\partial \phi}{\partial x}\right] + \frac{\partial^4 \phi}{\partial x^4} = 0, \qquad (31)$$

where α, β, γ, and δ are adjustable parameters. This equation models large-scale unidirectional flow. An example is the Kolmogorov flow in which an effectively two-dimensional viscous fluid is subjected to a unidirectional external force field periodic in one of the spatial directions. Such a flow can be realized in the laboratory using thin layers of electrolytic fluids moving in a periodic magnetic field.

In Eq. 31 $\phi = \phi(x,t)$ is the fluctuating part of the stream function (that is, the stream function minus the mean periodic field component), $\partial\phi/\partial t$ is the familiar local derivative for the fluid motion, ,8@ represents the classical linear damping of the fluctuations, $\gamma(\partial\phi/\partial x)^2$ is the convective derivative for the fluid motion in an unfamiliar form, and the last terms describe viscosity effects.

One can see the role of the local and convective derivative terms more directly by differentiating Eq. 31 with respect to x and considering the gradient of the stream function: $u(x,t) = \partial\phi/\partial x$, which is related to the velocity. In the differentiated equation, the two terms assume the form $\partial u/\partial t + 2\gamma u \partial u/\partial x$, familiar from, for example, the Korteweg-deVries equation (Eq. 10). Note that the convective term in Eq. 31 increases rapidly when ϕ is varying rapidly in x (that is, for large wavenumbers k), correctly suggesting that this term leads to a flow of energy from large to small spatial scales.

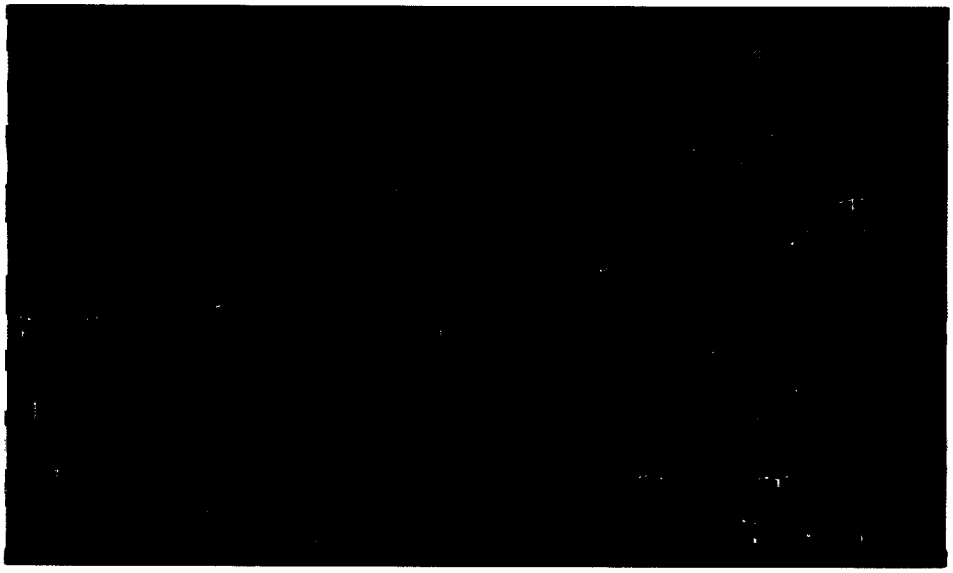
A careful examination of the viscosity effects-given by the final three terms in Eq. 3 l—reveals the interplay and competition essential to the pattern-forming properties of this model. The lowest-order diffusion term, $\alpha\partial^2\phi/\partial x^2$, has, since $\alpha > 0$, the wrong sign for stable diffusion and hence leads to an exponential growth of the solution for all wave numbers k. The higher order linear diffusion term $\partial^4\phi/\partial x^4$, controls the growth for large k. As a consequence, only a certain range of wave numbers ($0 < k < k_c$) exhibits the linear instability that leads to rapid growth. This *negative viscosity* region causes a flow of energy from small to larger spatial scales and thus creates the coherent structures observed in the equations. In turn, the growth of these structures is limited by the nonlinear terms-in particular, by the convective derivative terms-and the resulting competition between the negative viscosity and convective terms provides a mechanism for energy to cycle back and forth between small and large scales. Finally, the nonlinear viscosity term, $\delta\frac{\partial}{\partial x}(\partial\phi/\partial x)^3$, provides an important *local* variation in viscosity; in regions where $\partial\phi/\partial x$ is small, the effective local viscosity is negative, but as $\partial\phi/\partial x$ grows, the nonlinearity leads to a locally positive viscosity and to normal dissipation of energy at small scales.

For the KSS equation, recent analytic studies have shown that the full partial differential equation is strictly equivalent to a set of coupled ordinary differential equations corresponding to a finite-dimensional dynamical system. Further, the phase-space dimension of this dynamical system is proportional to the number of linearly unstable modes and hence increases linearly with the length of the system L. In addition, the finite dynamical system possesses a universal strange attractor with fractal dimension also proportional to L. These results are concrete examples of the mode-reduction program, and their attainment in an equation involving *local* negative viscosity effects marks a step forward in analytic understanding of turbulence. However, as in the case of the KS equation, the results are not of themselves sufficient to identify the natural coherent structures that arise in Eq. 31 nor to relate them directly to the reduced modes.

To search for the coherent structures, extensive numerical simulations of the KSS equation are currently in progress. Figure 19 depicts one solution (for $\alpha = 2$, $\beta = 0.15$, $\gamma = 1$, and $\delta = 0.58$) with a system size such that there are fifty unstable modes. The black cross-hatched structures are regions of (spatially homogeneous) chaos. Note that, with the horizontal axis representing time and the vertical axis representing position, these regions often propagate through the system, as indicated by the diagonal "motion" of the cross-hatched areas.

The most striking features in Fig. 19 are the orange horizontal bands, which intermittently appear and disappear at various locations and times within the system.

These are relatively quiescent, large-scale spatial subdomains and are the candidates for the coherent structures. Notice that the propagating chaotic regions do not penetrate these structures. However, as suggested in particular by the region around the long-lived coherent structure in the lower right comer, one may be able to describe interactions of the propagating chaotic regions with the coherent structures. Such interactions may involve phase shifts (as in the case of solitons) as well as creation and annihilation of both the propagating chaos and the coherent structures. At present, these and related



THE KSS SYSTEM

Fig. 19. This solution of the KSS equation (for α = 2, β = 0.15, γ = 1, and δ = 0.58 in Eq. 31) has both regions of chaos (crosshatched) and regions of relatively quiescent behavior (orange). Since time is represented by the horizontal axis and the spatial variable by the vertical axis, the diagonal 'motion" of crosshatched areas represents propagation of these chaotic regions through the system. (Photo courtesy of Basil Nicolaenko and Hughes Chate, Los Alamos National Laboratory.)

questions are under active investigation.

From our discussion it is clear that, although exciting results are beginning to appear, development of the paradigm of complex configurations and pattern formation will occupy researchers in nonlinear science for years to come. It is perhaps of interest to suggest a few of the broad questions that must be addressed.

- To what extent can the complex structures and patterns be thought of as superpositions of coherent structures, and in what "space" can these structures be superposed? In this regard, we know that for weakly perturbed soliton-bearing systems, the appropriate space in which approximate superposition holds is the inverse scattering transform space. Further, some recent studies suggest that certain turbulent flows can usefully be decomposed as sums of terms, each having vorticity parallel to the velocity.
- What is the dynamics of competition among patterns? How does this competition depend on the nature of the interactions among individual coherent structures? For what systems can one view the different patterns as local minima in a "pattern accessibility" space? What can one say about the basins of attraction in this space?
- In systems with constrained geometry-such as the fluid experiment of Figs. 15 and 16---can one understand quantitatively the observed selection of more symmetric patterns over less symmetric ones? Here the analogy to pinning phenomena in solid state systems may be useful.
- For what pattern-forming systems can one construct a hierarchy of equations in which successive levels of approximation lead downward from the Navier-Stokes equations through an approximate partial differential equation to a finite set of coupled ordinary differential equations? How can one match the solutions across various levels of this hierarchy? Such matching will be essential, in particular to understand what happens when the effective equations lower in the hierarchy break down.

• What can one say about pattern formation in fully-developed, three-dimensional turbulence? For the full Navier-Stokes equations, can anything analogous to the competing mechanisms in the KSS equation be identified, so that both a cycle involving a flow of energy from large to small spatial scales and the re-emergence of large-scale coherent structures can exist? If so, this cycle could form the basis for a "turbulence engine," which would explain at least the major features of transport in turbulent flows.

Since most of our remarks have focused on problems in fluid dynamics, it is important to re-emphasize the broad impact of our last paradigm. The complex fractal structures observed in ceramic cracks and in oil recovery problems, although treated for convenience in our discussion of fractals, are, in fact, more accurately viewed as examples of patterns. Similarly, dendritic growth is a solidification process critically dependent on a pattern selection mechanism that is itself exquisitely sensitive to anisotropy and extrinsic noise. The development of mesoscale textures—that is, patterns larger than the atomic scale but yet not macroscopic—remains an important issue for metallurgy.

In fact, in the microscopic theories of solid state materials, the mechanism underlying pattern dynamics is a question not yet fully resolved. Here, in distinction to the case of fluids, one does not have a fundamental model such as the Navier-Stokes equations to rely on, so one cannot naively assume diffusive coupling among the patterns. Instead, a variety of possible mechanisms must be looked at closely.

In biology, pattern formation and selection is ubiquitous, with applications from the cellular to the whole organism level. And in ecology, nonlinear reaction-diffusion equations suggest spatial patterns in predator-prey distributions and in the spread of epidemics.

To conclude this section, I will look at an intriguing feature of nonlinear pattern-forming systems—the property of pattern *self reproduction-using* a *cellular automaton.* Cellular automata are nonlinear dynamical systems that are discrete in both space and time and, importantly, have, at each site, a finite number of state values (allowed values of the dependent variable). Such systems were invented, and first explored, by John von Neumann and Stan Ulam. Currently, they are being studied both for their fascinating intrinsic properties and for a number of applications, including pattern recognition. They are also being used as novel computational algorithms for solving continuum partial differential equations (see "Discrete Fluids" for the example of lattice-gas hydrodynamics).

Figure 20 shows four stages in the growth of a self-reproducing pattern found in a cellular atuomaton with eight possible states per site. At each step in time, the new state of a given cell is determined by a small set of rules based on the current state of the cell and the state of its nearest four neighbors on a square lattice. The particular pattern shown generates copies of itself, forming a colony. On an infinite lattice the colony would continue to grow forever. Despite its simplicity and the rigidity of its predetermined rules, the self-reproduction of this automaton is intriguingly reminiscent of the development of real organisms, such as coral, that grow in large colonies.

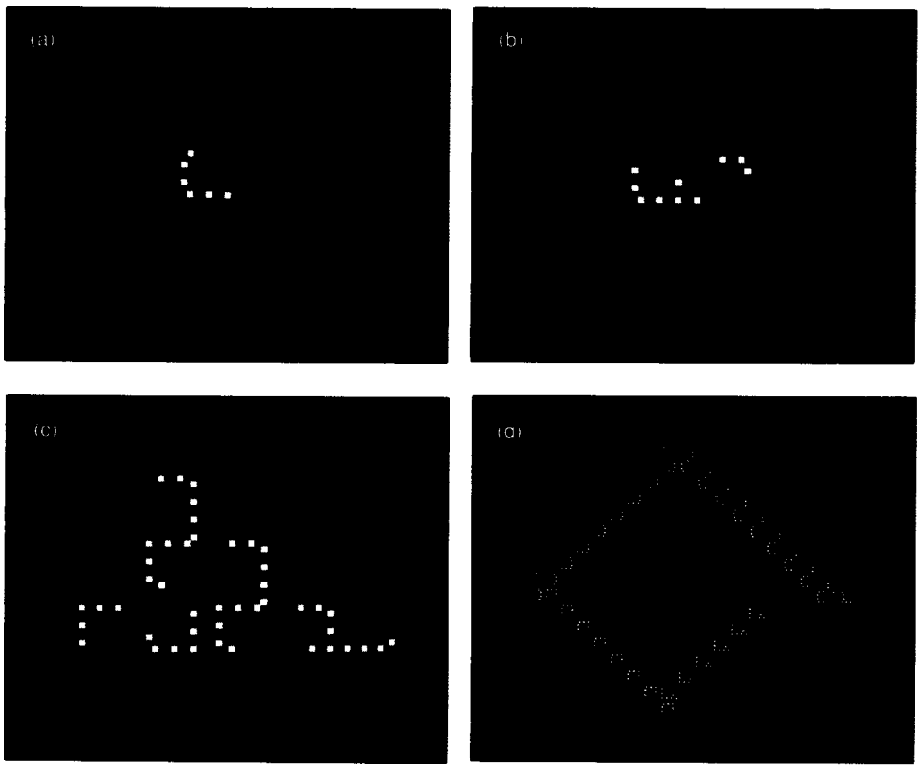
The Future of Nonlinear Science

From the many open questions posed in the previous sections, it should be clear that nonlinear science has a bright and challenging future. At a fundamental level issues such as the scaling structure of multifractal strange sets, the basis for the ergodic hypothesis, and the hierarchy of equations in pattern-forming systems remain unresolved. On the practical side, deeper understanding of the role of complex configurations in turbulent boundary layers, advanced oil recovery, and high-performance ceramics would provide insight valuable to many forefront technologies. And emerging solutions to problems such as prediction in deterministically chaotic systems or modeling fully developed turbulence have both basic and applied consequences. Further, the nonlinear revolution

promises to spread to many other disciplines, including economics, social sciences, and perhaps even international relations.

If, however, one had to choose just one area of clearest future opportunity, one would do well to heed another of Stan Ulam well-known *bons mots:*

> "Ask not what mathematics can do for biology,
> Ask what biology can do for mathematics. "



CELLULAR AUTOMATON

Fig. 20. This cellular automaton consists of a grid of square cells with each cell able to take on any of eight possible states (indicated by different colors). (a) The basic building block of a repeating pattern for this automaton is a hollow square occupying an area of 10 by 15 cells with a tail that develops (b) until it produces a second hollow square. (c) The pattern continues to grow in time until (d) it has produced a large colony of the original pattern. (Figures courtesy of Chris Langton, Los Alamos National Laboratory.)

If we replace "mathematics" with "nonlinear science," Stan's comment becomes even more appropriate to the present situation. We have already seen the beginnings of an understanding of many aspects of morphology, from fractal structures in ferns to nonlinear pattern-selection models for human digits. Similarly, the role of chaos in biological cycles, from heartbeats to cell densities, is rapidly being clarified. And the basic observation that incredibly complex behavior—including both pattern formation and self-reproduction-can emerge in systems governed by very simple rules has obvious implications for modeling biological phenomena.

But the greatest challenge is clearly to understand adaptation, learning, and evolution. *Adaptive* complex systems will have features familiar from conventional dynamical systems, including hierarchical structures, multiple basins of attraction, and competition among many metastable configurations. In addition, they must also have a mechanism for responding to, and taking advantage of, changes in their environment.

One approach to adaptation is to construct an explicit temporal hierarchy: one scale describes the actual dynamics and a second, slower time scale allows for changes in the nonlinear equations themselves. Models for the human immune system and for autocatalytic protein networks are among the prospective initial applications for this concept.

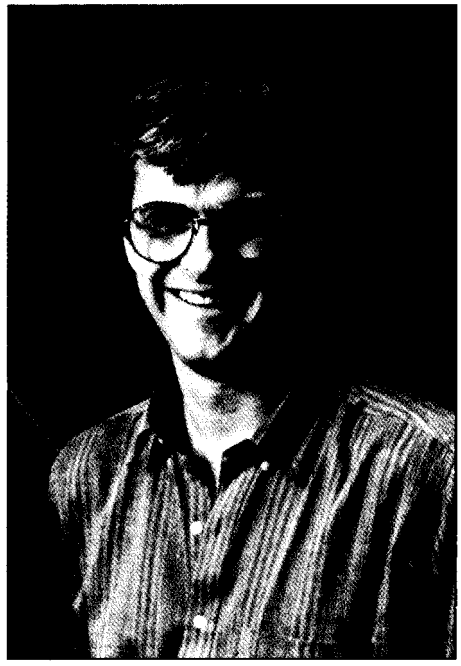
A second approach to adaptation, sometimes termed *connectionism,* is based on the idea that many simple structures exhibit complex collective behavior because of connections between the structures. Recent specific instances of this approach include mathematical models called *neural networks.* Although only loosely patterned after

true neurological systems, such networks show remarkable promise of being able to learn from experience. A related set of adaptive models, called *classifier systems*, show an ability to self-generate a hierarchy of behavioral rules: that is, the hierarchy is not placed a priori into the system but develops naturally on the basis of the system's experience. In general, connectionist models suggest a resolution of the long-standing issue of building a reliable computer from unreliable parts.

In all these future developments, the tripartite methodology incorporating experimental mathematics, real experiments, and novel analytic approaches will continue to play a critical role. One very exciting prospect involves the use of ultraspeed interactive graphics, in which enormous data sets can be displayed visually and interactively at rates approaching the limits of human perception. By using color and temporal evolution, these techniques can reveal novel and unexpected phenomena in complicated systems.

To insure the long-term success of nonlinear science, it is crucial to train young researchers in the paradigms of nonlinearity. Also, interdisciplinary networks must be fostered that consist of scholars who are firmly based in individual disciplines but are aware of, and eager to understand, developments in other fields.

In all these respects, nonlinear science represents a singularly appropriate intellectual legacy for Stan Ulam: broadly interdisciplinary, intellectually unfettered and demanding, and—very importantly-fun. ■



David Campbell, the Laboratory's first J. Robert Oppenheimer Fellow (from 1974 to 1977), is currently the Director of the Laboratory's Center for Nonlinear Science. He received his B.A. in chemistry and physics from Harvard College in 1966 and his Ph.D. in theoretical physics from Cambridge University in 1970. David has extended his activities in physics and nonlinear science to the international level, having been a National Academy of Sciences Exchange Scientist to the Soviet Union in 1977, a Visiting Professor at the University of Dijon. Dijon, France in 1984 and 1985, and a Ministry of Education Exchange Scientist to tbe People's Republic of China in 1986. He and his wife, Ulrike, have two children, Jean-Pierre and Michael.

Acknowledgments

I am grateful to the many colleagues who, over the years, have shared their insights and helped shape my perspective on nonlinear science. In the preparation of this article I have benefited greatly from the comments and assistance of Jim Crutchfield, Roger Eckhardt, Doyne Farmer, Mitchell Feigenbaum, Jim Glimm, Erica Jen, and Gottfried Mayer-Kress. I wish also to thank my generous coworkers in nonlinear science who permitted the use of the figures and pictures so essential to this article.

Further Reading

General Discussions of Nonlinear Science and Background

H. H. Goldstein and J. von Neumann. 1961. On the principles of large scale computing machines. In John *von Neumann.. Collected Works,* Volume V, edited by A. H. Taub, pp. 1–32. New York: Pergamon Press.

E. Fermi, J. Pasta, and S. Ulam. 1965. Studies of nonlinear problems. In *Enrico Fermi: Collected Papers,* Volume II, pp. **978–988.** Chicago: University of Chicago Press.

David Campbell, Jim Crutchfield, Doyne Farmer, and Erica Jen. 1985. Experimental mathematics: the role of computation in nonlinear science. *Communications of the ACM 28:374-384.*

Solitons and Coherent Structures

D. J. Korteweg and G. DeVries. 1895. On the change of form of long waves advancing in a rectangular canal and on a new type of long stationary waves. *Philosophical Magazine 39:422-443.*

Alfred Segger, Hans Donth, and Albert Kochendorfer. 1953. Theorie der Versetzungen in eindimensionalen Atomreihen. III. Versetzungen, Eigenbewegungen and ihre Weshselwirkung. *Zeitschrift* fur *Physik. 134: 171–193.*

J. K. Perring and T. H. R. Skryme. 1962. A model unified field equation. *Nuclear Physics 31:550-555*

N. Zabusky and M. Kruskal. 1965. Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Physical Review Letters 15:240-243.*

Alwyn C. Scott, F. Y. F. Chu, and David W. McLaughlin. 1973. The soliton: a new concept in applied science. *Proceedings of the IEEE 61: 1443–1483.*

A. R. Osborne and T. L. Burch. 1980. Internal solitons in the Andaman Sea. *Science 208:451-460.*

C. H. Tze. 1982. Among the first texts to explain the "soliton revolution." *Physics Today,* June 1982, 55–56. (This article is a review of *Elements of Soliton Theory* by G. L. Lamb, Jr., *Solitons: Mathematical Methods for Physicists* by G. Eilenberger, and *Solitons and the Inverse Scattering Transform* by M. Ablowitz and H. Segur.)

Akira Hasegawa. 1984. Numerical study of optical soliton transmission amplified periodically by the stimulated Raman process. *Applied Optics 23:3302–3309.*

David K. Campbell, Alan C. Newell, Robert J. Schrieffer, and Harvey Segur, editors. 1986. *Solitons and Coherent Structures: Proceedings of the Conference on Solitons and Coherent Structures held at Santa Barbara.* Amsterdam: North-Holland Publishing Co. (reprinted from *Physica D 18:Nos. 1–3).*

C. G. Slough, W. W. McNairy, R. V. Coleman, B. Drake, and P. K. Hansma. 1986. Charge-density waves studied with the use of a scanning tunneling microscope. *Physical Review B 34:994-1005.*

L. F. Mollenauer, J. P. Gordon, and M. N. Islam. 1986. Soliton propagation in long fibers with periodically compensated loss. *IEEE Journal of Quantum Electronics QE-22: 157–173.*

Deterministic Chaos and Fractals

Henri Poincare. 1952. *Science and Method,* translated by Francis Maitland. New York: Dover Publications, Inc.

W. A. Bentley and W. J. Humphreys. 1962. *Snow Crystals. New* York: Dover Publications, Inc.

M. V. Berry. 1978. Regular and irregular motion. In *Topics in Nonlinear Dynamics: A Tribute to Sir Edward Bullard,* edited by S. Jorna, A.I.P. Conference Proceedings, No. 46, pp. 16-120. *New* York: American Institute of Physics.

Edward N. Lorenz. 1979. On the prevalence of aperiodicity in simple systems. In *Global* Analysis: *Proceedings of the Biennial Seminar of the Canadian Mathematical Congress,* edited by M. Grmela and J. E. Marsden, pp. 53–75. New York: Springer Verlag.

Mitchell J. Feigenbaum. 1980. Universal behavior in nonlinear systems. Los *Alamos Science* 1 (Summer 1980):4-27 (reprinted in *Physica D 7:16-39, 1983).*

Robert H. G. Helleman. 1980. Self-generated chaotic behavior in nonlinear mechanics. In *Fundamental Problems in Statistical Mechanics,* edited by E. G. D. Cohen, pp. 165-233. Amsterdam: North-Holland Publishing Co.

B. A. Huberman, J. P. Crutchfield, and N. H. Packard. 1980. Noise phenomena in Josephson junctions. *Applied Physics Letters 37:750-753.*

J.-P. Eckmann. 1981. Roads to turbulence in dissipative dynamical systems. *Reviews of Modern Physics* 53x543454.

G. Mayer-Kress and H. Haken. 1981. The influence of noise on the logistic model. *Journal of Statistical Physics 26:149-171.*

Benoit Mandelbrot. 1983. *The Fractal Geometry of Nature. New* York: W. H. Freeman and Company.

David K. Umberger and J. Doyne Farmer. 1985. Fat fractals on the energy surface. *Physical Review Letters 55:661-664.*

Gerard Daccord, Johann Nittmann, and H. Eugene Stanley. 1986. Radial viscous fingers and diffusion-limited aggregation: fractal dimension and growth sites. *Physical Review Letters 56:336-339.*

Johann Nittmann and H. Eugene Stanley. 1986. Tip splitting without interracial tension and dendritic growth patterns arising from molecular anisotropy. *Nature 321 :663-668.*

James Gleick. 1987. *Chaos: Making a New Science. New* York: Viking Penquin, Inc.

David K. Campbell. 1987. Chaos: chto delat? To be published in *Nuclear Physics B,* Proceedings of Chaos, 1987.

Complex Configurations and Pattern Formation

J. D. Farmer, T. Toffoli, and S. Wolfram, editors. 1984. *Cellular Automata: Proceedings of an interdisciplinary Workshop, Los Alamos.* Amsterdam: North-Holland Publishing Co. (reprinted from *Physica D* 10:Nos. 1–2).

Alan R. Bishop, Laurence J. Campbell, and Paul J. Channell, editors. 1984. *Fronts, lntefaces, and Patterns: Proceedings of the Third Annual International Conference of the Center for Nonlinear Studies, Los Alamos.* Amsterdam: North-Holland Publishing Co. (reprinted from *Physica D 12:1-436).*

Basil Nicolaenko (Nichols). 1987. Large scale spatial structures in two-dimensional turbulent flows. To be published in *Nuclear Physics B,* Proceedings of Chaos, 1987.

Pierre Berge 1987. From temporal chaos towards spatial effects. To be published in *Nuclear Physics B,* Proceedings of Chaos, 1987.

Paul R. Woodward, David H. Porter, Marc Ondrechen, Jeffrey Pedelty, Karl-Heinz Winkler, Jay W. Chalmers, Stephen W. Hodson, and Norman J. Zabusky. 1987. Simulations of unstable flow using the piecewise-parabolic method (PPM). In *Science and Engineering on Cray Supercornputers: Proceedings of the Third International Symposium,* pp. 557–585. Minneapolis: Cray Research, Inc.

Alan C. Newell. The dynamics of patterns: a survey. In *Propagation in Nonequilibrium Systems,* edited by J. E. Wesfried. New York: Springer Verlag. To be published.

Adaptive Nonlinear Systems

Doyne Farmer, Alan Lapedes, Norman Packard, and Burton Wendroff, editors. 1986. *Evolution, games, and learning: models for adaptation in machines and nature: Proceedings of the Fifth Annual International Conference of the Center for Nonlinear Studies, Los Alamos.* Amsterdam: North-Holland Publishing Co. (reprinted from *Physica D 22:Nos.* 1-3).