10

Some Practical Fractals and Their Measurement

The equality that we demand is the most endurable degree of inequality.
—GEORG CHRISTOPH LICHTENBERG

In this chapter we shall pay a belated visit to the world of practical fractals. As Mandelbrot and others have so aptly observed, nature loves fractals at least as much as regular shapes. For every smooth curve or surface seen around us, there are as many, if not many more, that are highly irregular and often in fact fractal, with detailed structure on many size scales.

Why are fractals so prevalent in nature? The overriding reason is that a smoothly curved surface embodies an inherent length scale: the radius of curvature. And for such an inherent length there must of course be a reason. For example, the earth, seen from afar, is a ball with a curved surface with a radius of curvature, R. What is this R? It can be expressed in terms of the total mass M of the earth and its mean density ρ:

$$R = \left(\frac{3}{4\pi} \frac{M}{\rho}\right)^{1/3}$$

With $M \approx 6 \cdot 10^{24}$ kg and $\rho \approx 6 \cdot 10^{3}$ kg/m^{3}, the radius becomes about 6000 km. Thus, the earth ball's natural length scale is ultimately determined by the amount of aggregating dust, and its mean density, that formed the primordial earth 4.7 billion years ago.

But many objects and laws of nature, by luck, lack such a natural scale within the range of observation. Thus, whatever is true at one magnification must be true over a whole range of magnifications. In other words, the object must exhibit self-similarity—statistical, asymptotic, or even strict self-similarity. And if the object has any structure, then a similar structure must appear on many
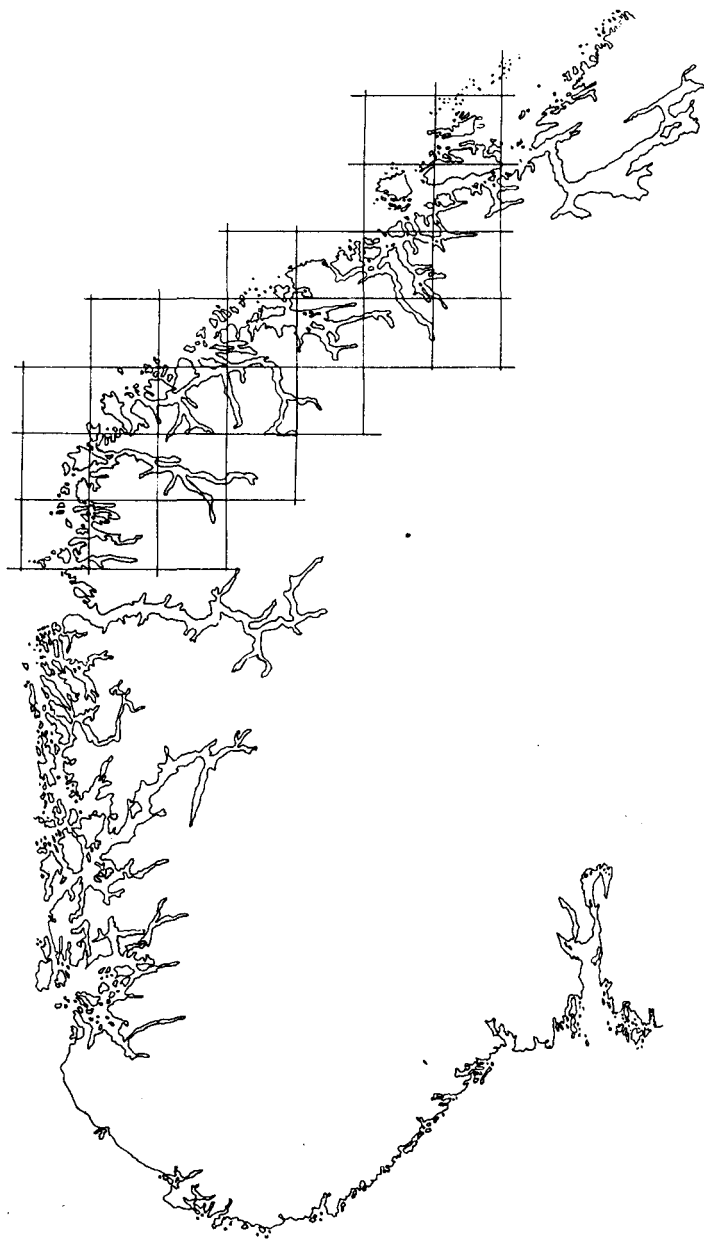
Figure 1 Determining the fractal dimension of the coast of Norway by counting how many boxes the outline of the coast penetrates [Fed 88].

size scales. In other words, lack of scales breeds self-similarity, and a self-similar object, exempting a few featureless entities such as the infinite straight line, must be fractal.

Let us look at Mother Earth again, not from outer space, but from distance at which her gross curvature becomes irrelevant. For many coastlines (as for some man-drawn boundaries) there is no natural length scale. The processes that shape many an interface between water and land are similar over a wide range of scales. Figure 1 shows a piece of the coast of Norway, the homeland of Jens Feder. from whose book *Fractals* [Fed 88] this illustration is drawn. There are large fjords and smaller fjords and ever littler inlets. And if we consult maps showing more and more detail and finally walk (or row) along the beach, we see that the little inlets harbor still littler inlets and so forth, down to the level of the seawater penetrating the spaces between individual pebbles.

How do we measure the length of such a fractal coast? Obviously, the more detail we "consult," the longer the apparent length L. In fact, if the coast is self-similar, we shall find a self-similar *power law* connecting L and the scale unit r employed in its measurement:

$$L(r) - r^\varepsilon \tag{1}$$

where the exponent ε is negative if L increases as r decreases. By contrast, for a smooth curve, the measured length approaches an asymptotic value as $r \to 0$— that is, the exponent ε is zero.

As we have seen in earlier chapters, fractal objects are characterized by a fractal dimension, such as the Hausdorff dimension

$$D = \lim_{r \to 0} \frac{\log N(r)}{\log (1/r)} \tag{2}$$

where $N(r)$ is the minimum number of disk of diameter r needed to cover the fractal. If we measure L in this manner, $N(r)$ equals $L(r)/r$ and the exponent ε in equation 1 is seen to equal $1 - D$. Thus, the Hausdorff dimension is given by

$$D = 1 - \varepsilon \tag{3}$$

which, for $\varepsilon < 0$, will exceed 1 (the value for a smooth curve).

Dimensions from Box Counting

The Hausdorff prescription of covering the fractal with disk is not always the most convenient way to measure a fractal dimension, nor is the Minkowski sausage recipe that we divulged in pages 41–43 in Chapter 1. Rather, for
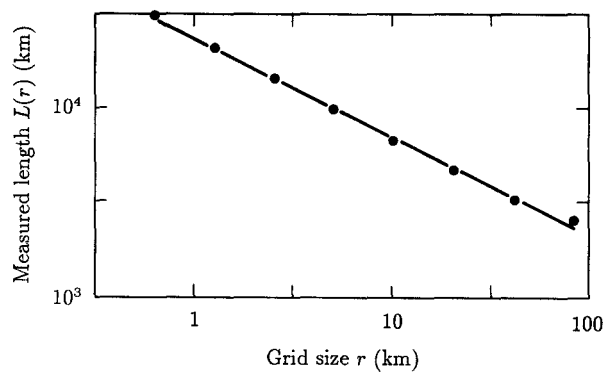
Figure 2 Measured length of Norwegian coast against grid size r. The slope of the straight line gives the fractal dimension of the coast $D \approx 1.52$ [Fed 88].
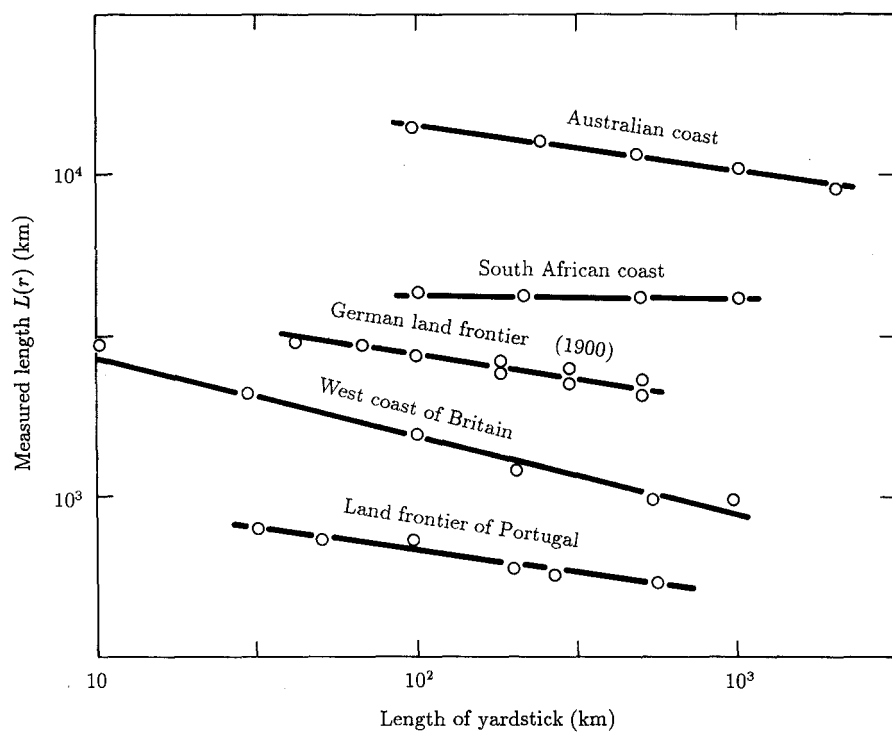

Figure 3 Apparent lengths and fractal dimension of different coasts and land frontiers [Man 83].

many practical fractals, *box counting* is the method of choice. In box counting, we superimpose a square grid over the fractal (see Figure 1) and count the number of boxes $N(r)$ that are penetrated by the fractal. Again, as r, the spacing between grid lines, becomes very small, we find that $\log N(r)/\log (1/r)$ converges to a finite value, the Hausdorff dimension D.

Figure 2 shows the result of Feder's counting the coast of Norway with grid sizes ranging from $r = 0.6$ km to 80 km. The measured length $L(r)$ decreases with increasing r by a factor of 12—from about 30,000 km to 2500 km. On a double logarithmic plot, all measurement points fall near a straight line with a slope of $\varepsilon = -0.52$. Thus, the Hausdorff dimension, according to equation 3, is $D = 1 - \varepsilon = 1.52$—halfway between the Euclidean dimension of a smooth curve and that of a smooth surface.

Figure 3 shows data publicized by Mandelbrot [Man 83] for the apparent lengths $L(r)$ of several other coasts and land frontiers, whose Hausdorff dimensions range from a smooth $D \approx 1$ for the coast of South Africa to a ragged $D \approx 1.3$ for the west coast of Britain. But no coast or country can match Norway's $D \approx 1.52$.

The Mass Dimension

For many purposes, the box-counting dimension is still not the most appropriate or convenient fractal measure. Look at a "Lichtenberg figure" (see Figure 4), one of the first physical fractals made by man. It is the electrical discharge pattern from a metallic tip placed on an insulator, first made visible in 1777 (the year of Gauss's birth in nearby Brunswick) by the Göttingen physicist and aphorist Georg Christoph Lichtenberg (1742–1799).[1] If we measure the bright area M of Lichtenberg's figure, we find that it increases with the characteristic radius R according to a simple, homogeneous power law:

$$M \sim R^D \qquad (4)$$

But the exponent D does not equal 2, as for a homogeneous figure in the plane (e.g., a circular disk, whose area M equals πR^2). Rather, the exponent D for the Lichtenberg figure lies between 1.7 and 1.9.

The exponent D in equation 4 can serve as another fractal dimension, more conveniently measured and easier to grasp than the dimensions introduced so

---

1.   Lichtenberg traveled twice to England, once as a guest of George III. There he encountered and came to admire British science and urbanity. His fame rests on the plethora of penetrating aphorisms with which he needled fellow scientists and citizens. Lichtenberg's major literary oeuvres are his illuminating comments on William Hogarth's engraving (*including Marriage à la Mode and, The Rake's Progress*).
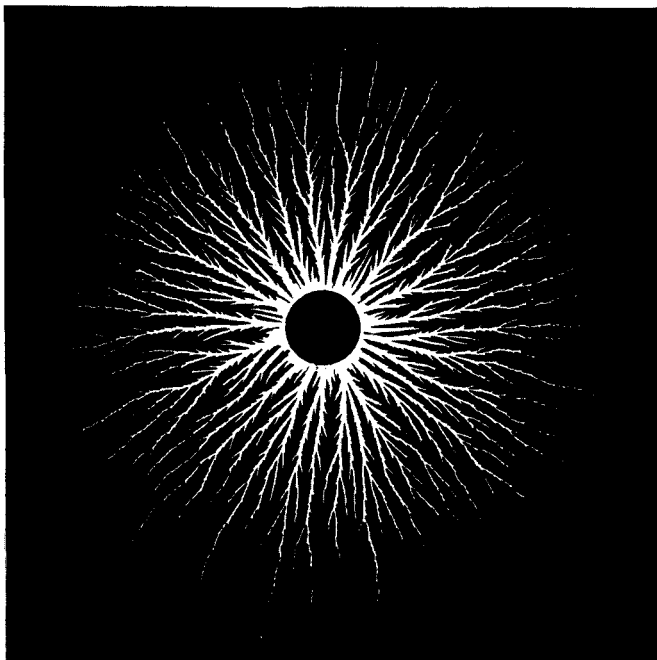
Figure 4   Another Lichtenberg figure: an electrical discharge pattern on the surface of a glass plate. Mass dimension $D \approx 1.9$ [NPW 86].

far. It applies to many fractals, from the man-made Cantor dust (see Figure 5) to the natural soft down (see Figure 6) that still fills a few (all too few!) pillows. In each case, the "mass" inside a radius R does not increase with the Euclidean dimension as an exponent but with some lesser power, such as about 1.6 for the down (depending on its price—smaller D costs *more!*).

In the Sierpinski gasket, too (see pages 17–19 in Chapter 1), D is smaller than 2 because the area M enclosed by a circle of radius R increases by a factor of 3 (not 4) every time the radius is doubled. Thus $M \sim R^D$ with $D = \log 3/\log 2 = 1.58 \ldots$.

For strictly self-similar mathematical fractals, such as the Sierpinski gasket and the Cantor dust, the mass dimension is the same as the Hausdorff dimension (and any other fractal dimension considered here). They are all given by the similarity dimension of the scaling law, defined by the initiator and the generator that generates the fractal. But for practical fractals, there are significant differences, as we saw when we introduced the Minkowski dimension in pages 41–43 in Chapter 1 to account for the number of vibrating modes of a drum with a fractal perimeter.
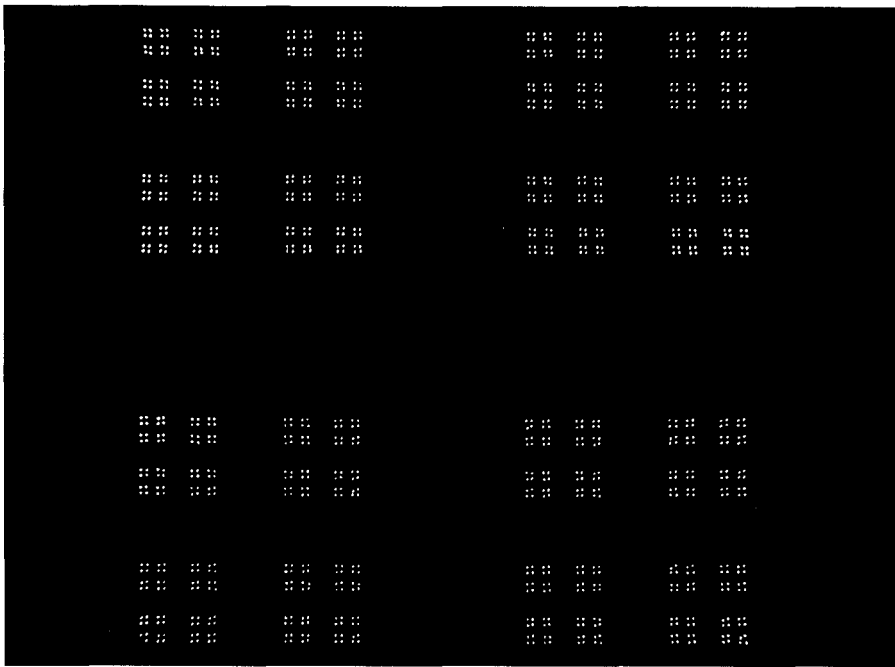
Figure 5 Cantor dust from the output of a digital-to-analog converter. Mass dimension $D \approx 1.26$.

The mass dimension is particularly appropriate to parameterize the packing of powders. Primary powder particles are apt to form clusters with a packing density of, say, p. Assume that these clusters have radii that are r times larger than those of the primary particles. Now very often these clusters will again cluster with the same or similar values of p and r, and so forth, for several generations; see Figure 7 [OT 86].

After n such generations, the density of the powder P equals p^n and the cluster radius R is equal to r^n. The total mass M is of course proportional to PR^d, where d is the Euclidean dimension in which the powder resides ($d = 3$ for most powders in a three-dimensional world). Thus,

$$M \sim PR^d = R^D$$

with the mass dimension $D = d + \log p/\log r$, which, since $p < 1$ and $r > 1$, is smaller than d. In Figure 7, d equals 2, r is about 7, and p is roughly 0.7, giving a mass dimension $D \approx 1.82$. For the Sierpinski gasket with $d = 2$, $r = 2$, and $p = \frac{3}{4}$, for comparison, we get a smaller mass dimension, namely, $2 + \log \frac{3}{4}/\log 2 \approx 1.58$, the same value as its Hausdorff dimension.
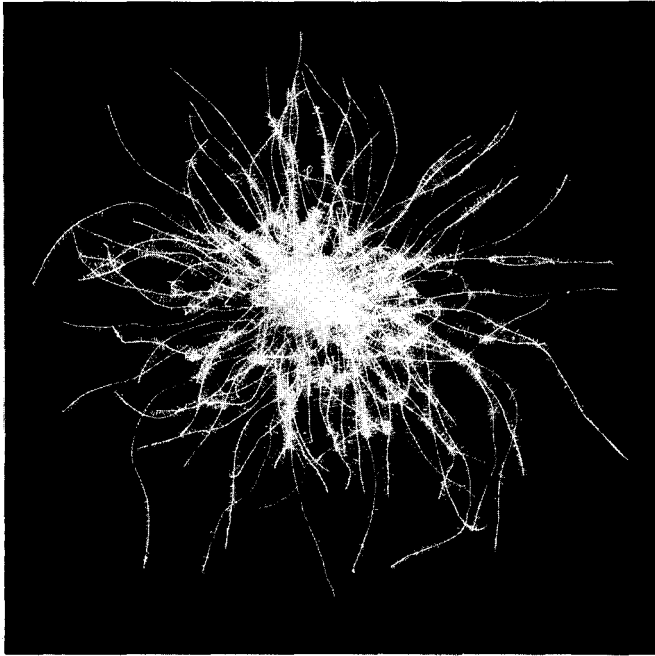
Figure 6   Natural down. Its softness results from its low mass dimension $D \approx 1.6$.
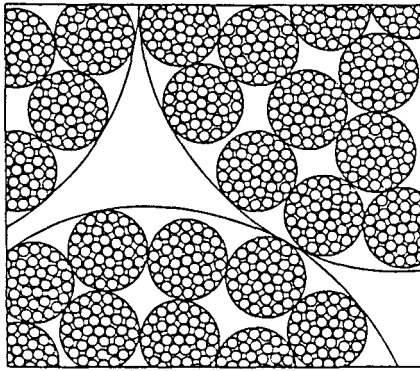

Figure 7   Three generations of a self-similar agglomerate in a powder [OT 86].

The mass dimension, as the name suggest, can be applied in higher Euclidean dimension, too, especially to spongy substances, such as the Menger sponge shown in Figure 3 of Chapter 8, with $D = \log 20/\log 3 \approx 2.73$, and crystals grown by *diffusion-limited aggregation* (DLA), for which $D = 2.4$ is a typical value [Mea 87]. In DLA, (see Figure 9 in Chapter 9 for the result of two-dimensional computer simulation), one "atom" at a time is released at a large distance from the growing aggregate and allowed to diffuse until it attaches itself to the aggregate once it comes under the short-range attractive forces of the atoms already captured. Simple probability tells us that new atoms will attach themselves preferentially near the tips of the outreaching "dendrites" rather than deep inside the crystal's "fjords."

Measurement of the mass dimension gives values of $D \approx 2.4$ for three-dimensional DLA and $D \approx 1.7$ in two-dimensional DLA. However, the exact values of D do depend on the physical and chemical parameters of the process and contain important clues for the manufacture of new materials and for practical applications in which fractal processes dominate.

Take viscous fingering (see Figure 10 in Chapter 9), produced by the surface instability as one liquid or gas invades the "territory" of another, more viscous liquid. (The reader is encouraged to produce his own viscous fingers with the help of water and glycerol, squeezed between two glass plates.)

Viscous fingering has also been observed as one liquid (say, water) replaces another (oil) in a porous medium (shale), a standard method of squeezing oil form bituminous rock (see Figure 8). The fractal mass dimension of the watery fingers depends sensitively on the liquid's viscosity, the porosity of the rock,
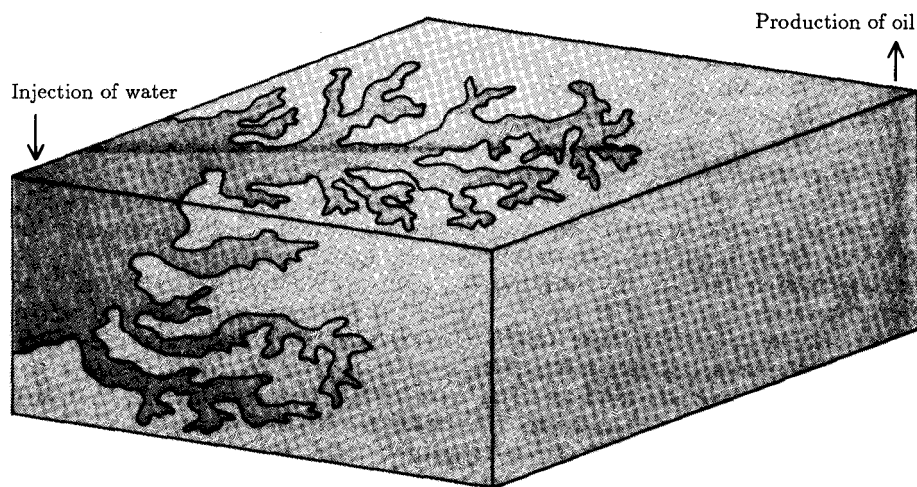

Figure 8   Squeezing oil from bituminous rock.

and wetting properties of the liquid and the rock surface [SWGDRCL 86]. If the water's edge were smooth ($D = 2$), most of the oil could be extracted from a distant well before the water reached the extracting pump. But, unfortunately for an oil-hungry world, D is much larger than 2, and the first water reaches the pump long before all the oil has been pumped out. However, the addition to the water of viscosity-increasing polymers reduces the fractal dimension and the oil industry hopes thereby to double the amount of oil that can be recovered by injecting water into shale. (The additives, the industry also hopes, will cost less than the value of the extra crude oil extracted.)

The Correlation Dimension

One of the most widely used fractal dimensions is the *correlation dimension*, because it is experimentally the most accessible, especially if the fractal comes as a "dust"—isolated points, thinly sprinkled over some region of space.

To determine the correlation dimension, one first counts how many points have a smaller (Euclidean) distance than some given distance r. As r varies, so does the relative count $C(r)$, defined as the total count divided by the squared number of points. The quantity $C(r)$ is also called the *correlation sum* (or *correlation integral*) [GP 83, Gra 83].

The correlation dimension is then defined by

$$D_2 := \lim_{r \to 0} \frac{\log C(r)}{\log r} \tag{5}$$

Figure 9 shows the experimental determination of D_2 for the strange attractor of the iterated *Hénon map*, which yields a straight-line dependence of $\log C(r)$ on $\log r$ over six orders of magnitude, with a slope $D_2 = 1.21$ [GP 83].

Infinitely Many Dimensions

The correlation dimension D_2 belongs to an infinite family of dimensions D_q defined by

$$D_q := \lim_{r \to 0} \frac{1}{q - 1} \frac{\log \sum_k p_k^q}{\log r} \qquad -\infty \leq q \leq \infty \tag{6}$$

where the sum is over all the cells of linear size r into which the space has been subdivided and p_k is the relative frequency or probability with which points of the dusty fractal fall inside cell k [HP 83a, Gra 83] (see also pages 202–205 in Chapter 9).
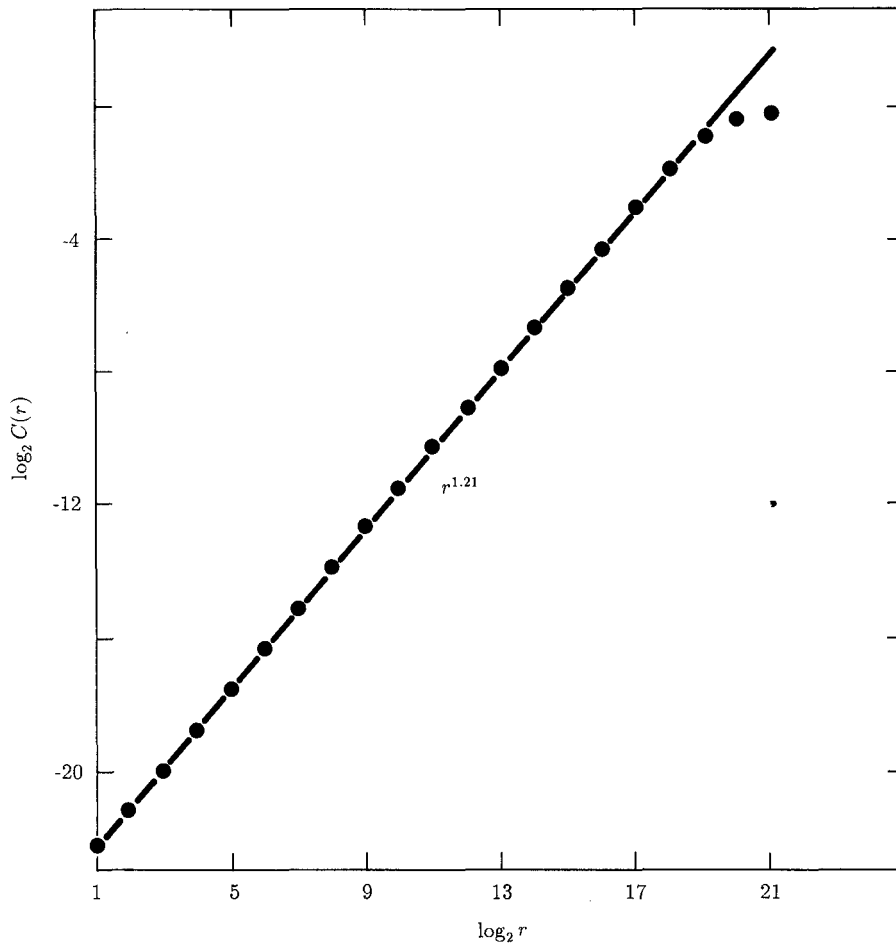
Figure 9   Determining the correlation dimension of a strange attractor [GP 83].

For $q = 0$, we recover our old friend, the box-counting or Hausdorff dimension D_H, because the sum $\sum_k p_k^0 = \sum_k 1$ simply counts how many boxes or cells are "invaded" by the fractal. Thus, $D_H = D_0$.

For $q = 2$, it is easy to show that, in the limit as $r \to 0$, the sum $\sum_k p_k^2$ equals the relative count $C(r)$, which yields the correlation dimension introduced in the preceding section [Schu 84].

For $q \to 1$, equation 6 yields the so-called *information dimension*

$$D_1 = \lim_{r \to 0} \frac{-\sum_k p_k \log p_k}{\log (1/r)} \tag{7}$$

—so called because the numerator in equation 7 is Shannon's entropy, as introduced in his "information theory" (following Boltzmann's invention of entropy in statistical mechanics). The dimension D_1 does in fact measure the loss of information in the dynamic development of chaotic systems.

For $q \to \infty$, only the highest probability p_{max}, in the sum in equation 6, counts. Hence

$$D_\infty = \lim_{r \to 0} \frac{\log p_{max}}{\log r}$$

Conversely, for $q \to -\infty$, the smallest probability p_{min}, controls the sum. Thus,

$$D_{-\infty} = \lim_{r \to 0} \frac{\log p_{min}}{\log r}$$

$D_{-\infty}$, depending, as it does, on the smallest probability, is difficult to measure for practical fractals. The sites with low probability are visited too infrequently.

Note that $D_{-\infty} \geq D_\infty$. In general, for any two dimensions with different q,

$$D_q \geq D_{q'}, \qquad \text{for} \qquad q < q' \tag{8}$$

Thus, D_q is a monotone nonincreasing function of q. Only in exceptional cases does D_q not depend on q at all and does it have the same value in the entire range $-\infty \leq q \leq \infty$.

One such exception is a strictly self-similar fractal generated from "non-erased" segments of equal length, such as those that generate the triadic Cantor dust. For the calculation of D_q according to equation 6, we consider the generator to consist of $N = 2$ segments with equal probabilities $p_1 = p_2 = \frac{1}{2}$. Since the triadic Cantor set is a strictly self-similar fractal, the limit as $r \to 0$ is superfluous; we can calculate D_q with the value $r = \frac{1}{3}$ for the generator. This gives

$$D_q = \frac{1}{q-1} \frac{\log [2(1/2)^q]}{\log (1/3)} = \frac{\log 2}{\log 3}$$

which is a value that is indeed independent of q.

If the generator consists of segments of *different* lengths r_k, then, for $D_q = D$ to hold for all q, the probabilities p_k must be proportional to $r_i^{D_q}$. To show this, we use equation 14 from Chapter 9, the formula for self-similar fractals:

$$\sum_i p_i^q r_i^\tau = 1 \tag{9}$$

where $\tau = (1 - q)D_q$.

With $p_i \sim r_i^{D_q}$ in equation 9, we get

$$\sum_i r_i^q = 1$$

which means, of course, that D_q is independent of q.

For example, for $r_1 = \frac{1}{2}$ and $r_2 = \frac{1}{4}$, $D = \log \gamma / \log (\frac{1}{2}) \approx 0.694$, where $\gamma = (\sqrt{5} - 1)/2$ is the golden mean. Thus, $p_1 = (\frac{1}{2})^D = \gamma$ and $p_2 = (\frac{1}{4})^D = \gamma^2$. Probabilities $p_i = r_i^D$, except for $r_i = r$, are, of course, rather artificial, and D_q in general *does* depend on q. For example, for $r_2 = r_1^2$ and $p_1 = p_2 = \frac{1}{2}$, equation 9 gives

$$D_q = \frac{1}{1 - q} \frac{\log [\sqrt{1 + 2^{q+2}} - 1/2]}{\log r_1} \qquad q \neq 1$$

which ranges from $D_{-\infty} = \log 2 / \log (1/r_1)$ down to $D_{\infty} = \log 2 / \log (1/r_2)$.

Note that, for $r_1 > \frac{1}{2}$, $D_{-\infty}$, which measures the densest part of the fractal, exceeds 1. In general, for constant p_k, $D_{-\infty}$ is determined by the longest segment, r_{max}, and D_{∞} by the shortest segment, r_{min} (see Chapter 9).

The Determination of Fractal Dimensions
from Time Series

A large quantity of the data that humans take in comes in the form of a "time series," that is, a temporal sequence of measured values, such as electroencephalographic (EEG) potentials (see Figure 10). Is this just random noise, or is it deterministic chaos, generated by some underlying deterministic, albeit chaotic, process?

In order to decide this often crucial question, one constructs d-dimensional data vectors from d measurements spaced equidistant in time and determines the correlation dimension D_2 of this d-dimensional point set. If the data were truly
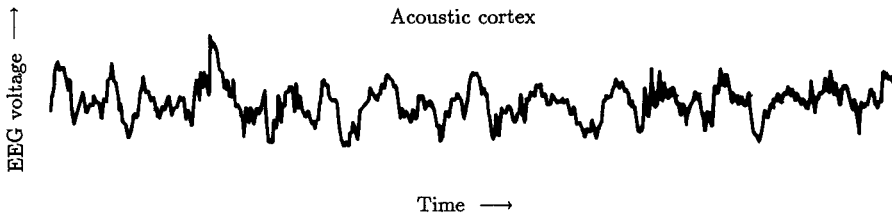

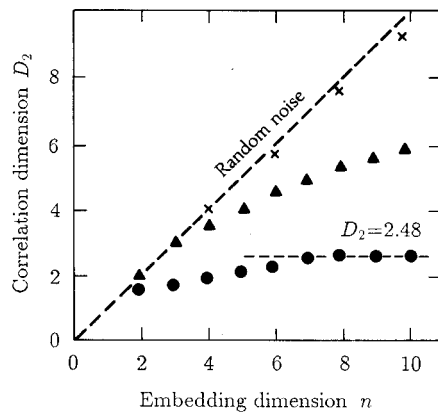**Figure 10**   An electroencephalogram [Rös 86].

Figure 11   Correlation dimension as a function of embedding dimension [LH 86].

random, then, as d is increased, the calculated D_2 would go up accordingly. But for a deterministic system, no matter how chaotic it appears to the "naked eye," the calculated correlation dimension will not increase further, once the so-called *embedding dimension* d exceeds the correlation dimension D_2 of the data (see Figure 11). In this manner the author's student J. Röschke found that the seemingly noiselike EEGs recorded from the acoustic cortex of cats are in fact not noise at all, but deterministic chaos whose correlation dimension depends on the state of the cat's wakefulness [Rös 86].

   This "embedding" method of determining the fractal dimension of experimental data, and thereby distinguishing deterministic chaos from random noise, has been applied successfully to a wide variety of physical, meteorological, biological, and physiological observation [HP 83b].

Abstract Concrete

Self-similarity is not only amenable to *measurement*; self-similarity can also be employed profitably in the *design* of fractal structures and materials with increased durability or lower cost (or both). A case in point is the construction of field walls on many New England farms; see Figure 12, a snapshot taken by the author during a biking tour through Connecticut. There are large stones whose gaps are filled by smaller stones, whose interstices, in turn, are filled by still smaller stones. As a result of this roughly self-similar composition, the wall keeps standing upright without the customary edifying intervention of expensive cement to fill and fixate the cracks. If the number of stones as a function of their size is selected according to a power law, what would be a good exponent?

Figure 12   The hierarchical construction of a New England field wall.

In a column in *Nature* entitled "Abstract Concrete," David Jones argues that by employing ever finer particles, from the coarsest gravel to the finest dust, the volume to be filled by expensive binder can be made arbitrarily small, thereby cutting cost—or allowing more expensive, high-tenacity binding materials like epoxy or even polyimide to be used [Jon 88]. Likewise, many other composite materials, such as fiberglass, could probably be improved by a self-similar composition.

Fractal Interfaces Enforce Fractional Frequency Exponents

Finite electrical circuits constructed of passive "lumped elements," like resistors and capacitors, have input impedances that are rational functions of frequency. For example, a capacitor, with capacitance C, has an impedance $Z = (i\omega C)^{-1}$, where $i = \sqrt{(-1)}$ and ω is the radian frequency (2π times the frequency). Thus, for a capacitor, $Z \sim \omega^{-1}$, which is a rational function of ω.

Electrical engineers found out a long time ago that a finite electrical circuit, constructed in any way from a finite number of lumped elements, always has an

impedance that is a rational function of frequency. This is a consequence of the fact that rational functions form a group under composition—that is, if $R(x)$ and $S(x)$ are rational functions, so is $S(R(x))$.

This rationality is unfortunate in a way, because the *characteristic impedance* Z_0 of a transmission line (such as a TV cable), which is needed for a "matching" echo-free connection, is *not* a rational function of frequency. Instead, Z_0 typically involves square roots like $\omega^{-1/2}$. Thus, no finite network can match such a characteristic impedance exactly; matching networks are always approximations (and the standards of approximation, alas, vary from country to country).

However, *infinite* networks can produce an irrational frequency dependence. For example, the infinite ladder network shown in Figure 13 has an input impedance Z_0 that is best written as a continued fraction:

$$Z_0 = R + \frac{1}{G +} \frac{1}{R +} \frac{1}{G +} \cdots$$

A closed form for the value of Z_0 can be obtained by exploiting the periodicity of this continued fraction, which permits us to write

$$Z_0 = R + \frac{1}{G + (1/Z_0)}$$

This is a quadratic equation for Z_0. The physically meaningful solution (the one with positive real part) is

$$Z_0 = \frac{1}{2}(R + \sqrt{(R^2 + 4R/G)}) \tag{10}$$

For $R = 1/G = 1$ ohm (Ω) and a finite number of elements, Z_0 equals a rational number, namely, the ratio of two successive Fibonacci numbers: 1/1, 2/1, 3/2, 5/3, ... But for an infinite number of elements, as equation 10 shows, Z_0 is no longer rational. In fact, it equals the reciprocal golden mean $(\sqrt{5} + 1)/2 = 1.618\ldots$, an irrational number.
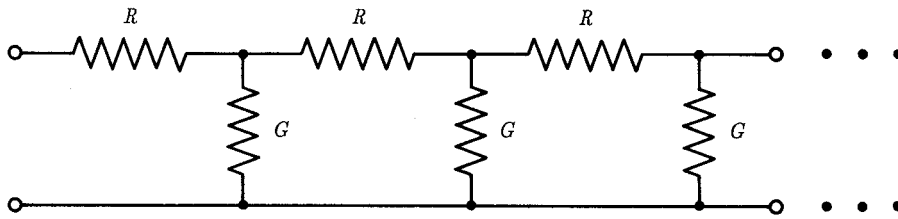

Figure 13   Electrical ladder network composed of resistances R and conductances G.

In order to realize an irrational *frequency* dependence, we can make G a capacitor, with reciprocal impedance $G = i\omega C$. For small frequencies, $\omega \ll 1/RC$, Z_0 then depends on ω in simple power-law manner $\omega^{-\beta}$, with the fractional exponent $\beta = \frac{1}{2}$:

$$Z_0 = R\left(i\frac{\omega}{\omega_c}\right)^{-1/2} \tag{11}$$

where $\omega_c = 1/RC$ is the upper "cutoff frequency" of this approximation.

However, the real world of electrical conduction goes beyond simple half-integer exponents like $\beta = \frac{1}{2}$. This is noteworthy because *uniform* networks are described by *periodic* continued fractions, which lead invariably to square roots and half-integer exponents. Hence, the observation in a physical system of a nonstandard frequency dependence with a fractional value of 2β implies some kind of *nonuniform* structure.

In fact, such nonstandard frequency behavior is frequently observed in electrical conduction across *rough surfaces*, such as between an electrode and the electrolyte of a car battery. Indeed, *simple models of the current-carrying interface as fractals* have established a unique relationship between the fractal geometry of the interface and the frequency exponent η.

Figure 14A shows a highly schematized, two-dimensional model of an interface between an electrode (white) in contact with an electrolyte (black) [Liu 85]. The model is based on a Cantor set whose generator consists of two segments of equal length r. In Figure 14A, r equals 0.3. To model the roughness of the interface, the "grooves" in the electrode have increasing depth, becoming deeper by a constant amount with each generation of constructing the fractal, as shown in Figure 14A.

Figure 14B, shows the treelike electrical circuit representing current conduction through this interface. Note that the resistances increase by a factor $1/r > 2$ with each generation, reflecting the fact that the grooves filled by the electrolyte become narrower and narrower. This network is in fact a *self-similar* Cayley tree, that is, a tree with a constant branching ratio (here equal to 2) and a scale factor for the resistances equal to $1/r$, representing the resistance increases in the progressively narrower grooves.

The input impedance of this tree is given by the continued fraction

$$Z(\omega) = R + \cfrac{1}{i\omega C +} \cfrac{2}{R/r +} \cfrac{1}{i\omega C +} \cfrac{2}{R/r^2 +} \cdots \tag{12}$$

Here the 2s in the numerators stem from the 2:1 branching ratio of the tree.

This continued fraction, too, would be periodic if it were not for the fact that the scaling factor r is different from $\frac{1}{2}$. It can be written in the following closed form:

$$Z(\omega) = R + \cfrac{1}{i\omega C +} \cfrac{2r}{Z(\omega/r)} \tag{13}$$

(A)
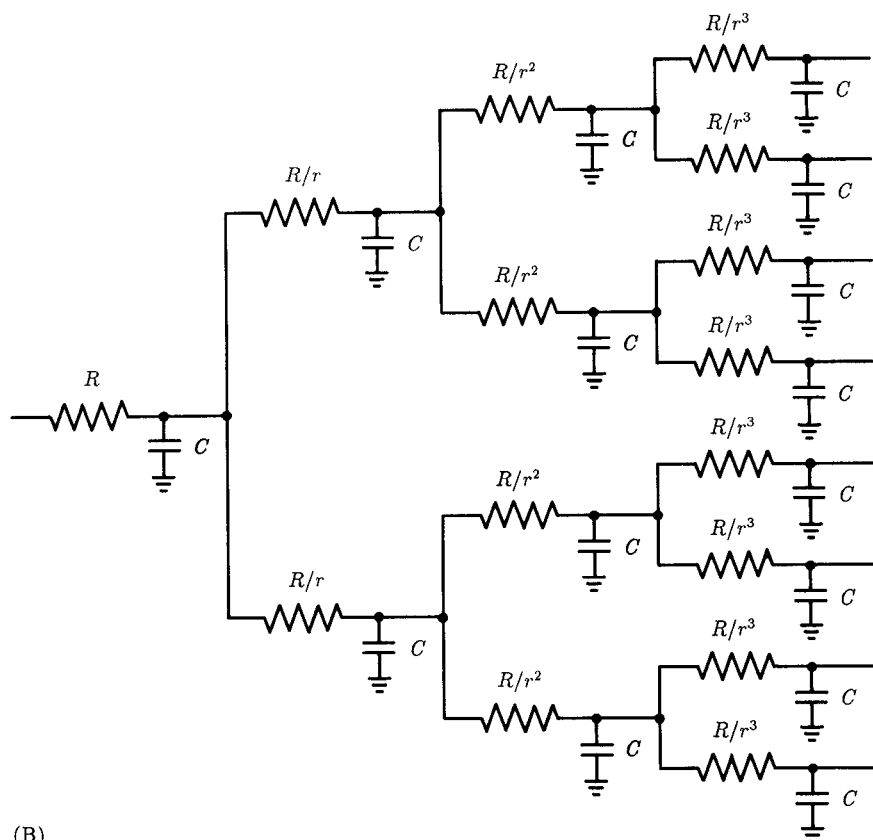
(B)

Figure 14 (A) Middle-third-erasing model of interface between electrode (white) and electrolyte (black) in a car battery. (B) Treelike electrical circuit representing current flow through the fractal interface shown in part A [Liu 85].

Note that $Z(\infty) = R$, as one would expect from the circuit diagram in Figure 14B. For finite frequencies, $|Z(\omega)| > R$.

For

$$R \ll \frac{|Z(\omega)|}{2r} \ll \frac{1}{\omega C}$$

which implies $\omega \ll 1/RC$, equation 13 simplifies to the scaling law

$$Z(\omega) = \frac{1}{2r} Z\left(\frac{\omega}{r}\right) \tag{14}$$

which is solved by the power law

$$Z(\omega) \sim \omega^{-\beta}$$

with the exponent given by

$$\beta = 1 - \frac{\log 2}{\log (1/r)} = 1 - D \tag{15}$$

Here D is the Hausdorff dimension of the Cantor set used in the model (Figure 14A). Because $0 \leq D \leq 1$, the exponent β too lies between 0 and 1; it is not necessarily restricted to $\beta = \frac{1}{2}$.

The Hausdorff dimension of the one-dimensional electrode-electrolyte interface in two dimensions equals $D + 1 = 2 - \beta$. For a two-dimensional interface with isotropic roughness, the relations for β and D are unchanged. The corresponding tree model would employ a branching ratio of 4 instead of 2, and a resistance scale factor equal to $1/r^2$ instead of $1/r$. However, the Hausdorff dimension of the interface becomes $2D + 1 = 3 - 2\beta$. These, then, are the anticipated relations between conductor geometry and frequency exponent. Note that the rougher the interface (large D), the smaller the frequency exponent β.

Intuitively, the reason for the increase in impedance as the frequency is lowered is that the current reaches deeper and deeper into narrower and narrower crevices of the interface before being shunted away by the capacitances. For the treelike fractal interface, the penetration depths scale with an exponent β which can assume a range of values depending on the roughness of the interface.

If it were not for the exponential growth in the number of needed components, treelike networks, like the one shown in Figure 14B, would also be useful for generating noises or filtering signals with a fractional-exponent frequency dependence. However, the $1/f$ noises observed in many electronic materials may in some cases be generated by a fractal composition that is amenable to the modeling discussed here.

The Fractal Dimensions of Fracture Surfaces

God made the bulk;
surfaces were invented
by the devil.
—WOLFGANG PAULI

Of course, fractal surfaces are not limited to rough electrodes. *Fracture* is another omnipresent source of two-dimensional fractals. In a classic study, Mandelbrot, Passoja, and Paullay investigated the structure of fractured samples of a low-carbon steel which they found to be self-affine [MPP 84]. By plating the fracture surface with a thick layer of nickel and subsequent planing, they could create little islands of steel that grew in size as the planing parallel to the surface progressed. Figure 15 shows the areas A of these steel islands versus their perimeters L in a double logarithmic plot. The data are well fitted over four orders of magnitude by the power law $A \sim L^{1.56}$, implying that the perimeters of the steel islands are self-similar. The value of the exponent, 1.56, is quite distinct from the Euclidean law $A \sim L^2$. The exponent 1.56 also means that the Hausdorff dimension D of the perimeter equals $2/1.56 = 1.28$. This value is close to that of the coast of Britain ($D \approx 1.3$), but considerably greater than that of the "fractal hexagon" ($D = \log 9/\log 7 \approx 1.13$) that we encountered in Chapter 1, pages 13–15. For the fractal hexagon, too, the reader may recall, the perimeter had to be raised to the power $2/D \approx 1.77$ (not 2) to properly predict the "hexagon's" area from its perimeter.

The value $D = 1.28$ for the steel islands, incidentally, implies that the self-affine fracture surface itself has a fractal dimension of $D + 1 = 2.28$, a value typical for rough mountains. A vertical cut through such a fracture mountainscape would produce a profile with a fractal dimension $2.28 - 1 = D$. Mandelbrot and his collaborators confirmed this result by measuring the spatial frequency spectrum $P(f)$ of such fracture profiles. They found $P(f) \sim f^{-4.5}$. From the frequency ex-
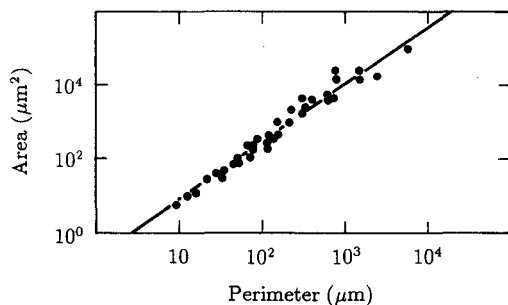

Figure 15 Areas of steel "islands" plotted against their perimeters in fractured low-carbon steel [MPP 84].

ponent $\beta = 4.5$, the fractal dimension is obtained by the formula $D = 3 - (\beta - 1)/2 = 1.25$ (see page 137 in Chapter 5), in good agreement with their independent "island" measurements. Thus, metal fracture is free from an inherent scale over several orders of magnitude.

Interestingly, the fractal dimension and the impact energy required for fracture were found to be related to the temperature used in tempering the steel. The precise metallurgical basis of this dependence of fracture energy and topography on heat treatment remains to be rendered intelligible, though.

The Fractal Shapes of Clouds and Rain Areas

Heavy rains interfere with microwave transmission—the medium of choice, for the last half-century, of long-distance telephone transmission. A patch of dense rain between two microwave towers, ubiquitous in the United States and other countries, necessitates the rerouting of communications links. Small wonder, then, that telephone engineers have shown great interest in the temporal distributions of precipitation and the geometric shapes of rain areas. Rain comes, of course, from clouds (although there seem to be people, quite erudite by the way, who confessed surprise when first told so).

Thus, nothing seems more natural than to study the statistics of rain and clouds together, as Lovejoy among others has done [Lov 82]. Figure 16 shows
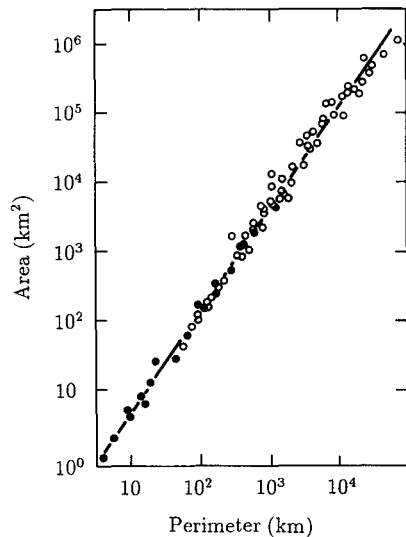

Figure 16   Areas of rain patches (filled circles) and clouds (open circles) plotted against their perimeters [Lov 82].

the area A of rain patches (filled dots) and clouds (open circles) versus the perimeter L. On a double logarithmic plot, like Figure 16, the data hew close to a straight line over a combined range of six orders of magnitude, from 1 km^2 to 1 million km^2! The slope is just under 1.5, implying a fractal dimension D of the perimeters equal to $2/1.5 \approx 1.35$. Again, as in fracture and so many other natural phenomena, there seems to be no natural length scale.

The numerical value of the exponent is in good agreement with a thermodynamic model by Hentschel and Procaccia [HP 84]. Lovejoy and Mandelbrot proposed a mathematical model in which rain areas are assumed to be generated by the superposition of individual rain patches with hyperbolic size distribution [LM 85]. Their model is capable of generating eminently realistic images of clouds. Color Plate 5 shows a photograph, taken by the author, of a natural cloud that rivals the realism of computer-generated nebulosities. Color Plate 6A and B shows two other fractals observed in nature—one generated by natural decay, the other by natural growth.

Cluster Agglomeration

In diffusion-limited aggregation (DLA), aggregates of molecules grow by adding one molecule at a time. Another important growth process that leads to fractal structures is the agglomeration of aerosols and colloids. Figure 17 shows an electron microscope image of a gold colloid grown by cluster aggregation by Weitz and Oliveria [WO 84]. The fractal structure of this colloid is so sparse that, even in the two-dimensional projection shown here, the cluster is transparent.

The agglomeration of clusters is illustrated in Figure 18. Initially, individual particles are distributed roughly uniformly within a finite volume (Figure 18A). They are then allowed to migrate randomly as in Brownian motion. When two particles touch each other, they stick together and from then on move together as a little "cluster." When these little clusters run into each other, they stick together and form larger clusters as observed by Meakin [Mea 83]. In such agglomeration processes, larger and larger clusters are formed as shown in Figure 18B and C. The large clusters are in fact statistically self-similar fractals. In two-dimensional computer simulations, Hausdorff dimensions D near 1.4 are found. In three dimensions, Kolb, Botet, and Jullien found $D \approx 1.8$ [KBJ 83].

Experimentally, the fractal dimension can be determined by scattering of light, x rays, or neutrons from the fractal. For spatial frequencies (reciprocal wavelengths) f in the range $1/R \ll f \ll 1/r$, where R is the size of the entire fractal and r that of the individual particle, one expects the scattered intensity $I(f)$ to follow a simple power law: $I(f) \sim f^{-D}$ (see the next section). The fractal dimension D determined in this manner by Schaefer and his collaborators for silica particles is about 2.1; see Figure 19 [SMWC 84]. This value, being considerably higher than the computer simulation results, points to a different mechanism for the agglomeration of silica. Simulations in which clusters form only *slowly*
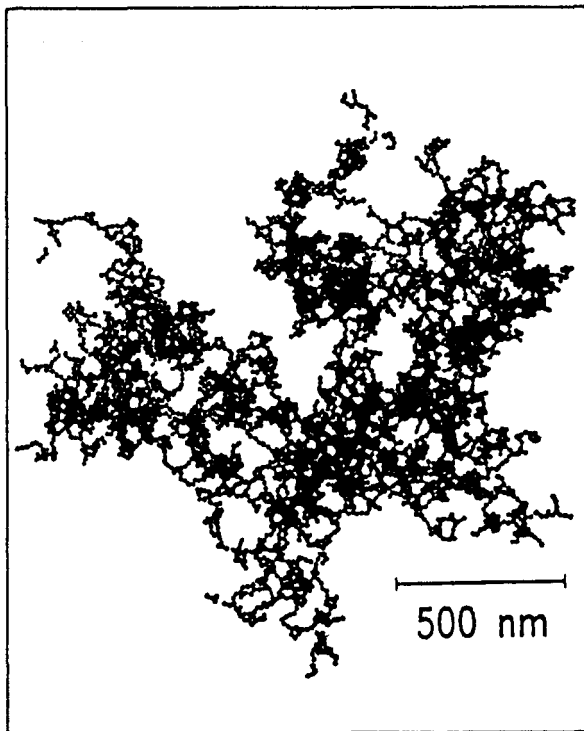
Figure 17 Gold colloid grown by cluster aggregation [WO 84].

after multiple collisions give $D \approx 2.0$, in better agreement with the results obtained by scattering [FL 84, SO 85].

Agglomeration plays a decisive role also in electrolytic deposition and catalytic reactions. Cluster formation is likewise prevalent in the spreading of epidemics, gossip, and opinions. Grassberger found that, in opinions surveys, fractal structures can lead to strongly biased and therefore false results [Gra 85].

Diffraction from Fractals

For incoherent diffraction from a fractal that consists of independent particles, such as a colloid, the scattered intensity $I(f)$ as a function of spatial frequency f is proportional to the total "mass" M contained in a volume of radius $\rho = 1/f$. With $M \sim \rho^D$, one obtains

$$I(f) \sim f^{-D} \tag{16}$$
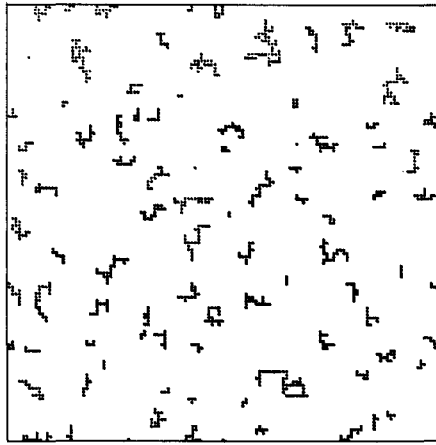
(A)
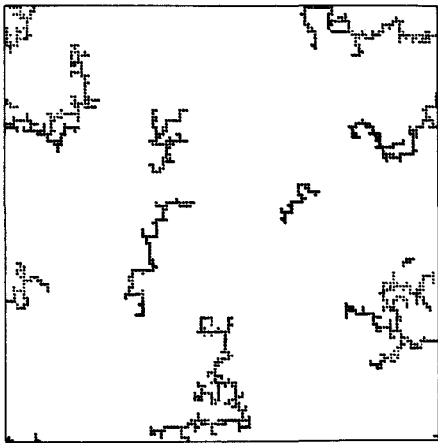
(B)
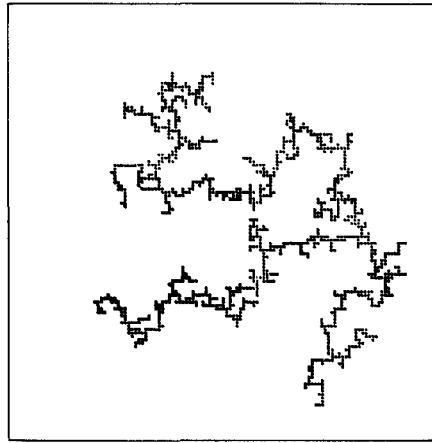
(C)

Figure 18   Agglomeration of clusters: (A) uniform random distribution; (B) formation of small clusters following random motion; (C) formation of larger clusters [Mea 83].

For *surface* fractals, $I(f) \sim S$, where S is the surface area. But $I(f)$ can also be written as $M^2F(f\rho)$, where F is some universal function [MH 87]. With $S \sim \rho^{D_s}$ and $M \sim \rho^d$, this implies $F(fR) \sim (f\rho)^{D_s - 2d}$ and

$$I(f) \sim f^{D_s - 2d} \tag{17}$$

where D_s is the fractal dimension for the scattering surface and d is the Euclidean

embedding dimension. For $d = 3$, we thus have

$$I(f) \sim f^{D_3 - 6} \tag{18}$$

For a smooth surface, D_s equals 2 and $I(f)$ is proportional to f^{-4}, a well-known classical result for the scattering regime considered here. As already noted, equation 18 permits the determination of D_s of a fractal surface by wave diffraction.

Figure 19 shows that scattered intensity of x rays, scattered from the colloidal aggregate of silica that we mentioned before, as a function of spatial frequency in reciprocal angstroms (10^{-10} m). The measurements are restricted to small scattering angles, because at large angles the x rays would resolve the *molecular* structure, as opposed to the cluster structure that is of interest here. For $f < 1/r$, where $r = 27$ Å is the radius of the nonfractal monomers that form the fractal colloid, the scattered intensities fall on a straight line with a slope of -2.1 in a double logarithmic plot, which is thus the Hausdorff dimension of the fractal colloid.

For $f > 1/r$, there is another linear regime of the scattered intensities with a slope of -4. This slope is the one that is theoretically expected for the nonfractal monomers that form the fractal. The experimental result is therefore an indication that the monomers remain intact in the aggregate.   •

Here then we have a quintessential application of wave diffraction to the analysis of a fractal structure, that of colloids. Understanding the processes that govern colloidal aggregation has been a long-standing aim in several branches
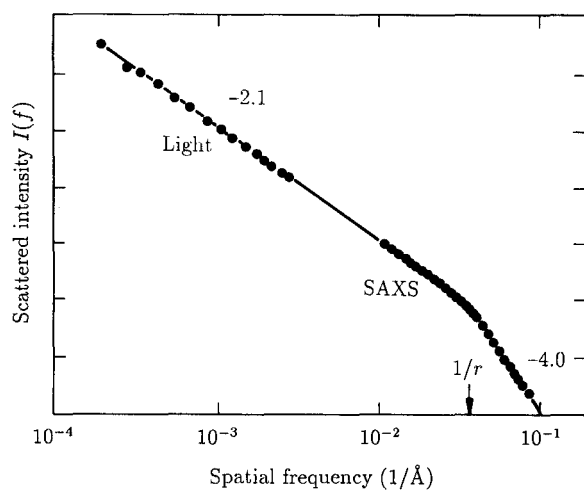

Figure 19   Determining the fractal dimension of silica particles. Intensity of x rays scattered from a colloidal aggregate of silica. SAXS = small-angle x-ray scattering [SMWC 84].

of physics and chemistry, and for a wide variety of commercial applications. Wave diffraction, always the tool of choice for structure analysis, is now being successfully extended from homogeneous bodies to fractal substance.

Another fractal structure that has been elucidated by small-angle x-ray and neutron scattering is lignite, or "brown coal." Lignite is pervaded by microscopic pores with a fractal inner surface. These pores and their surfaces are, of course, what makes "active" coal active and interesting for air filters and other purifying applications. Theory predicts that the scattering intensity $I(f)$ should be proportional to f^{-4} for smooth pores. For rough pores, with a fractal surface, $I(f)$ according to equation 18 should be proportional to f^{D-6}, where D is the fractal dimension of the inner surfaces of the pores. This law should hold for spatial frequencies corresponding to the reciprocal surface roughness of the pores.

Experimental results by Bale and Schmidt gave an exponent of -3.44 with an error smaller than 1 percent over an intensity range exceeding seven orders of magnitude [BS 84]. Thus, the pore surfaces are fractal, and their fractal dimension equals $6 - 3.44 = 2.56$.

Interestingly, the same power law, $I(f) - f^{D-6}$, would also obtain for *smooth* pores with a self-similar *size* distribution. Specifically, if the number of pores $N(r)$ with a radius larger than r is proportional to r^{-D}, then, according to Pfeifer and Avnir, the scattered intensities would follow the law $I(f) - f^{D-6}$ [PA 83]. However, the spatial frequency range would be different: it would be given by the pore *sizes*, not by their roughness. This is another example of how relatively simple things can become if scaling similarity prevails and if one properly exploits it.

Several years ago, the author proposed various number-theoretic concepts (quadratic residues, primitive polynomials and primitive roots in finite number fields, and Zech logarithms) as design principles for *reflection phase gratings* with very broad scattering of the incident energy over wide frequency bands [Schr 90]. The frequency bands efficiently scattered by such gratings can be further widened by recruiting self-similarity for their design, resulting in fractal diffraction gratings [D'An 90].

11

C H A P T E R

Iteration, Strange Mappings, and a Billion Digits for π

What good your beautiful proof on [the transcendence of] π: Why investigate such problems, given that irrational numbers do not even exist?

—LEOPOLD KRONECKER
to Ferdinand Lindemann

Apart from power laws, *iteration* is one of the prime sources of self-similiarity. Iteration here means the repeated application of some rule or operation—doing the same thing over and over again. (As the cartoon, Figure 1, illustrates, sometimes even a single repetition can lead to self-similarity.) In this chapter we continue to explore some of the strangely attractive consequences of iteration—one of our recurring themes.

A concept closely related to iteration is *recursion*. In an age of increasing automation and computation, many processes and calculations are recursive, and if a recursive algorithm is in fact repetitious, self-similarity is waiting in the wings. Think of the recursive calculation of the golden mean $(\sqrt{5} - 1)/2 = 0.618 \ldots$, obtained by the rule "add 1 and take the reciprocal." Beginning with 1, we obtain $1, \frac{1}{2}, \frac{2}{3}, \frac{3}{2}, \frac{5}{8}, \frac{8}{13}$, and so on, a sequence of fractions that converges to the golden mean. (These fractions, in fact, approach the golden mean with an error that decreases geometrically, forming an asymptotically self-similar error sequence with a similarity factor equal to the golden mean squared.) Recursion is also one of the main themes of Hofstadter's *Gödel, Escher, Bach* [Hof 80].

A very early example of an iterative algorithm is Euclid's method for determining the greatest common divider (GCD) of two natural numbers: divide the larger number by the smaller, take the reciprocal of the remainder, and *iterate*
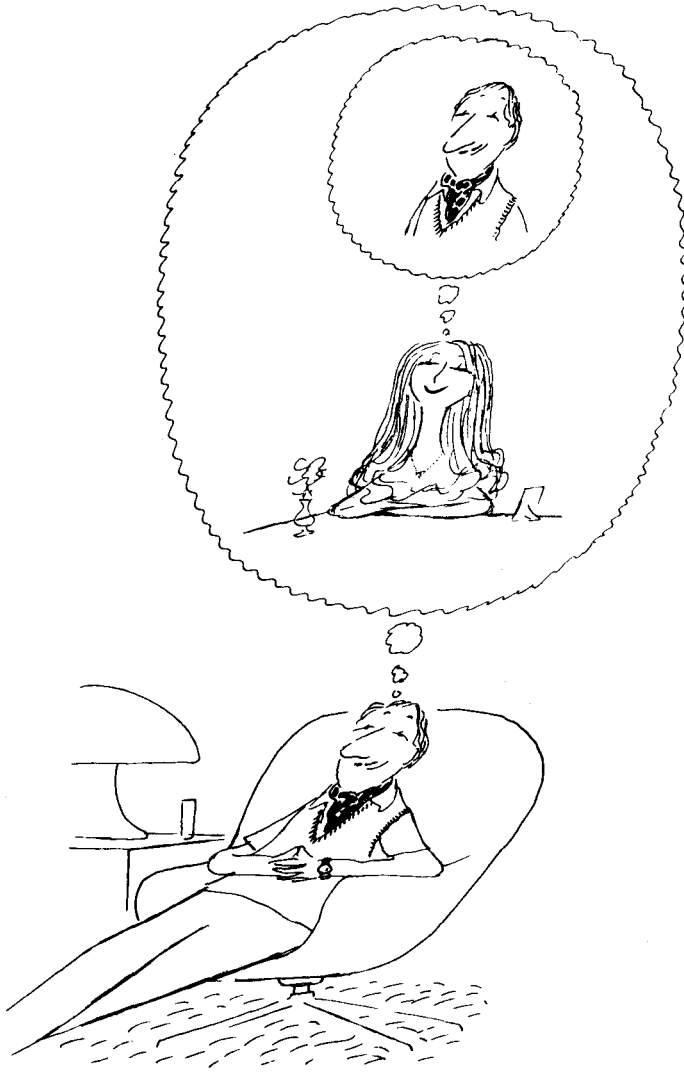
Figure 1 Iteration leads to cartoon "self-similarity." (Drawing by Sempé; © 1985 The New Yorker Magazine, Inc.)

until the remainder is zero. The denominator of the last quotient is the GCD. For example, for 182 and 78, $\frac{182}{78} = 2 + \frac{26}{78}$ and $\frac{78}{26} = 3 + 0$. Thus, 26 is revealed as the GCD of 182 and 78. (The principle behind Euclid's method is the simple fact that the integers in the remainder fraction, (26 and 78), have the same GCD as the original integers, (182 and 78.)

Another ancient problem that is solved by a recursive algorithm is the "tower of Hanoi," in which a stack of disks of different sizes has to be transferred

from one peg to another peg in such a manner that no larger disk ever comes to rest on a smaller one in the restacking process. A third peg is used as a "parking lot" for disks; otherwise the task would, of course, be impossible. For a large number of disks, trial and error would lead nowhere, but a simple recursive recipe gives the optimum solution without any guesswork.

We shall see how a simple iterative scheme, invented by Newton, for finding the zeros of a function produces basins of attraction of these zeros that are intimately intertwined self-similar multifractals. On a more mundane level, Newton's iteration is applied to calculate reciprocals and roots in high-precision arithmetic, using only multiplication. Thus $1/z$ is given by the recursion $x_{n+1} = 2x_n - zx_n^2$, and $z^{-1/2}$, for example, is calculated by iterating $x_{n+1} = x_n(3 - zx_n^2)/2$. (Initial values x_0 have to be chosen aptly, of course, to make the recursion converge to the desired result.)

The efficacy of iteration is also at the basis of the efficient algorithms for the fast Fourier transform (the famous FFT) and the less well-known *Hadamard transform* [HS 79]. The important point is that the matrices describing these transforms can be factored into the *direct* product of smaller matrices [Schr 90]. Specifically, a Fourier or Hadamard matrix with 2^n rows and 2^n columns can be iteratively factored into n 2×2 matrices, and it is precisely this iterative decomposition that results in the substantial computational savings by a factor $2n/2^n$. Not surprisingly, Hadamard matrices generated in this manner by iteration show self-similarity when represented as graphic images or optical masks.

The iterated *baker's transformation* is revealed as playing musical chairs with binary digits while simulating nonlinear transformations in two-dimensional spaces. Arnol'd's cat map leads to another chaos-producing recursion for modeling area-preserving transformations. (The *logistic parabola*, the paradigm of nonlinear mapping, and its two-dimensional generalization, the *Hénon map*, are analyzed in Chapter 12.)

But recursion, properly controlled, engenders beauty, too, be it in the form of handsome designs for needlework or shapely trees and flowers. And in mathematics proper, recursion weaves some veritable wonders—such as the computation of a billion digits of π in just 15 steps, a computational coup that is based on the work of Srinivasa Ramanujan (1887–1920), the great Indian mathematician whose awesome intuition remains beyond human comprehension.

Looking for Zeros and Encountering Chaos

Some 300 years ago, Isaac Newton (1642–1727) suggested finding the zeros of a function $f(z)$ by an *iteration*, based on drawing tangents. Given an approximate value of z_n to the solution of $f(z) = 0$, Newton finds the next approximation by calculating

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)} \tag{1}$$

where $f'(z_n)$ is the derivative of $f'(z)$ at $z = z_n$. For equation 1 to be applicable, the slope $f'(z_n)$ of the tangent must be different from zero.

For $f(z) = z^2 - 1$, equation 1 reads

$$z_{n+1} = \frac{1}{2}\left(z_n + \frac{1}{z_n}\right) \tag{2}$$

Not surprisingly, for an initial value z_0 with a positive real part, z_n converges toward the positive root of $z^2 - 1 = 0$, namely, $z = 1$. Similarly, for a negative real part of z_0, the solution converges to the negative root, $z = -1$.

But what happens for purely imaginary $z_0 = ir_0$, where $r_0 \neq 0$ is real? Interestingly, it does not converge at all; the iteration in equation 2 cannot make up its mind, so to speak, and hops all over the imaginary axis, according to the mapping

$$r_{n+1} = \frac{1}{2}\left(r_n - \frac{1}{r_n}\right) \tag{3}$$

For example, the golden mean $0.618\ldots$ maps iteratively into -0.5, 0.75. $-0.291\overline{6}$, $1.56845\ldots$ and so on. But some r_0 behave quite differently, such as $r_0 = 1 + \sqrt{2}$, which maps into 1, 0, and ∞, a kind of fixed point, if rather distant.

How can we inject order into this chaotic mapping? A trigonometric substitution will do the trick:

$$r = -\cot(\pi\alpha) \tag{4}$$

which turns equation 3 into the exceedingly simple iteration

$$\alpha_{n+1} = 2\alpha_n \bmod 1 \tag{5}$$

where "mod 1" means subtracting the integer part and keeping only the fractional part, which lies in the half-open interval $[0, 1)$. For example, 2.618 mod 1 equals 0.618.

In terms of the new variable α, the chaotic mapping of r becomes totally transparent. If we express α_n as a binary fraction, then the digits of α_{n+1} will be simply the same as those of α_n shifted one place to the left. A 1 that moves to the left of the binary point is dropped. Thus, a periodic binary number α_0 will lead to periodic *orbits* (i.e., periodic sequences of iterates). For example, $\alpha_0 = \frac{1}{3} = 0.\overline{01}$ will map into $0.\overline{10} = \frac{2}{3}$, which will map right back into $\alpha_2 = 0.\overline{01} = \alpha_0$. (Indeed, $r_0 = -\cot(\pi/3) = -1/\sqrt{3}$ is mapped by equation 3 into $r_1 = 1/\sqrt{3}$, which maps back to $r_2 = -1/\sqrt{3} = r_0$.)

Similarly, *preperiodic* binary fractions, which begin aperiodically and end in a periodic tail, such as $\alpha_0 = \frac{5}{6} = 0.1\overline{10}$, lead to *preperiodic* orbits:

$$r_0 = \sqrt{3}, \quad r_1 = \frac{1}{\sqrt{3}}, \quad r_2 = -\frac{1}{\sqrt{3}}, \quad r_3 = r_1, \quad \text{etc.}$$

Terminating binary fractions are but a special kind of preperiodic binary fraction, the periodic tail being $\bar{0}$. Where will *they* end up? Continued left-shifting and taking fractional parts (remainders modulo 1) will, sooner or later, produce $\alpha = 0$, which corresponds to $r = \infty$. For example, $\alpha_0 = \frac{7}{8} = 0.111$ will map into $0.11 = \frac{3}{4}$ and, then $0.1 = \frac{1}{2}$, which maps into 0. In fact, the corresponding $r_0 = -\cot(7\pi/8) = 1 + \sqrt{2}$ maps into 1, 0, and ∞, as we already noted.

We also see by this analysis that any irrational α_0 will lead to an aperiodic "orbit" along the imaginary axis in the z plane. Thus, the simple mapping given by Newton's iteration for the function $z^2 - 1 = 0$ has rather strange consequences for initial values on the imaginary axis: in terms of the corresponding values of α, numbers are classified into three categories:

1 Periodic binary rational numbers

2 Preperiodic binary rational numbers

3 Irrational numbers

Periodic and preperiodic binary rational numbers α_0 converge on a fixed point or lead to periodic orbits. By contrast, irrational numbers α_0, an uncountable set, give aperiodic orbits: the same value never occurs twice, nor is α_n ever rational. Surprisingly, the simple map in equation 5 even makes a subtle distinction between different kinds of irrational numbers—not the usual number-theoretic distinction between algebraic (such as $\sqrt{2}$) and transcendental irrational numbers (for example, π), but between *normal* and *nonnormal* numbers, including "Liouville" numbers.

A normal number (in a given number system) is defined as a number in which every possible block of digits is equally likely to occur. For example, on the evidence of its first 100 million digits, π appears to be normal in the decimal system [Wag 85]. This means, for example, that somewhere in the decimal expansion of π a string of eight 7s will occur—in fact, there is a good probability that this will occur in the first 10^8 digits of π (or any other normal decimal number). For up-to-date evidence, see the book by Klee and Wagon [KW 89]. Iteration of normal numbers under the rule of equation 5 gives rise to chaotic orbits.

How awesome an object a normal number is can perhaps best be appreciated by the following reflection. The entire contents of the *Encyclopaedia Britannica* can be coded into a single decimal number (about 10^{10} digits long). And somewhere in the expansion of a normal number this block of digits will occur. In fact, the contents of the encyclopedia will occur infinitely often! (But don't ask where!)

Are there any *nonnormal* irrational numbers? In fact there are uncountably many in any number system! For example, the Cantor numbers (see Chapter 7) are nonnormal in the ternary system, because they eschew the digit 1. In subsequent chapters we shall get to know the Morse-Thue constant $= 0.01101001\ldots$ and the rabbit constant $= 0.10110101\ldots$ neither of which can boast a triple 1 in its binary expansion. Thus, they cannot be normal binary numbers.

But there are even stranger nonnormal numbers, such as the binary

$$L = \sum_{k=1}^{\infty} 2^{-k!} = 0.110001000000000000000001000\ldots$$

and other Liouville numbers[1] that are irrational yet very close to rational numbers.

In general, a Liouville number β is defined as an irrational number for which rational numbers p and q exist such that

$$\left| \beta - \frac{p}{q} \right| < \frac{1}{q^n} \tag{6}$$

for *any* n. In fact, to satisfy inequality 6 for the Liouville number L, one sets $q = 2^{n!}$. The resulting approximation error is then $2^{-(n+1)!} + 2^{-(n+2)!} + \cdots$, which (for $n > 1$) is smaller than $2^{-n!n} = q^{-n}$. (For $n = 1$, one sets $q = 4$ and notes that $\left| L - \frac{3}{4} \right| < \frac{1}{4}$.) The fact that for algebraic irrational numbers of degree n the absolute difference in inequality 6 *exeeds* $1/q^{n+1}$ demonstrates the existence of numbers, such as L, that "transcend" the algebraic numbers, namely, the *transcendental numbers.*

The iterates according to equation 5 of normal binary numbers α_0 will, by their definition, fill the unit interval densely and with equal probabilities for any subintervals of equal lengths. Thus their stationary distribution under the iteration, called the *invariant distribution,* is in fact flat. By contrast, *non*normal numbers will do nothing of the kind. For example, iterates of L will accumulate at 0 and all negative powers of 2 (i.e., $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$, and so on.)

Thus, we have to face the curious fact that the tiniest distinctions—the differences between rational and irrational numbers and, among the irrational numbers, between normal and nonnormal numbers—make a decisive difference in the final fate of a numerical iteration. Ordinarily, one should think that physics, and certainly the *tangible* world at large, would be untouched by the purely mathematical dichotomies between rational and irrational numbers or normal and nonnormal numbers. But in reality this is just not so. While, true enough, everything in the real world can be adequately described by rational numbers, it so happens that a mathematical *model* that distinguishes between different kinds of

---

1.   Named after Joseph Liouville (1809–1882), who exhibited the first transcendental number of which L is a prototype. The proofs that e and π are transcendental came later from Charles Hermite (1822–1901) and Ferdiand Lindemann (1852–1939), respectively.

Another (existence) proof of transcendental numbers was furnished by Cantor, who showed that algebraic numbers (i.e., the solutions of polynomial equations with rational coefficients) form a countable set. Since the continuum, by contrast, is an uncountable set, there must be numbers (in fact, uncountably many) that "transcend" the algebraic irrational numbers, namely, the so-called transcendental numbers.

numbers is not only extremely useful but catches the true, possibly hidden, spirit of the problem.

More specifically, two different initial conditions of a physical system that are completely indistinguishable by any finite measurement precision will sooner or later lead to a total divergence as the system evolves in time or space. The essential condition for this to happen is that the corresponding iteration, called *Poincaré mapping* (see chapter 14), be sufficiently *nonmonotonic*, such as equation 5, which has a sawtooth nonmonotonicity. The rate of divergence is measured by the so-called *Lyapunov exponent* $\lambda := \log (\alpha_{n+1}/\alpha_n)$ for $n \to \infty$. In our bare-bones example based on the iteration in equation 5, $\lambda = \log 2 \approx 0.693$ (if we take the natural logarithm).

Although the iteration in equation 5 may look unrealistically simple, it captures the essence of innumerable nonlinear problems which show period-doubling bifurcations and thus follow one of the two outstanding routes to chaos (see Chapter 12). (The other route is by quasi periodicity modeled by the so-called *circle map*, discussed in Chapter 14.)

The Strange Sets of Julia

Newton's method (equation 1) applied to the cubic equation $f(z) = z^3 - 1$ gives the iteration

$$z_{n+1} = \frac{2z_n^3 + 1}{3z_n^2} \tag{7}$$

which we already encountered in pages 38–40 in Chapter 1.

Now the mapping is even crazier. For one thing, the naive conjecture that all z_0 will converge toward the closest of the three roots is false. For example, $z_0 = -1$ will converge on 1, the root most distant from it.

Color Plate 3 shows the intimately intertwined basins of attraction of the three roots—a real, or rather imaginary, crazy qulit. In fact, it can be shown that for the mapping in equation 7 two basins (colors) can never meet if the third is not present also. This may sound impossible, and in fact it would be—if it were not for the fractal nature of the boundaries as intimated in Color Plate 3, which also shows the attractive self-similarity resulting from Newton's iteration.

Why are the three basins of attraction not simply shaped like three pieces of pie or sectors, each 120° wide? After all, Newton's iteration for $z^2 = 1$ leads to two basins that are half planes. That this cannot be so for $z^3 = 1$ becomes clear when one looks at the point $z = -2^{-1/3}$, which is mapped to the origin, $z = 0$, by equation 7. The neighborhood of the point $z = 0$ contains points from all three basins of attraction (because of the 120° rotational symmetry of the problem). Thus, because the inverse map of equation 7 is continuous, the point

$z = -2^{-1/3}$ must also be surrounded by points from all three basins of attraction. In fact, its infinitesimal neighborhood must be a scaled-down version of the "cloverleaf" at the origin. As a consequence, there are points below the negative real line that belong to the basin of attraction for the attractor $z_1 = \omega := \exp(i\pi/3)$, which lies *above* the real line. Similarly, there are points above the negative real line that will converge on $z_2 = \omega^2$, which lies *below* the real line. Thus, the basins cannot be simple sectors; the basins "nibble" at each other's pies.

In fact, quite generally for $z^n = 1$, Newton's method will produce at the origin basins that form a cloverleaf with n leaves. A preimage of the origin that falls on the boundary between two attractors will therefore intermingle *all* basins at that point, and for $n > 2$, this intermingling must result in a fractal boundary, because in two dimensions only two attractors can meet in a smooth boundary. Furthermore, since *all* boundary points are preimages of the origin, they are boundary points of *all* n basins. Such strange sets of boundary points are customarily called *Julia sets*. By definition, the Julia set J of a rational function $R(z)$ is the set of points z for which $R(z)$ is *not normal*. (Normal points are those for which $R(z)$ is equicontinuous in a neighborhood of z.)

The Julia set of a rational function has the following astounding property. If z_k is a periodic attractor and A_k its basin of attraction, then $J = \partial A_k$ for all k. Here ∂A_k is the *boundary* of the basin A_k, that is, all those points in whose neighborhood, no matter how small, one finds points both in A_k and outside A_k. Hence, if we have a point z that lies on the boundary of A_1, say, we know that it also lies on the boundaries of $A_2, A_3, \ldots$

Another important property of a Julia set J is that periodic repellors of $R(z)$ are dense in J. In fact, if z_r is a periodic repellor, then J is the closure of the set of all the preimages of z_r. As a result, the dynamics on a Julia set are chaotic, that is, sensitive to initial conditions, as we already saw in pages 239–241 for the Julia set of $R(z) = (z + 1/z)/2$. A Julia set always contains an uncountable number of points, but it is not necessarily fractal.

Julia sets, consisting of repellors, are difficult to plot because of the chaotic sensitivity, which requires unrealistic numerical precision. However, another property of J comes to the plotter's aid: for any z in J, the *inverse orbit* $R^{-n}(z)$, $n = 1$, $2, 3, \ldots$, is dense in J. For points in J, $R^{-n}(z)$ is attractive and so there are no problems with numerical divergence. However, $R^{-1}(z)$ is in general multivalued, so that clever algorithms are needed to cover the entire Julia set uniformly. Such algorithms are described by Peitgen and Saupe in *The Science of Fractal Images* [PS 88]. Figure 2 shows the dustlike Julia set for the mapping of equation 7 generated by such an algorithm.

The simple example of a rational mapping with three attractors and its Julia set discussed here is not devoid of physical implications. Consider, for example, a pendulum consisting of an iron bob at the end of a string. Below the pendulum are three magnets to which the iron bob is attracted. After some oscillations, the pendulum will come to rest with the bob directly above one of the three magnets. But will the bob always go the attractor nearest to its initial position?
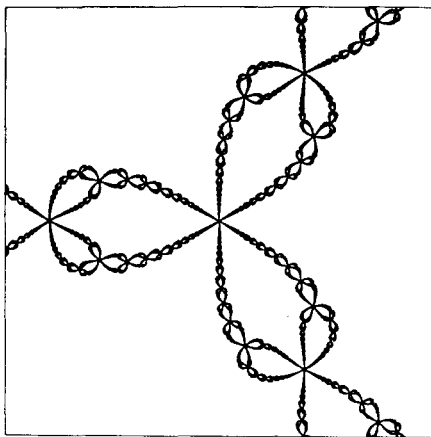
Figure 2   Julia dust for Newton iteration [PS 88].

It will not. Try it! The iron will, for many initial conditions, follow a rather tortuous route, and the end position will seem totally unpredictable—and it often not only will *seem* unpredictable but will be unpredictable, unless the initial position can be given with a totally unrealistic precision. In other words, the basins of attraction of the three magnets are bounded by fractal sets—just as the Julia set in our Newton iteration.

A Multifractal Julia Set

As we have seen in the preceding section, Julia sets of rational functions with more than two attractors are fractals—in fact, they are *multi*fractals. Such sets have been traditionally analyzed by generating individual points of the set by numerical backward iteration. However, for some Julia sets, *analytical* methods, which offer much higher accuracy and require less computing, are feasible. A prime example is the recent analysis by Nauenberg and Schellnhuber of the multifractal properties of the Julia set associated with the Newtonian map (equation 7) for the solution of $z^3 - 1 = 0$ [NS 89].

   The first order of business is to construct the *support* of the fractal set. For the original Cantor set the support is the unit interval: all members of the set "live" on the straight unit interval. For our Newtonian Julia set, by contrast, the support is already a highly complicated manifold, an infinitely nested "cobweb." To construct this spidery support, Nauenberg and Schellnhuber note that one of the three preimages of the negative real line is the interval $-\infty < z \leq -2^{-1/3}$, called M_0 in Figure 3A. The two other preimages are obtained from M_0 by rotations of $\pm 120°$.
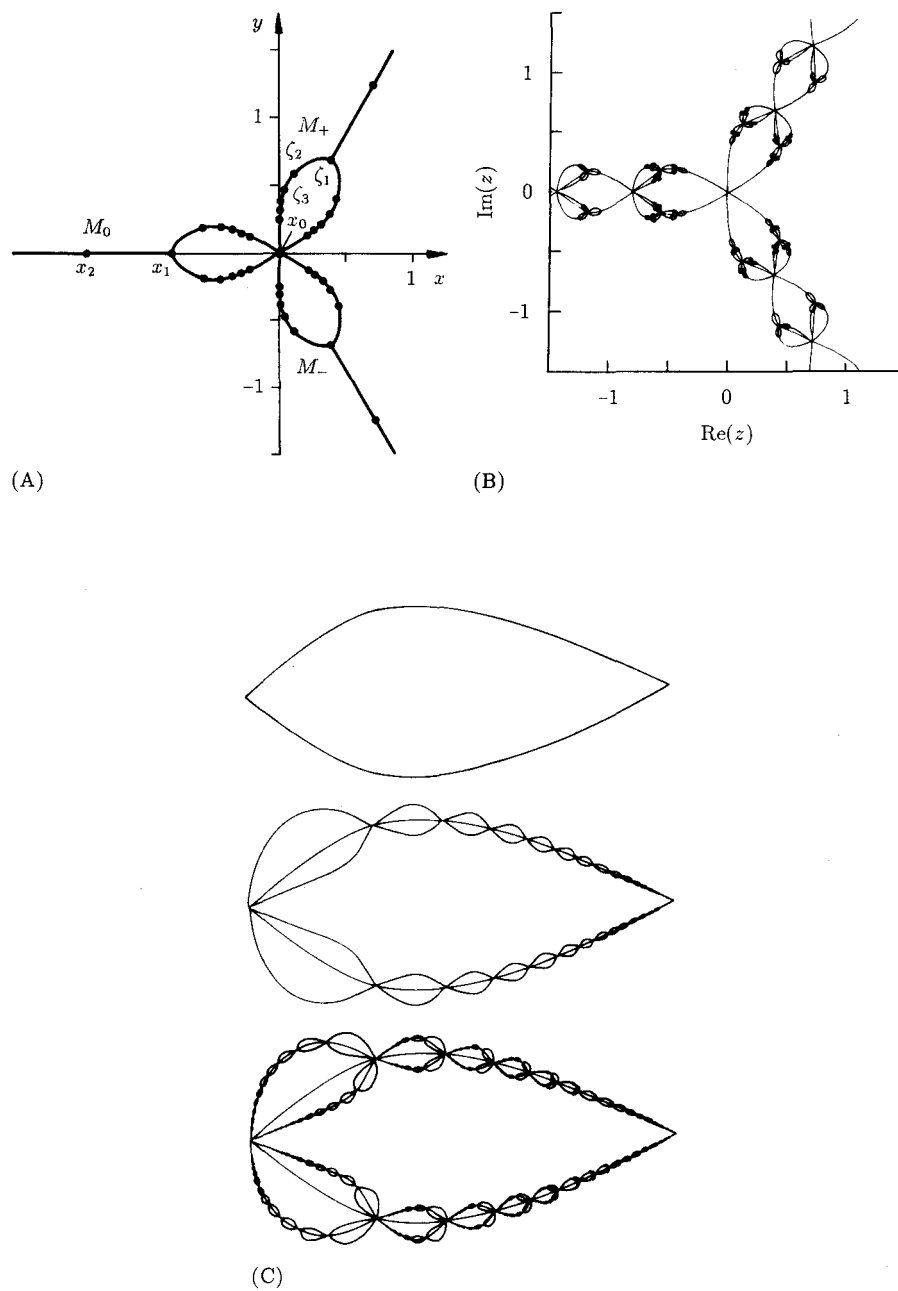
(A)

(B)

(C)

Figure 3 (A) Constructing the Julia flower: the first-generation preimages. (B) Support of the Julia-flower set at fourth stage of construction. (C) Julia petal: the first three steps of construction [NS 89].

The two preimages M_+ and M_- of M_0 that do not fall on the negative real line are given by

$$M_\pm = \xi \pm i(\xi^{1/2} - \xi^2) \qquad 0 \le \xi \le 2^{-4/3} \tag{8}$$

The other four arclike second-order preimages are obtained by $\pm 120°$ rotations. The six arcs together form the central "flower" shown in Figure 3A. The six arcs and the three straight-line spikes at the end of the petals together form the first generation of the support of the fractal. The second generation consists of $3 \cdot 9 = 27$ parts, and so on. Figure 3B shows the support at the fourth stage of construction, made up of the first four generations obtained by backward iteration from the central flower. The nested character of this Julia fractal is already clearly evident.

The higher-order preimages x_n of $z = 0$ falling on the negative real line are given by the recursion

$$x_{n+1} = \frac{x_n}{2} + x_n \cosh\left[\frac{1}{3}\cosh^{-1}(1 - 2x_n^{-3})\right] \tag{9}$$

where $x_1 = -2^{-1/3}$. The asymptotic scaling factor of the x_n is $\frac{3}{2}$, as can be deduced from the derivative of the forward iteration, $dN/dz \to \frac{2}{3}$ for large z.

The preimages $\zeta_n = \xi_n + i\eta_n$ of the x_n on M_+ are given by

$$\xi_{n+1} = -x_n \sinh^2\left[\frac{1}{3}\sinh^{-1}((-x_n)^{-3/2})\right] \tag{10}$$

and $\eta_{n+1} = \pm(\xi_{n+1}^{1/2} - \xi_{n+1}^2)$ according to equation 8. The arc lengths l_n on M_+ between ξ_n and ξ_{n+1} scale asymptotically as $l_n \sim (\frac{2}{3})^{n/2}$. The longest arc has length $l_1 = 0.3834$.

For the calculation of the fractal dimensions D_q, we focus on one of the three petals of the central flower (see Figure 3C). In the first generation the petal consists of just two large arcs. As the next step, we consider the infinite succession of smaller arcs whose vertices, lying on these arcs, are given by the first-order preimages of the points x_n on the negative real line. Note that each arc generates a succession of *double* arcs. The bottom of Figure 3C shows the result of the third step, in which each smaller arc has sprouted an infinite succession of even smaller double arcs.

The Julia set that we are interested in, namely, the common boundary points of three basins of attraction, consists of the *branch* points of this support. (This is analogous to the original Cantor set, which is given by the *end* points of the surviving intervals.) The fractal dimensions D_q of this Julia set are given by the expression

$$2\sum_m p_m^q l_m^\tau = 1 \qquad \tau = (1 - q)D_q \tag{11}$$

(see Chapter 9) where the summation is over successive generations of the support-generating process and the factor 2 reflects the fact that each mother arc gives birth to twin daughter arcs. If we choose the three preimages with equal weight, then $p_m = (\frac{1}{3})^m$. The l_m are the arc lengths computed with equations 8 to 10.

Of all the dimensions, D_∞ is the easiest to compute, because for $q \to \infty$, $\tau \to -\infty$, leaving only the largest l_m in equation 11 to make a contribution. The largest l_m is $l_1 \approx 0.3834$. Hence, with $p_1 = \frac{1}{3}$.

$$D_\infty = \frac{\log p}{\log l_1} = 1.146 \ldots$$

For $q = \pm \infty$, the factor 2 in equation 11 is irrelevant. The other extremal dimension, $D_{-\infty}$, is also easy to calculate. Here the *smallest* l_m in equation 11 dominate the sum, that is, l_m for $m \to \infty$. With $l_m \sim (\frac{2}{3})^{m/2}$ for large m and $p_m = (\frac{1}{3})^m$, equation 11 yields

$$D_{-\infty} = \lim_{m \to \infty} \frac{\log 3}{\frac{1}{2} \log \frac{3}{2}} = 5.419 \ldots$$

Note that $D_{-\infty}$, unlike the Hausdorff dimension D_0, has no simple geometric meaning. Its value, being larger than 2 (the Euclidean dimension of the space embedding the Julia fractal), is therefore no contradiction.

Interestingly, in the present approach, the numerical value of $D_{-\infty}$ is given by a simple analytical fact: namely, that the l_m are proportional to $(\frac{2}{3})^{m/2}$. By contrast, numerical methods are not even feasible for $q = -\infty$, because the computer would have to run forever to explore the sparsest regions of the Julia set that are characterized by $D_{-\infty}$. Thus, one is compelled to estimate $D_{-\infty}$ from D_q for large q. But this access is hampered by a frustratingly slow numerical convergence. The other dimensions, including the Hausdorff dimension D_0, are likewise obtained by elementary calculations. For good approximations, only a few of the l_m in equation 11 need to be calculated explicitly; for the remaining terms, the approximation $l_m = 0.1986(\frac{2}{3})^{m/2}$ suffices. The result for the Hausdorff dimension is $D_0 = 1.429 \ldots$, a considerably more accurate result than the one obtained by number-crunching methods on the basis of 1 million points of the Julia set. Note that $D_0 < 2$, as behooves a two-dimensional "dust." For the important information dimension (see Chapter 9),

$$D_1 = \frac{\sum_m p_m \log p_m}{\sum_m p_m \log l_m}$$

just *two* rough values of l_m, $l_1 \approx 0.38$ and $l_2 \approx 0.18$, suffice for a good estimate, $D_1 \approx 1.2$.

The Beauty of Broken Linear Relationships

Figure 4 shows the result of applying the piecewise linear "sawtooth" map $\tilde{b} = 8b$ mod 1 to the local brightness $0 \leq b \leq 1$ of the photograph of a human face and quantizing the result $\tilde{b}$ to 0 if $\tilde{b} > b$ and to 1 for $\tilde{b} \leq b$ [Schr 69]. Figure 5 shows the beautiful pattern that results from the iteration of another simple piecewise linear function [PR 84].

Such maps are useful mathematical models of *deterministic* diffusion [Schu 84]. While traditional (thermodynamic) diffusion is a typical random process, piecewise linear maps, such as that shown in Figure 6, cause a particle to drift in a seemingly random manner, although the drifting is a strictly deterministic process. This drifting is another example of a chaotic motion which depends strongly on the precise (and not exactly knowable) initial coordinate value. As in traditional diffusion, the spatial correlation between two initially close particles decays to zero as time evolves [Gro 82]. As a consequence, the mean squared drift $\overline{x^2}$ increases linearly with time, as in ordinary diffusion:

$$\overline{x^2} = 2Dt \qquad \text{for } t \gg 1$$

Figure 4   Sawtooth map of "Karen," an early computer graphic by the author [Schr 69].
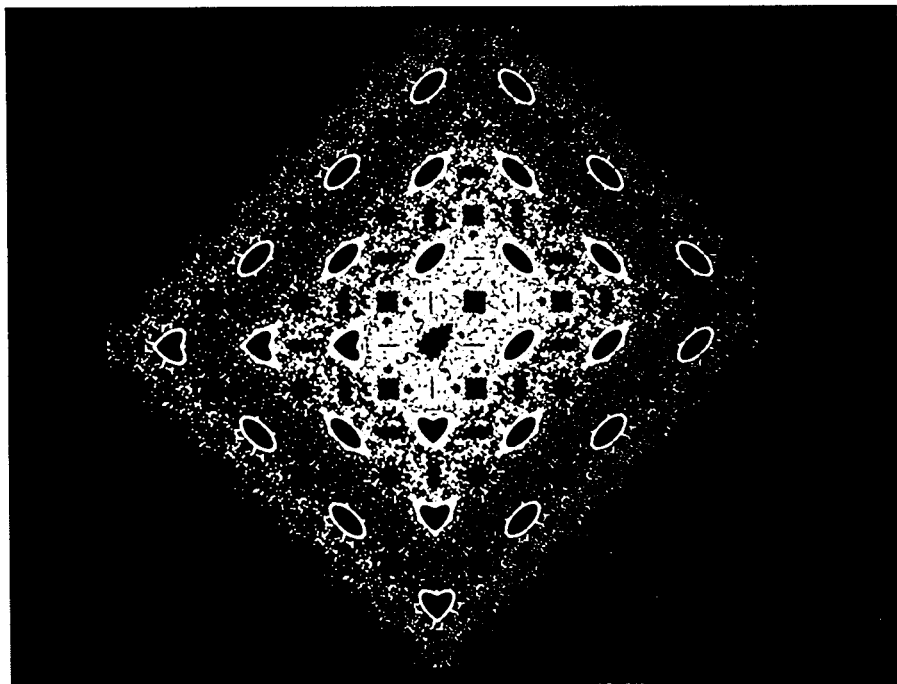
Figure 5 Embroidery by broken-linear relationship [PR 84].

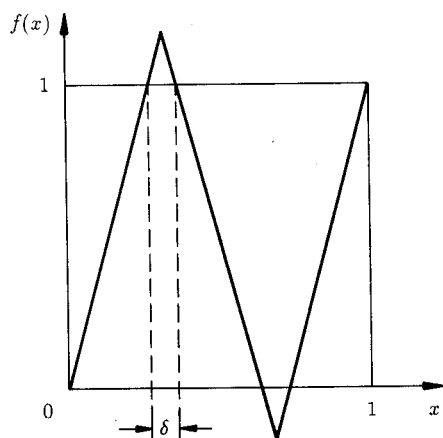***Figure 6*** Broken-linear map for modeling deterministic diffusion [Gro 82].

with a diffusion constant D that is proportional to the width δ over which the iterated function "sticks out" from the unit square.

The Baker's Transformation and Digital Musical Chairs

Bakers mix their dough by rolling it out, folding it over, and rolling it out again, in a seemingly endless iteration—roll-fold-roll-fold . . .—until they are satisfied that they have achieved a sufficiently uniform mixture of the dough's ingredients.

A mathematically sanitized version of the dough rolling and folding, called the *baker's transformation*, is illustrated in Figure 7. It is a useful model of all kinds of mixing processes, including the chaotic mixing of fluids. Arithmetically, a point (x, y) in the unit square is transformed to a point $(2x, y/2)$ by rolling out and to $(\langle 2x \rangle_1, y/2 + \lfloor 2x \rfloor/2)$ by subsequently cutting the rolled-out dough in half and putting the right half on top of the left half. (This operation is mathematically simpler than the folding over.) Here, as before, the pointed brackets
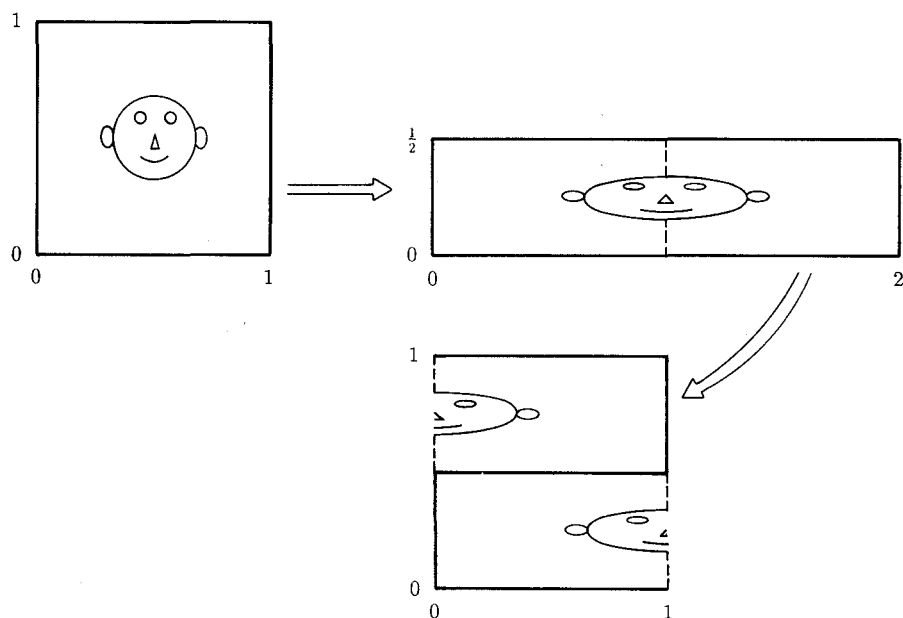

Figure 7   Baker's map, a recipe for chaotic mixing. The unit square is first "rolled out" to a rectangle. The right half is then cut off and stacked on top of the left half.

mean "take the fractional part" and the open-top brackets "round" down to the nearest integer."

If we express x and y in binary notation, then the baker's transformation becomes particularly simple: the digits of x are left-shifted by one place, the digits of y are right-shifted by one place, and the leftmost digit of x becomes the leftmost digit of y. In fact, the binary digits of x and y play "musical chairs." For example, the coordinate pair

$$x_1 = 0.10110001\ldots$$
$$y_1 = 0.01110100\ldots$$

is mapped into

$$x_2 = 0.0110001\ldots$$
$$y_2 = 0.101110100\ldots$$

which goes into

$$x_3 = 0.110001\ldots$$
$$y_3 = 0.0101110100\ldots$$

and so on. Thus, any terminating fraction for x will asymptotically approach the origin $(0, 0)$, which is therefore an attractor for such x values.

Periodic binary x values will converge on periodic orbits. For example, $x = \frac{1}{3} = \overline{0.01}$, $y = \frac{2}{3} = \overline{0.10}$ will lead to a periodic orbit of period length 2. By contrast, normal binary numbers will follow chaotic orbits, in which even initially close points are soon separate and follow independent orbits; see Figure 8. Both chaotic and nonchaotic flows are discussed in Ottino's "The Mixing of Fluids" [Ott 89].
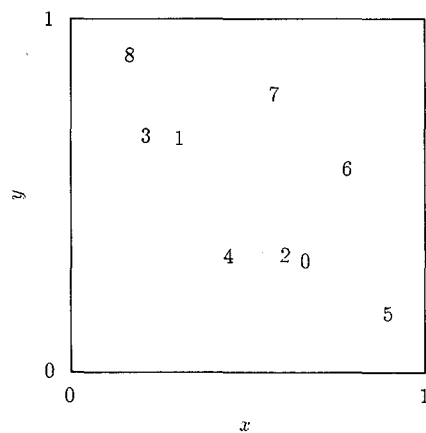

Figure 8 A chaotic orbit for the baker's map starting *near* the period -2 point $x_0 = \frac{2}{3}$, $y_0 = \frac{1}{3}$.
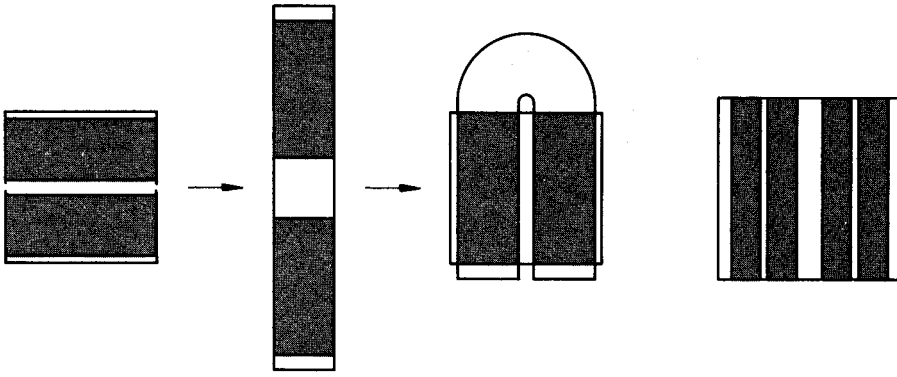
Figure 9   Smale's horseshoe map. A space is stretched in one direction, squeezed in another, and then folded. When the process is repeated several times, a pair of points that end up together may have begun far apart.

Sometimes the baker throws away some dough during every iteration, in which case the remaining dough turns into a Cantor dust in one direction. This generalized baker's transformation is a simple model of the phase spaces of nonlinear dynamic systems that contract in some directions and thus sport strange attractors, such as the Julia sets discussed in pages 243–248.

Related transformations are Smale's horseshoe map [Sma 67] (see Figure 9) and Hénon's map [Hen 76] (see Figure 10), which are characteristic of dissipative physical systems with strange attractors.

Where is the self-similarity in Hénon's map? Take another look at Figure 10D and the Hénon attractor after 10^4 iteractions shown in Figure 11. While strange attractors may be strange in many ways, they do maintain self-similar order in their chaotic orbits.

Arnol'd's Cat Map

Another important two-dimensional map, for the description of Hamiltonian nonlinear systems, is Arnol'd's area-preserving cat map,

$$x_{n+1} = x_n + y_n \bmod 1$$
$$y_{n+1} = x_n + 2y_n \bmod 1$$

(12)

illustrated in Figure 12A and B. The notation "mod 1" means, as before, that only fractions in the half-open unit interval [0, 1) are considered. The distortion engendered by this transformation is reminiscent of a maladjusted, out-of-sync
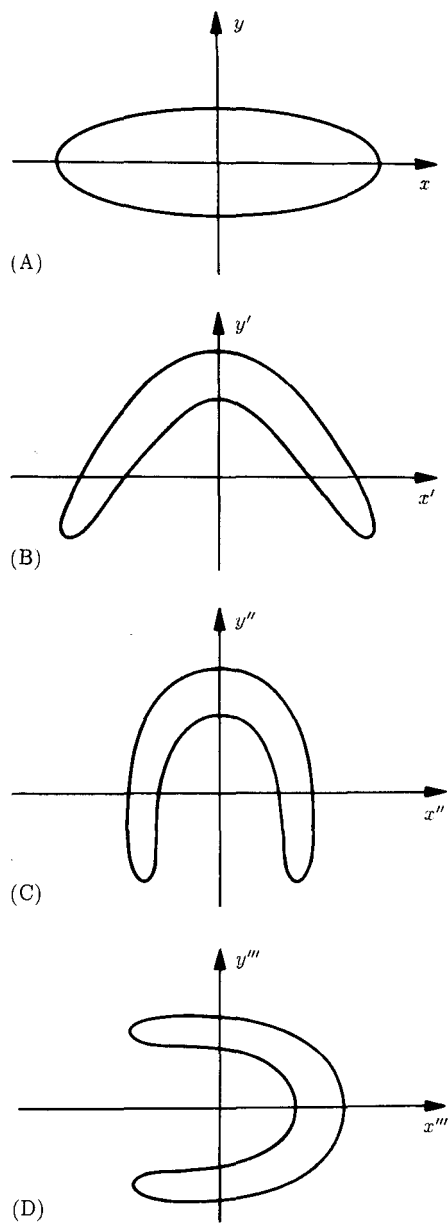
Figure 10   Hénon's map. (A) Initial ellipse. (B) Area-preserving bending: $x' = x$, $y' = 1 - ax^2 + y$. (C) Contraction in the x direction: $x'' = bx'$, $y'' = y'$. (D) Rotation by 90°: $x''' = y''$, $y''' = -x''$ [Schu 84].
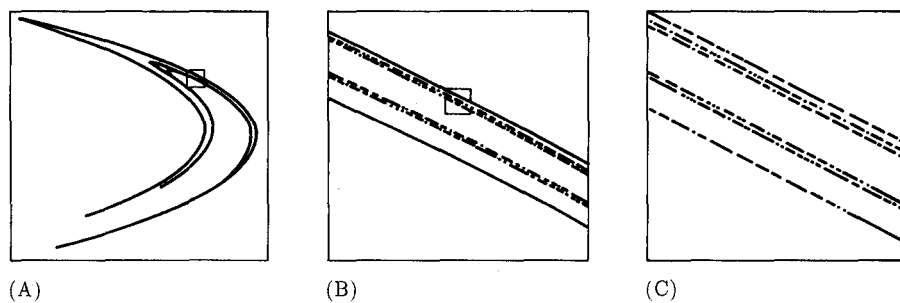
Figure 11   Self-similarity of the Hénon attractor. (A) The entire attractor. (B) Enlargement of portion shown by small square in part A. (C) The result of another enlargement. Note the similarity of the streak patterns between parts B and C, attesting to the ultimate self-similarity of the Hénon attractor [Far 82].
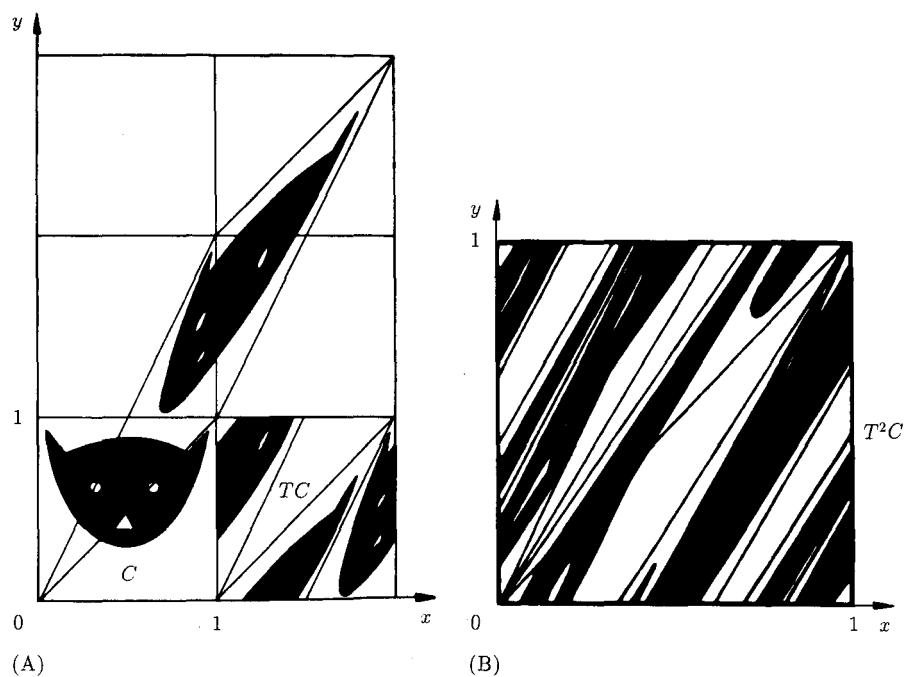
Figure 12   (A) Arnol'd's cat map; (B) iterated cat map.

television set. The cat map's two eigenvalues, λ_1 and λ_2, are the squared reciprocal of the golden mean γ and its inverse: $\lambda_1 = (3 + \sqrt{5})/2 > 1$ and $\lambda_2 = \lambda_1^{-1} = (3 - \sqrt{5})/2 < 1$. Because one eigenvalue is larger than 1 and the other is smaller than 1, its fixed points are all *hyperbolic*, meaning that the map is expanding in one direction (corresponding to the eigenvalue λ_1) and contracting in the orthogonal direction (corresponding to λ_2). Hyperbolic fixed points are therefore neither repellors nor attractors; they are both, depending on the direction of approach. In geometric representations, hyperbolic fixed points are *saddle* points: a ball rolls down the mountain *toward* a saddle point (mountain pass), but then it will move *away* from the saddle point toward the valley. The mountain pass first *attracts* and then *repels* water running down the mountain in its direction.

Maps with hyperbolic fixed points are the hallmark of chaotic motion in energy-conserving physical assemblies, called Hamiltonian systems [Arn 89]. For example, $(x, y) = (0.4, 0.2)$ is a fixed point of the once-iterated cat map; that is, it belongs to an orbit with period length 2. It is transformed into $(0.6, 0.8)$, which is mapped back into $(0.4, 0.2)$. A perturbation of the initial point $(0.4, 0.2)$ in the direction corresponding to the eigenvalue λ_2, $\Delta x/\Delta y = -(\sqrt{5} + 1)/2 = -1/\gamma$, is contractive. Indeed, the initial even-numbered iterates x_{2n} of $x_0 = 0.4 + 1/100\gamma$, $y_0 = 0.2 - 1/100$ are, beginning with x_2, approximately 0.402, 0.4003, 0.40005, 0.400007, 0.400001, 0.4000002. Thus, the point $(0.4, 0.2)$ acts like an attractor when approached from the direction that corresponds to λ_2. But following this convergence, using a 12-digit calculator, which introduces a small perturbation in the λ_1-direction, numerical inaccuracies cause eventual divergence, as evidenced by subsequent iterates: 0.4000005, 0.400003, 0.40002, 0.4002, 0.401, 0.407, 0.45, 0.75, and so on to chaos. The point is that, because of this attractive-repelling nature, any numerical calculation, or any physical system modeled by the cat map, will sooner or later show chaotic motion, except for rare initial conditions with measure zero. Such initial conditions, of course, cannot be physically realized; they correspond to unstable equilibria—such as a pencil standing on its tip.

One of the advantages of the cat map is that its iterates are easy to analyze. In matrix form, the cat map (equation 12) is given by the matrix

$$T \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

which transforms the column vector (x_n, y_n) into the column vector $(x_{n+1}\ y_{n+1})$, where all x and y are taken modulo 1.

Given the recursion for the Fibonacci numbers $F_n = F_{n-1} + F_{n-2}$, $F_1 = F_2 = 1$, it is easy to show by induction that the nth iterate of the cat map is

$$T^n = \begin{pmatrix} F_{2n-1} & F_{2n} \\ F_{2n} & F_{2n+1} \end{pmatrix}$$

which inherits its area preservation from that of T. (Indeed, the determinant of T^n, $F_{2n-1}F_{2n+1} - F_{2n}^2$, equals 1.) Since T^n is a symmetric matrix, it has only real eigenvalues, namely, $\lambda_1^{(n)} = \lambda_1^n > 1$ and $\lambda_2^{(n)} = \lambda_2^n < 1$.

Fixed points of T^n correspond to orbits of period length n and any divisors of n. For $n = 2$, for example,

$$T^2 = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

To find the fixed points of period length 2, one has to solve the equations

$$2x + 3y = x \bmod 1$$

$$3x + 5y = y \bmod 1$$

Apart from the solution $x = y = 0$ (which has period length 1), eliminating y results in $5x = 0 \bmod 1$, that is, $x = k/5$, where k is an integer. The only allowed values for $0 < x < 1$ are $k = 1, 2, 3, 4$, each of which, in fact, yields a solution. The corresponding values of y are $y = k'/5$, with $k' = -2k \bmod 5$. These four periodic points of period length 2 form two orbits, namely, $(\frac{2}{5}, \frac{1}{5}) \to (\frac{3}{5}, \frac{4}{5})$, which we already encountered, and $(\frac{1}{5}, \frac{3}{5}) \to (\frac{4}{5}, \frac{2}{5})$.

We leave the highly instructive analysis of the complete orbit structure of the cat map to the interested reader.

A Billion Digits for π

Iteration is one of the most powerful mathematical tools. To calculate the value of π by means of the Gregory-Leibniz series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

to an accuracy of just three decimal places, one needs to sum 500 terms. By contrast, a recursion involving the arithmetic-geometric mean doubles the number of correct digits with every iteration [BB 87, KW 89]. And there are iterative algorithms, based on Ramanujan's work, that multiply the number of decimal digits for each iteration by 4 or even 5 [BBB 89]. Thirteen iterations of such an algorithm have yielded more than 134 million digits of π, and just two more iterations would give π to an accuracy exceeding two *billion* digits. This is a relative error of 10^{-10^9} (not just 10^{-9}), an awesome accuracy.

Of course, nobody needs π that accurately. Among the most accurate measurements in physics are those exploiting the Mössbauer effect, with a precision of, say, 14 decimal digits, corresponding to an error of 1 second in 3 million years. And 39 digits for π suffice to calculate the circumference of the known universe from its radius to within the diameter of the hydrogen atom. However, computations of the digits of π have become benchmark tests for supercomputers and superfast algorithms. Needless to say, these calculations do not run on pocket calculators; for these do not have enough memory and their displays are too limited for the purpose at hand.

One of Ramanujan's astonishing results is the following formula:

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801} \sum_{n=0}^{\infty} \frac{(4n)![1103 + 26390n]}{(n!)^4(396^{4n})} \tag{13}$$

whose very first term gives π with a relative accuracy better than $3 \cdot 10^{-8}$. Each additional term adds about eight more decimal digits (i.e., multiplies the accuracy by 100 million).

Very rapid approximations to π are based on a seminal paper by Ramanujan which established a close connection with the transformation theory for elliptic integrals [Ram 14]. One of the recursive algorithms resulting from this connection is the following. Let $\alpha_0 = 6 - 4\sqrt{2}$ and

$$y_{n+1} = \frac{1 - (1 - y_n^4)^{1/4}}{1 + (1 - y_n^4)^{1/4}} \qquad y_0 = \sqrt{2} - 1$$

Then

$$\alpha_{n+1} = (1 + y_{n+1})^4 \alpha_n - 2^{2n+3} y_{n+1}(1 + y_{n+1} + y_{n+1}^2)$$

approaches $1/\pi$ with an error smaller than $16 \cdot 4^{n+1} \exp(-2\pi \cdot 4^{n+1})$. The first approximation, α_1, has already an accuracy of 9 digits, and α_2 has 40 correct digits. The number of correct digits of α_n is greater than $2 \cdot 4^n$; asymptotically it *quadruples* with each iteration and exceeds 1 billion after 15 steps.

There is even a *quintic* algorithm which multiplies the number of correct digits by a factor of 5 with each step [BBB 89].

How random are the first billion digits of π? Gregory and David Chudnovsky of Columbia University found that the digits of π are more random than the strings of digits produced by standard pseudorandom number generators, which, as finite-state algorithms, are, of course, ultimately periodic. Specifically, when used to generate Brownian motions, the digits of π generate random walks that seem to satisfy the iterated-logarithm law (in contrast to the usual pseudorandom number generators).

Even so, there is no mathematical proof that π is a normal number, in which all groups of digits occur with the same asymptotic probability. In fact, given

that π is determined by such compact and rapidly converging fomulas as equation 13, experts suspect the π is *not* normal in some appropriate base.

Bushes and Flowers from Iterations

Iteration of simple rules is one of the more potent prescriptions for generating not only mathematical fractals but interesting biological shapes too [Pru 87]. A preferred paradigm for these applications is the so-called turtle algorithm for producing line drawings:

A *state* of the "painting turtle" is defined as a triplet (x, y, α), where the Cartesian coordinates (x, y) represent the turtle's *position* and the angle α, called the turtle's *heading*, is interpreted as the direction in which the turtle is facing.

$d = 2 \quad \delta = 90°$

F-F-F-F

f→fffff

F→F-f+FF-F-FF-Ff-FF+f-FF+F+FF+Ff+FFF
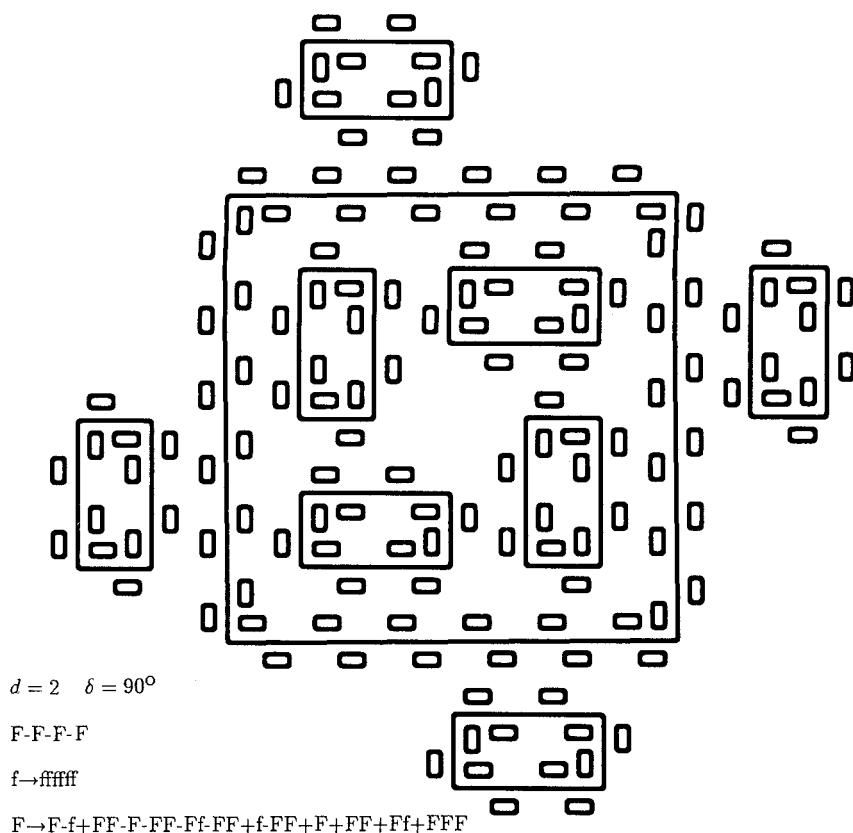
Figure 13   Nested "Islands and Lakes" drawn by the turtle algorithm [Man 83].

Given the *step size d* and the *angle increment* δ, the turtle can respond to commands represented by the following symbols:

F Move forward a step of length d. This motion changes the state of the turtle to (x', y', α) where $x' = x + d \cos \alpha$ and $y' = y + d \sin \alpha$. The turtle also leaves a trail, drawing a line segment between points (x, y) and (x', y').

f Move forward a step of length d without drawing a line.

+ Turn right (clockwise) by angle δ. The next state of the turtle is $(x, y, \alpha + \delta)$.

− Turn left by angle δ. The next state of the turtle is $(x, y, \alpha - \delta)$.

Let v be a string of commands, (x_0, y_0, α_0) the initial state of the turtle, and d and δ fixed parameters. The picture (set of lines) drawn by the turtle responding to the string v is called the *turtle interpretation* of v. Figure 13 shows as an example

$$d = 4 \qquad \delta = 22.5^o$$
$$\text{F} \rightarrow \text{FF+[+F-F-F]-[-F+F+F]}$$

Figure 14   "Breezy Bush" (watch its self-similarity) [Pru 97].

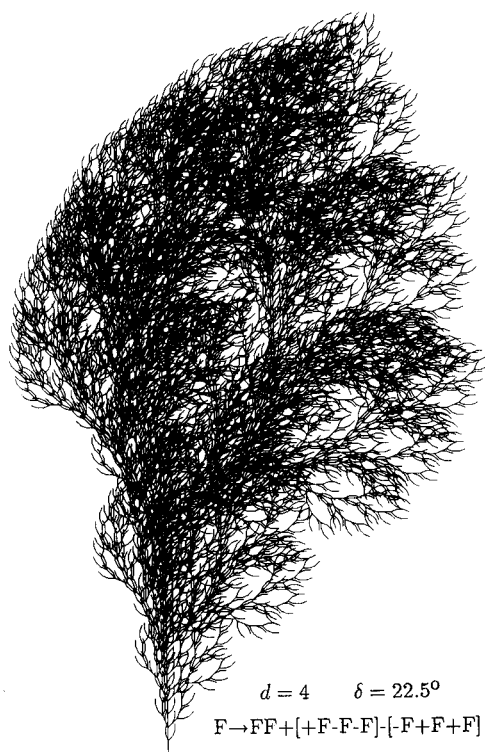the drawing of the fractal "Islands and Lakes," starting with the unit square, given by the command string F — F — F — F. According to these rules, the turtle interprets a character string as a sequence of line segments, connected "head to tail."

In a further elaboration, Lindenmayer introduced a notation for representing graph-theoretic trees using strings with *brackets* [Lin 68]. The motivation was to formally describe branching structures found in many plants, from algae to trees. An extension of the turtle interpretations to bracketed strings uses two additional symbols interpreted by the turtle:

[Remember the current turtle state (x, y, α) for later retrieval.

]   Recall the turtle state at the *corresponding* "open" bracket ([) and continue executing the instructions to the right of the "closed" bracket (]). Note that brackets can be nested.

An example of a bracketed string and its turtle interpretation of a bush in a breeze is shown in Figure 14.

Color Plate 7A through F shows six plantlike structures generated by Prusinkiewicz in this manner [Pru 87]. By adding color to the constructions, Prusinkiewicz has obtained breathtakingly beautiful and realistic-looking flowers and bushes, as evidenced in the color plates, which are reproduced here with his kind permission.

In another approach to image coding, called *iterated function systems*, Barnsley has created sunflowers, ferns, and forests from an astonishingly small set of parameters (see pages 28–30 in Chapter 1).

12

C H A P T E R

A Self-Similar Sequence, the Logistic Parabola, and Symbolic Dynamics

I work in statistical mechanics, but I am
not interested in getting to the moon.
—MARSTON MORSE

In this chapter we shall delve deeper into the self-similarities engendered by iteration. We shall focus our attention on the so-called *logistic parabola*, a simple quadratic equation that models the waxing and waning of warring species and restrained growth processes—restrained by a lack of "logistical" resources and supplies. This simple law has found widespread application in many fields. Its iteration gives rise to numerous universal features and self-similarities. We shall attempt to illuminate such signal attributes of the logistic parabola as *stable* and *unstable orbits, deterministic chaos, tangent bifurcations, intermittency,* the *hierarchy of orbits,* the *bifurcation of chaos bands,* and *invariant distributions.* We shall also touch upon some of the mathematical tools that exploit the self-similarity inherent in the quadratic map and explore some noteworthy transformations that shed a lot of light on this and other iterations.

One of the predictions of the logistic parabola, borne out by observation of many natural phenomena, is the occurrence of periodic cycles, especially those of period lengths 2, 4, 8, 16, 32, and so on. This is the famous *period-doubling* "scenario," which is born of self-similarity and begets, in the end, total *chaos,* albeit *deterministic* chaos. We shall attempt to bridge the seemingly impossible gulf between the self-similarity of the binary integers and period doubling in the logistic parabola, forging a strong bond, dubbed *symbolic dynamics,* between these two fundamental phenomena. But, as we shall also see, deterministic chaos is intimately linked to a simple operation on the digits of (binary) numbers: an incessant left shift until their totally unpredictable tails are exposed to full view.

After having reveled long enough in the *real* world of the logistic parabola, we shall follow Mandelbrot in an imaginative leap in the *imaginary* direction into the complex plane, where much becomes plain that was obscure on the real line. In other words, we shall *complexify* to *simplify*. In the process, we will discover the *Mandelbrot set* and its intricate self-similar designs as an added reward.

We begin our excursion into complexity with a recursive exercise in the discrete world of the real integers.

Self-Similarity from the Integers

The self-similar properties that can be squeezed from the integers are far from exhausted by the Fibonacci numbers and Pascal's triangle (see Chapter 17, on cellular automata). Consider the sequence of the nonnegative integers 0, 1, 2, 3, 4, 5, 6, 7, . . . in binary notation:

$$0, \quad 1, \quad 10, \quad 11, \quad 100, \quad 101, \quad 110, \quad 111, \ldots$$

and take the "digital root" (i.e., the sum of the digits modulo 2) of each binary number. This yields the sequence

$$0, \quad 1, \quad 1, \quad 0, \quad 1, \quad 0, \quad 0, \quad 1, \ldots$$

which is called the *Morse-Thue* (MT) *sequence*, in honor of the Norwegian mathematician Axel Thue (1863–1922), who introduced it in 1906 as an example of an aperiodic, recursively computable string of symbols, and after Marston Morse of Princeton (1892–1977), who discovered its significance in the symbolic dynamics in the phase-space description of certain nonlinear physical systems [Thu 06, Mor 21].

Interestingly, the MT sequence can also be generated by iterating the mapping $0 \rightarrow 01$ and $1 \rightarrow 10$, that is, the mapping in which each term is followed by its complement. Starting with a single 0, we get the following successive "generations":

$$0$$
$$0 \quad 1$$
$$0 \quad 1 \quad 1 \quad 0$$
$$0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1$$

and so on. A sequence generated in this manner is called a self-generating sequence [Slo 73].

Alternatively, each generation is obtained from the preceding one by appending its complement:

0

0   1

0   1   1   0

0   1   1   0   1   0   0   1

and so on. This is simply a consequence of the fact that the mapping $0 \to 01$ and $1 \to 10$ immediately gives rise to the mapping $01 \to 0110$ and $10 \to 1001$, and so on, where each higher-order map mimics the original generating rule "copy and append the complement." In other words, the original mapping rule is *inherited* by all successive generations. This kind of inheritance is an important consequence of iterated mappings and often leads to self-similar structures. Such generating processes are also called *inflation*, a term which (in its noneconomic and noncosmological sense) is associated with Penrose tilings and their fascinating scaling properties [GS 87].

The infinite sequence obtained from the iterated map $0 \to 01$ and $1 \to 10$ is invariant under this mapping; inflation leaves it untouched. The MT sequence is in fact *self-similar*: retaining only every other term of the infinite sequence (indicated by underlining), beginning with the first term, reproduces the sequence:

$$\underline{0} \quad 1 \quad \underline{1} \quad 0 \quad \underline{1} \quad 0 \quad \underline{0} \quad 1 \ldots$$

Similarly, retaining every other *pair* also reproduces the sequence:

$$\underline{01} \quad 1 \quad 0 \quad \underline{10} \quad 0 \quad 1 \ldots$$

as does the "renaming" of each pair, quadruplet, octet, and so on, by its left most digit (i.e., regenerating it from its "amputated" first digit). This "skipping" process is equivalent to what has been called *deflation* in tiling or *block renaming* in renormalization theories. These schemes follow simply from the inverse, $01 \to 0$ and $10 \to 1$, of the original mapping $0 \to 10$ and $1 \to 10$. Naturally, if inflation reproduces a given infinite sequence, so does the corresponding deflation. The fact that proper subsets can be equivalent to the entire set is, of course, a well-known property of infinite sets.

The self-similarity of the MT sequence is very easy to understand. Retaining only every other term of the infinite sequence is equivalent to multiplying the original numbers in the underlying integer sequence by 2. Since, in the binary representation, this means a left shift of the digits, the digital roots are not changed—by definition this *is* the MT sequence. (If we retain every other term beginning with the *second* term, the MT sequence, for similar reasons, turns into its own complement.)

Of course, there is nothing magic about the number 2 and the binary number system. The successive integers written in ternary notation:

$$0, \quad 1, \quad 2, \quad 10, \quad 11, \quad 12, \quad 20, \quad 21, \ldots$$

have digital roots (sums of digits modulo 3) that form a self-similar sequence with a similarity factor of 3:

$$p_k = 0, \quad 1, \quad 2, \quad 1, \quad 2, \quad 0, \quad 2, \quad 0, \ldots$$

Indeed, p_{3k} equals p_k because the ternary representation of $3k$ is the same as that of k except for a left shift. What iteration generates the p_k? (And what are the self-similarities of the sequences $\{p_{3k+1}\}$ and $\{p_{3k+2}\}$?)

Another interesting property of the MT sequence is its *aperiodicity*. We leave the proof, which is not too difficult, to the reader. Although aperiodic, the sequence is anything but random; it in fact has strong short-range and long-range structures. For example, there can never be more than two adjacent terms that are identical. And of course, any terms whose indices (beginning with the index 0 for the initial 0 in the sequence) differ by a factor of 2^n are identical.

This strong internal structure is reflected in the Fourier spectrum of the sequence (see Figure 1), which shows pronounced peaks, in spite of the fact that the sequence is aperiodic. The reader may want to show that the two strongest peaks occur at the frequencies one-third and two-thirds of the "sampling frequency."

A particularly convenient starting point for deriving the Fourier transform of the MT sequence in the ± 1 alphabet ($m_k = 1, -1, -1, 1, 1, -1, \ldots$) is its generating function

$$H(z) := \sum_{k=0}^{\infty} m_k z^k$$

The invariance of the MT sequence m_k under the substitution $1 \rightarrow 1, -1$ and $-1 \rightarrow -1, 1$ implies the functional equation

$$H(z) = (1 - z)H(z^2)$$

which in turn yields the generating function

$$H(z) = (1 - z)(1 - z^2)(1 - z^4)(1 - z^8) \cdots$$

Except for replacing z by z^{-1}, the generating function is identical with the "z transform," which is commonly used by electrical engineers to describe the transfer functions of digital filters.
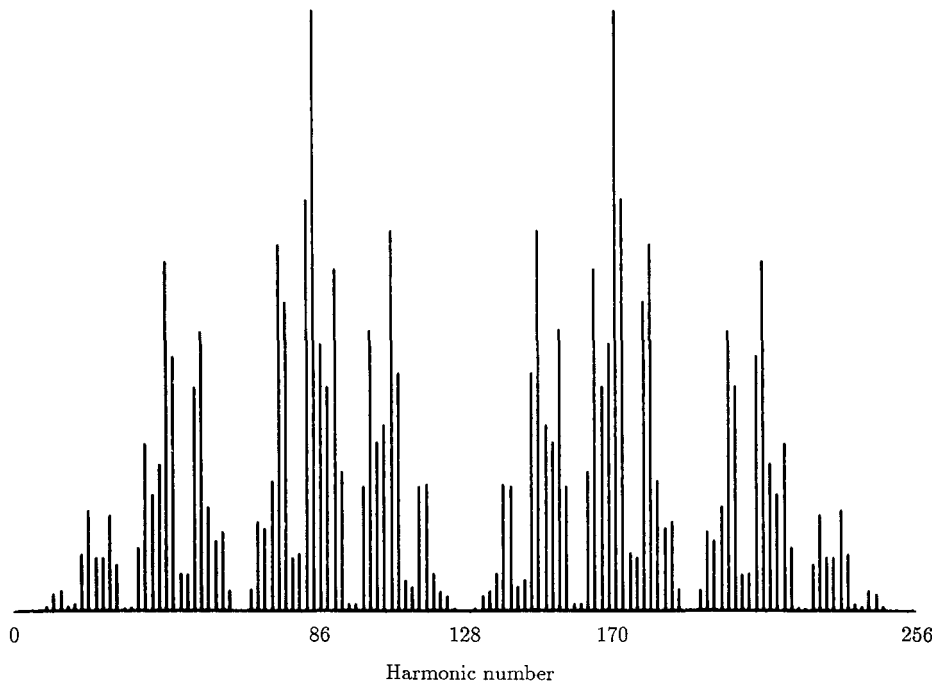
Figure 1   Fourier amplitude spectrum of the Morse-Thue sequence (first 256 terms repeated periodically) [Schr 90].

By setting $z = \exp(i\omega)$, where ω is the radian frequency, the generating function yields the Fourier transform

$$M(\omega) = \prod_{k=0}^{\infty} [1 - \exp(i\omega 2^n)]$$

which obeys the scaling law $M(\omega) = [1 - \exp(i\omega)]M(2\omega)$. This scaling law, together with the symmetries $M(-\omega) = M^*(\omega)$ and $M(\omega + 2\pi) = M(\omega)$, determines the self-similar structure of the spectrum.

In physics, the MT sequence occurred originally in the symbolic dynamics for certain nonlinear dynamic systems. Marston Morse proved that the trajectories of dynamic systems whose phase spaces have a negative curvature everywhere can be completely characterized by a *discrete* sequence of 0s and 1s—a stunning discovery. This means that some complicated curve in $\mathbb{R}^n$, which, after all, represents an uncountably infinite set in a high-dimensional space, can be mapped into a discrete binary sequence! With the help of the MT sequence, Morse also proved the existence (under certain rules) of infinitely long chess games.

In the following section we shall study a particularly simple and instructive case of symbolic dynamics and its relation to the Morse-Thue sequence.

The Logistic Parabola and Period Doubling

Suppose in an ecological, economic, or other growth process the measure x_{n+1} of the next generation (the number of animals, for example) is a linear function of the present measure x_n:

$$x_{n+1} = rx_n$$

where $r > 0$ is the *growth parameter*. If unchecked, the growth will follow a geometric ("exponential") law:

$$x_n = r^n x_0$$

which for $r > 1$ will tend to infinity.

But growth is often limited by limited resources. In other words, the larger x_n, the smaller the growth factor r. The simplest way to model the decline in the growth factor is to replace r by $r(1 - x_n)$, so that, as x_n approaches some limit (1, in our case), the growth factor goes to 0. Thus, we get the growth law

$$x_{n+1} = f(x_n) = r(1 - x_n)x_n \tag{1}$$

which is called the *quadratic map* or, because of its use in logistics and its parabolic shape, the *logistic parabola* (see Figure 2).
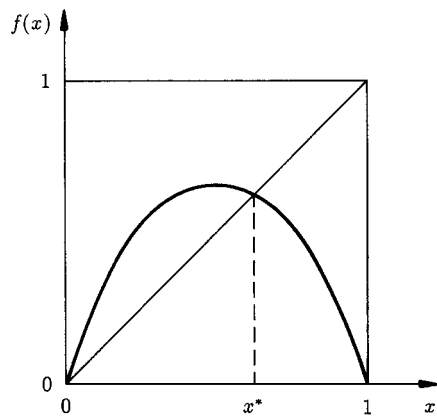
Figure 2   Quadratic map, also known as the *logistic parabola*. Note fixed point at $x = x^*$.

The logistic equation was introduced in 1845 by the Belgian sociologist and mathematician Pierre-François Verhulst (1804–1849) to model the growth of populations limited by finite resources [Ver 1845]. (The designation *logistic*, however, did not come into general use until 1875. It is derived from the French *logistique*, referring to the *lodgment* of troops.)

In its original form, the logistic equation was written as $x(t) = K/[1 + \exp(a - bt)]$, where x is the population whose growth pattern is being studied as a function of time t. The constants a and b set the origin and the scale of the time variable. Depending on these constants, the initial growth of x is approximately exponential, the growth rate reaching a maximum for $t = a/b$ and then tapering off to zero. The constant K determines the asymptotic value of x.

In another form, the Verhulst equation for the growth rate is $dx/dt = rx(K - x)/K$, which for $x \ll K$ leads to an exponential growth of x. But as x approaches K, the growth rate drops down to zero. Equation 1 is a recursive form of this equation in which time t has been replaced by the discrete variable n. The most important attribute of the Verhulst equation and its corollaries is their *nonlinearity*, which allows the modeling of nonlinearities and their consequences, such as chaotic dynamics, in many fields.

The quadratic map (equation 1) has two fixed points: $x = 0$ and, for $r > 1$, $x = x^* = 1 - 1/r$ (look again at Figure 2). The derivative of this map is

$$f'(x) = r(1 - 2x)$$

which equals r for $x = 0$ and $2 - r$ for the other fixed point, $x^* = 1 - 1/r$. Fixed points are stable as long as $|f'| < 1$. Thus the fixed point $x = 0$ is stable for $r < 1$. The fixed point $x = 1 - 1/r$ exists and is stable in the range $1 < r < 3$, because $|f'(x = 1 - 1/r)| < 1$ there.

In fact, for $r = 2$, $f'(x) = 0$ at the fixed point $x^* = 1 - 1/r = \frac{1}{2}$. Such fixed points are said to be *superstable*, because convergence to the fixed point is very rapid, as can be observed on any pocket calculator. In general, superstable *orbits* occur whenever $x = \frac{1}{2}$, for which $f'(x) = 0$, is a member of the orbit. (*Orbit* is the technical term for a succession of iterates x_n.) The parameter values for the superstable orbits of period length 2^k are called R_k; $r = R_0$ gives the superstable fixed point (period length 1).

For $r = 3$, the slope of equation 1 is -1 at the fixed point $x = \frac{2}{3}$. This fixed point is "indifferent," meaning that nearby values are neither attracted nor repelled. What happens for $r > 3$? The fixed point becomes unstable, splitting or *bifurcating* into an orbit of period length 2: $x_0, x_1, x_2 = x_0$ (see Figure 3). For example for $r = R_1 = 3.2360679775$, there is a stable (in fact, superstable) orbit of period length 2: $0.5 \rightarrow 0.8090169943 \ldots \rightarrow 0.5$, and so on (see Figure 4).

The value of R_1 is obtained from setting $f(f(0.5)) = 0.5$ as the solution of the cubic equation $R_1^3 = 4R_1^2 - 8$, which happens to have a *quadratic* irrational solution related to the golden mean $\gamma = 0.618 \ldots$: $R_1 = 2/\gamma = \sqrt{5} + 1$.
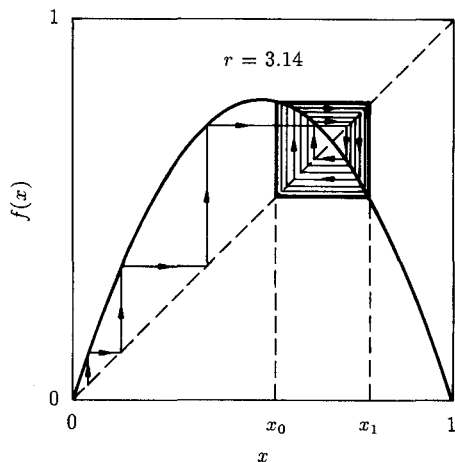
Figure 3   Fixed point turned unstable, leading to orbit of period length 2.

An orbit having a period length of 2, or a 2-orbit, for short, means, of course, that $f(f(x))$—abbreviated $f^{(2)}(x)$—has a fixed point (and because $f^{(2)\prime}(x) = 0$, the orbit is superstable). This is indeed the case (see Figure 4B). In fact, $f^{(2)}(x)$ has *two* fixed points, both superstable, at $x_0 = 0.5$ and $x_1 = 0.809 \ldots$.

If r is increased further, then these two fixed points of $f^{(2)}(x)$ will in turn become unstable. Indeed, they will become unstable at precisely the same value of r. Is this a coincidence? No, because, according to the chain rule of differentiating.

$$\frac{d}{dx} f(f(x)) \bigg|_{x=x_0} = f'(f(x)) \bigg|_{x=x_0} \cdot f'(x) \bigg|_{x=x_0}$$

or, with $f(x_0) = x_1$,

$$\frac{d}{dx} f(f(x_0)) = f'(x_1) \cdot f'(x_0)$$

Hence:

$$f^{(2)\prime}(x_0) = f^{(2)\prime}(x_1)$$

As a consequence of this equality, if x_0 becomes unstable because $\left| f^{(2)\prime}(x_0) \right| > 1$, so does x_1 at precisely the same value of the parameter r. This means that both fixed points of $f^{(2)}(x)$ will bifurcate at the same r value, leading to an orbit of period 4. In other words, now $f^{(2)}(f^{(2)}(x)) = f^{(4)}(x) := f(f(f(f(x))))$ will have a fixed point—in fact, four fixed points. For $r = R_2 = 3.498561699 \ldots$, the four fixed points of $f^{(4)}(x)$ are $x_0 = 0.500$, $x_1 = 0.874 \ldots$, $x_2 = 0.383 \ldots$, and
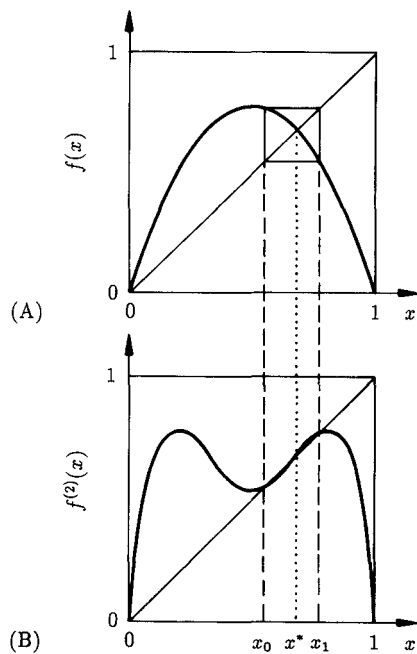
Figure 4 (A) The quadratic map for an orbit of period length 2. (B) The once-iterated map $f^{(2)}(x)$ has two stable fixed points, x_0 and x_1.

$x_3 = 0.827 \ldots$. These also form a superstable orbit of period length 4 of $f(x)$: $x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_0$, and so on.

Again, because of the chain rule of differentiation, the four derivatives are the same at all four points of the orbit. Thus if, for a given value of r, the magnitude of one of the derivatives exceeds 1, then the magnitude of all four derivatives will. Hence, all four iterates will bifurcate at the same value of r, leading to an orbit of period 8. This bifurcation scenario will repeat again and again as the growth parameter r is increased, yielding orbits of period length 16, 32, 64, and so on *ad infinitum*, ending up in a "chaotic" orbit of infinite period length for $r = r_\infty = 3.5699 \ldots$.

These period-doubling bifurcations are also called *pitchfork bifurcations*, because of the resemblance to a pitchfork when the values of the iterates are plotted as a function of the parameter (see Figure 5). Two prongs of the fork are the new iterates after bifurcation, and the central prong (shown as a dashed line in Figure 5) is the old (now unstable) iterate, which has turned from attractor to *repellor*.

Period doubling is a very common phenomenon, encountered in a wide variety of physical, ecological, and economic systems. Think of predators and
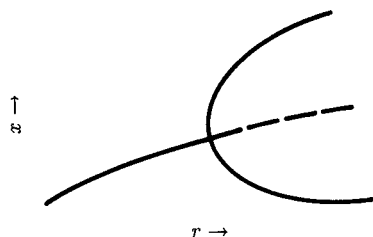
Figure 5 Pitchfork bifurcation: outer "prongs" are the new iterates after bifurcation, and dashed "prong" is the old iterate, now unstable.

their prey, such as foxes and rabbits. For sufficiently low fox reproduction rates, corresponding to $r < 3$ in equation 1, there may be enough rabbits to satiate the foxes' appetites, enabling a stable equilibrium between the number of foxes and rabbits to ensue. But if the fox reproduction rate is increased above a certain limit, corresponding to $r > 3$, the foxes will devour so many rabbits that, in the next season, there will not be enough rabbits to go around, so that the fox population will decrease. This will give the rabbits a chance to recover and become more plentiful again, allowing the number of foxes to increase too, giving rise to a two-season cycle.

At which values r_n of the growth parameter r do these bifurcations, from period length 2^{n-1} to 2^n, take place? And for which values R_n do we get superstable orbits of equation 1 with period length 2^n? What happens to the iterates x_0, $x_1, \ldots, x_{2^n-1}$ as n goes to infinity? To answer these questions we have to exploit the self-similarities that *must* be hiding somewhere in the iterated quadratic map.

Self-Similarity in the Logistic Parabola

Let us consider the superstable orbits of equation 1 with period lengths $P = 1$, 2, 4, 8, and so on. The parameter values $r = R_n$ that give superstable orbits of period length 2^n are much better defined, both experimentally and theoretically, than the points of bifurcation, $r = r_n$. The fast convergence to the final orbit gives better numerical estimates, and one always knows one member of the orbit a priori: $x_0 = 0.5$. By contrast, numerical determination of r_n, a bifurcation value, is somewhat trickier.

The period-doubling process is characterized by self-similarities that facilitate its analysis. To demonstrate one of these self-similarities, compare $f(x)$ for the parameter value $r = R_1$ for the superstable orbit with period length $P = 2$ (see Figure 6A) with the function $f^{(2)}(x)$ for $r = R_2$ for the superstable orbit with $P = 4$ (Figure 6B). The resemblance between the dashed square and its contents in
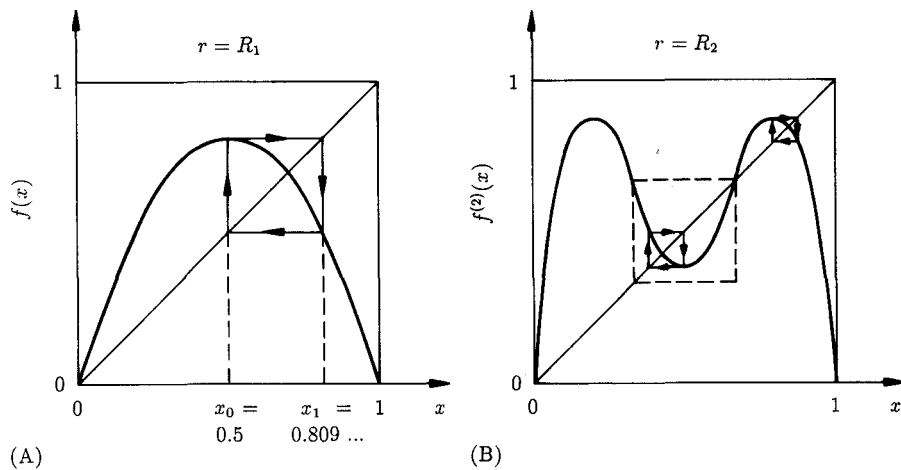
Figure 6   (A) The quadratic map $f(x)$ for the superstable orbit of period length 2. (B) The once-iterated map $f^{(2)}(x)$ for period length 4. Note the similarity between the contents of the small dashed square and part A. This self-similarity is characteristic for period doubling and facilitates its analysis [Schu 84].

Figure 6B and the large square and *its* contents in Figure 6A is striking. The discrepancy between the parabola in Figure 6A and the fourth-order curve inside the dashed square in Figure 6B is in fact quite small, as can be seen in Figure 7, which shows both the parabola (the solid curve) and the rescaled and inverted fourth-order curve (dashed curve).

Figure 7   The difference between the quadratic map (solid curve) and the rescaled iterated quadratic map (dashed curve).

The rescaling factor for this transition from period length 2 to period length 4 is easily calculated to equal $-(2 + 2/R_1) = -(2 + \gamma) = -2.618\ldots$, wh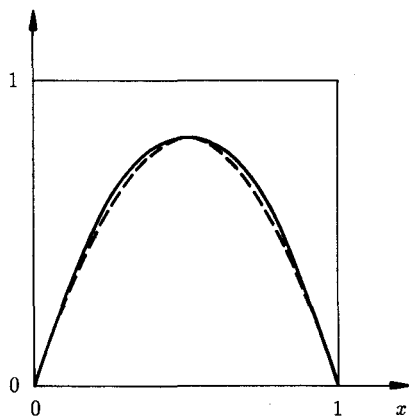ere γ is the golden mean. (The minus signs account for the fact that period doubling engenders an upside-down orientation.) In going from the superstable orbit of period length $P = 4$ ($r = R_2$) to that for $P = 8$ ($r = R_2$), almost precisely the same "scenario" is repeated for $f^{(2)}(x)$ and $f^{(4)}(x)$ as that shown for $f(x)$ and $f^{(2)}(x)$ in Figure 6, only the scaling factor is slightly different from -2.618. In fact, the same scenario (a popular buzzword in this context) is repeated every time the parameter value is changed from $r = R_n$ to $r = R_{n+1}$. As $n \to \infty$, the scaling factor quickly converges on its asymptotic value of $-2.5029\ldots$, which is not very far from its initial value of $-2.618\ldots$. In the limit, the initial parabola of the quadratic map becomes a transcendental function, given by an infinite power series $g(x)$ originally derived by Mitchell Feigenbaum:

$$g(x) \approx 1 - 1.52763x^2 + 0.104815x^4 - 0.0267057x^6 + \cdots$$

Here the x coordinate has been shifted so that the maximum of $g(x)$ is at $x = 0$, instead of $x = 0.5$, and has a value of 1. The function $g(x)$ is the *fixed-point function* of the period-doubling transformation for quadratic maps. It obeys the scaling law $g(x) = \alpha g(g(x/\alpha))$, which also determines $\alpha = 1/g(1)$. The derivation of $g(x)$ as a universal function for *all* maps with a quadratic maximum by Feigenbaum, via a renormalization theory, is instructive but not exactly easy [Fei 79].

Numerically, the scaling parameter α can be obtained from any of the numerous self-similarities of the iterates $x_m^{(n)}$ generated by the quadratic map. A particularly attractive method is to calculate the value of the iterate $x_{P/2}^{(n)}$ at the half period for a superstable orbit of period length $P = 2^n$, starting with $x_0 = 0.5$. For the parameter value $r = R_{n-1}$, $x_{P/2}^{(n-1)} = x_0$; but for $r = R_n$, $x_{P/2}^{(n)}$ misses x_0 by a small amount, the difference $|x_{P/2}^{(n)} - x_0|$ scaling asymptotically with α as n is increased to $n + 1$. More precisely,

$$\alpha_n := \frac{x_{P/2}^{(n)} - x_0}{x_{P/2}^{(n+1)} - x_0} \to \alpha \qquad \text{for } n \to \infty$$

With a programmable calculator, one first determines R_n and R_{n+1} (by adjusting R until $x_P = x_0 = 0.5$, for $P = 2^n$ and $P = 2^{n+1}$) and then reads out the value of $x_{P/2}$. In this manner one quickly obtains $\alpha_7 \approx -2.502905$, with a relative deviation from α of about 10^{-6}.

The Scaling of the Growth Parameter

We have just learned that period doubling is asymptotically self-similar, with a scaling factor for the variable x equal to $-2.5029\ldots$. How do the parameter values r—say, those for the superstable orbits R_n—scale? Numerical evidence

suggests that the differences $R_{n+1} - R_n$ become smaller and smaller according to the following geometric law:

$$R_{n+1} - R_n \approx \frac{R_n - R_{n-1}}{\delta} \qquad \text{for } n \to \infty$$

where δ is a *universal* constant, the famous (and probably transcendental) Feigenbaum constant (originally found by S. Grossmann and S. Thomae [GT 77]). This magic number has earned the epithet "universal" because it applies, as Feigenbaum has shown, to many different nonlinear maps, independent of the details of the mapping, as long as the absolute maximum of the mapping is quadratic. The convergence of $\delta_n := (R_n - R_{n-1})/(R_{n+1} - R_n)$ to δ is very rapid: $\delta_1 \approx 4.7$, $\delta_2 \approx 4.68, \ldots, \delta_6 \approx 4.66918$. The asymptotic value is

$$\delta = 4.6692016091029\ldots$$

and the accumulation point of the growth parameter for the period doublings is

$$R_\infty = 3.5699456\ldots$$

With these two values and another constant, one has $R_n \approx R_\infty - 1.542\delta^{-n}$.

The two scaling parameters α and δ are related to each other. A simplified theory yields $\delta \approx \alpha^2 + \alpha + 1 \approx 4.76$ [Fei 79].

At the critical value of the growth parameter $r = R_\infty$ the period has become infinite. In other words, the orbit is now *aperiodic*, comprising a point set of infinitely many values of x that never repeat. However, other x values are attracted to this point set, which is in fact a Cantor set (see Figure 8). Note the approximate
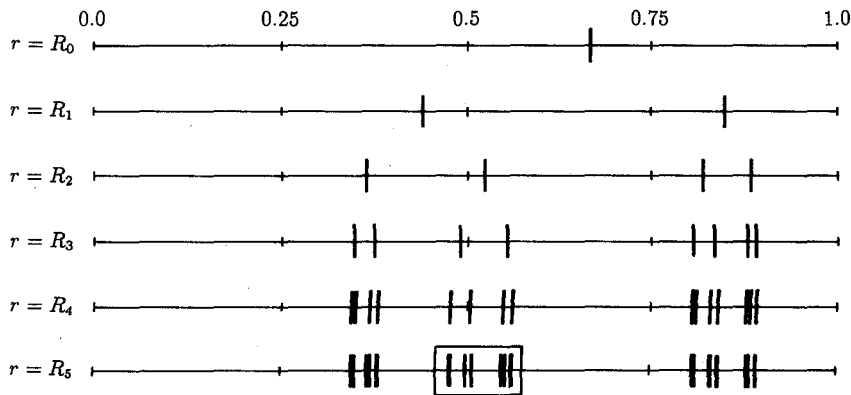

Figure 8   Self-similar Cantor set of iterates for period doubling. The iterates in the small box for $r = R_5$ are a scaled-down version of all the iterates for $r = R_3$.

self-similarity of this point set: the left half of the bottom line is the mirror image of the line above it compressed by a factor of about 2.5, and the right half of the bottom line is the line above it compressed by a factor of 2.5^2. The Hausdorff dimension $D = 0.538 \ldots$ of this set was derived analytically and numerically by P. Grassberger [Gra 81]. This and similar attractors in higher dimensions have been called *strange*, although once one knows about Cantor sets they are really not so strange after all.

Assuming that the limit set is strictly self-similar, we can use the well-known formula (see Appendix A) for the Hausdorff dimension D of a self-similar Cantor set with two different remainders, s_1 and s_2,

$$s_1^D + s_2^D = 1$$

to calculate a good approximation to the Hausdorff dimension of the strange attractor of the logistic parabola at the period-doubling accumulation point. With $s_1 = 1/2.5 = 0.4$ and $s_2 = s_1^2$, and setting $z = 0.4^D$, we obtain from

$$z + z^2 = 1$$

$z = \gamma \approx 0.618$ and $D \approx \log \gamma/\log 0.4 \approx 0.525$, a surprisingly good approximation to the more precise value $0.538 \ldots$. As so often, self-similarity may be only approximate, but ignoring the lack of exact scaling still gives good results that can be bettered only by more involved computation.

The Fourier spectrum of the periodic sequence x_m, too, shows pronounced self-similarity (see Figure 9). Let a_k^n be the Fourier coefficients of the x_m for a period length $P = 2^n$. In going from an orbit of period length $P = 2^n$ by a period-doubling bifurcation to an orbit of length 2^{n+1}, the new Fourier coefficients with
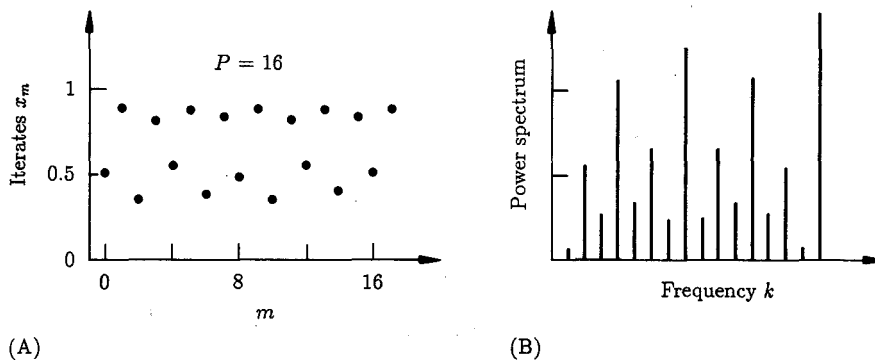

(A)   (B)

Figure 9   (A) Iterates x_m for the quadratic map at period length 16. (B) Fourier power spectrum (on a logarithmic scale) of the iterates x_m [Schu 84].

an even index a_{2k}^{n+1} are approximately equal to the old Fourier coefficients: $a_{2k}^{n+1} \approx a_k$ (because $x_{n+P} \approx x_n$). The odd-index coefficients a_{2k+1}^{n+1}, which describe the subharmonics that appear in the spectrum as a result of the period doubling, are determined by the difference $x_{n+P} - x_n$. A detailed analysis shows that the squared magnitudes of the Fourier coefficients, $|a_{2k+1}^{n+1}|^2$, are roughly equal to an adjacent component from the previous orbit scaled down by a factor of $8\alpha^4/(1 + \alpha^2) \approx 40$, corresponding to 16 decibels (dB) in logarithmic units [Fei 79]. (The number of decibels is, by definition, 20 $\log_{10}$ of a magnitude ratio or 10 $\log_{10}$ of a squared magnitude ratio, such as a spectral *power* ratio.)

An early confirmation of period-doubling bifurcations occurred in a hydrodynamic ("Rayleigh-Bénard") experiment by Libchaber and Maurer in which the Reynolds number played the role of the growth parameter r [LM 80]. In studying the forced nonlinear oscillations of bubbles in water, Lauterborn and Cramer found a similar behavior: the appearance of more and more subharmonics until the onset of chaos, called *cavitation noise* in this context [LC 81]. As a result of such experiments, the destructive mechanism of cavitation, a much dreaded source of failure in ship propellers, is now well understood.

Self-Similar Symbolic Dynamics

Instead of listing the sequence of iterates x_n themselves, it often suffices to state whether they fall to the left (L) or the right (R) or *on* the maximum or center (C) of the map. The sequence of symbols L, R, C is then called the *symbolic dynamics* for a given orbit. Thus, the superstable orbit of period 2 has the symbolic dynamics or "kneading sequence" *CRCRCR* . . . . Restricting the notation to a single period, we write simply *CR*.

It is not too difficult to show that the next superstable orbit, the one with period length 4, is obtained as follows. First one writes two periods of the orbit with period 2, *CRCR*, and then changes the second C to L if the number of R's to the left of it is odd. Otherwise the second C is changed to R. Thus, the superstable orbits have the following symbolic dynamics:

		Period 1:	C
Period 1	$\rightarrow$	Period 2:	$CC \rightarrow CR$
Period 2	$\rightarrow$	Period 4:	$CRCR \rightarrow CRLR$
Period 4	$\rightarrow$	Period 8:	$CRLRCRLR \rightarrow CRLRRRLR$

and so on. The orbit of period 8 is often more conveniently written as $CRLR^3LR$. In the same vein the superstable orbit of period 16 is written as $CRLR^3LRLRLR^3LR$.

This algorithm of counting the number of previous R's and checking whether it is even or odd is directly related to the fact that the slope of the quadratic map is *negative* for the right half of the map. Thus, each time the iterate x_n falls into the right half ($x_n > 0.5$), there is a sign change in how small differences in x_n are propagated, and an *odd* number of sign changes *is* a sign change (while an even number is not). This is one of the most important properties not only of the quadratic map but of all unimodal ("one-hump") maps. As a result these maps have a "universal" ordering of their symbolic dynamics as the growth parameter is changed.

More specifically, for $r = R_n$, that is, period length $P = 2^n$, the iterate $x_P^{(n)}$ equals x_0 by definition. In changing the growth parameter r from R_n to R_{n+1}, $x_P^{(n)} - x_0$ will be positive (negative) if $x_m^{(n)} - x_0$ was positive an even (odd) number of times for $m = 1, 2, \ldots, P - 1$. This is the reason for the aforementioned rule $C \to R$ (or L) for an odd (even) number of preceding R's.

These sequences are self-similar in the following sense. Retaining only every other symbol (starting with C) reproduces the sequence for the superstable orbit with half the period, except that L and R are interchanged. Thus, "pruning" the symbolic dynamics we derived for the orbit with period 16 results in $CLRL^3RL$, which is the complement of $CRLR^3LR$ that describes the orbit of period 8.

As in the Morse-Thue sequence, retaining every other term produces a similar, albeit complemented, sequence. Is there a closer connection between the Morse-Thue sequence and the symbolic dynamics of the superstable orbits? There is indeed. To see this, let us replace R by 1 and C and L by 0. With this notation, the superstable orbits of periods 1, 2, 4, and 8 are

$$0$$
$$0 \quad 1$$
$$0 \quad 1 \quad 0 \quad 1$$
$$0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1$$

which unfortunately does *not* look like the beginning of the Morse-Thue sequence. However, the running partial sums modulo 2 of these orbits (beginning at the left) do reproduce the Morse-Thue sequence:

$$0$$
$$0 \quad 1$$
$$0 \quad 1 \quad 1 \quad 0$$
$$0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1$$

and so on. Computing running partial sums modulo 2 is, of course, equivalent to keeping track of the number of preceding R's.

Conversely, the kneading sequences for unimodal maps in the binary notation, 01011101 ..., are obtained from the Morse-Thue sequence by taking sums (or differences) modulo 2 of adjacent elements.

This connection allows us to write down directly (without iteration) the kneading sequence of period 2^m for any m. The rule—I encourage readers to derive it for themselves—is simply this: the kth term in every such symbolic sequence is obtained by writing $k = 2^q \cdot j$, where j is odd. Then the parity of q determines the choice between L and R: for odd q the symbol is L; for even q it is R (and of course for $k = 0$ the symbol is C by definition). Thus, for example, the ninety-sixth symbol in the kneading sequence is L, because $96 = 3 \cdot 2^5$, and 5 is odd.

One general result of this rule is that all terms with an odd k are R (including the last term of each periodic orbit with $P = 2^n$). All terms whose k is a power of 2 ($k = 2^q$) alternate between L and R.

The irrational number constructed with the help of the Morse-Thue sequence interpreted as a binary fraction, $0.0110100110010110 \ldots = 0.4124 \ldots$, which may be called the *Morse-Thue constant*, is intimately related to the period-doubling bifurcation scenario and the Mandelbrot set (see pages 295–299).

Periodic Windows Enbedded in Chaos

Eyesight should learn from reason.
—JOHANNES KEPLER

Figure 10A shows the "behavior" of the iterated logistic parabola, that is, the values of its iterates x_n as the parameter r is increased from 3 to 4. There is a cascade of period-doubling bifurcations followed by chaos (the dense bands) interleaved with periodic "windows." In fact, the r values for periodic windows are dense, but one sees only a few in the treelike plot (Figure 10A), sometimes called a "Feigenbaum" ("figtree") plot. Most prominent is the period-3 window starting at $r = \sqrt{8} + 1$ (see Figure 10A). Once the period length 3 has been observed, we know from the work of Li and Yorke that all possible periods appear [LY 75].

Note that in the period-3 window the period doubling occurs again, leading to orbits of period length 6, 12, 24, and so on, and renewed chaos in which another period-3 window is embedded, and so forth *ad infinitum* in another self-similar cascade (see Figure 10B).

Interestingly, the chaos bands also show bifurcation, called *reverse bifurcation*, as we decrease r from its highest value $r = 4$. For r somewhat below 3.68, the single chaos band splits into two (see Figure 11), and for r near 3.6 the two chaos bands split into four, and so forth, mimicking the period-doubling bifurcations observed for increasing r. In fact, each chaos band is torn asunder by the surviving "ghost" of the corresponding period-doubling bifurcation.
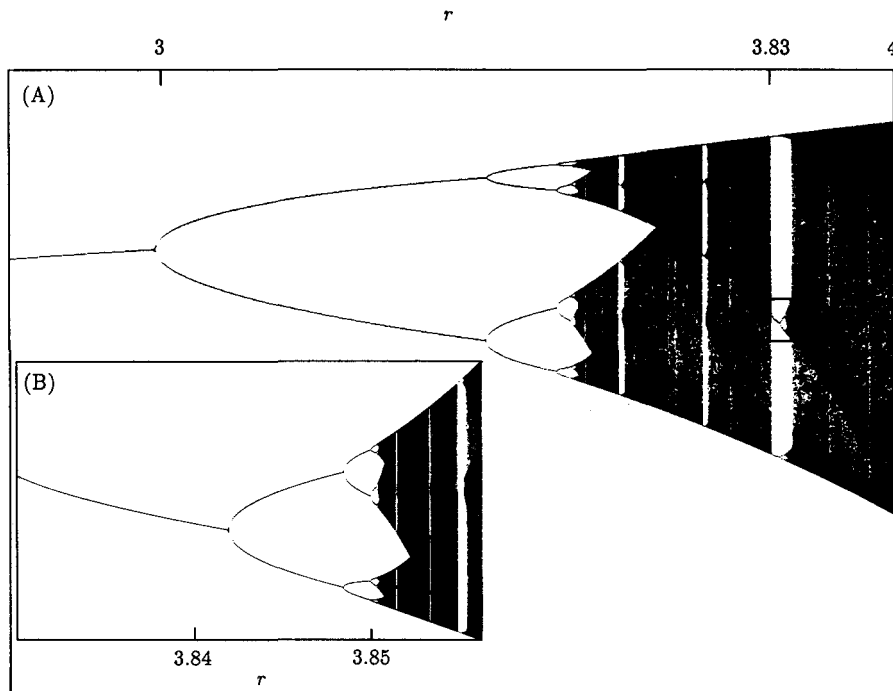
Figure 10   (A) Iterates of the quadratic map plotted against the growth parameters. The period-doubling bifurcation cascade is on the left, followed by chaos bands, on the right, in which the iterates are chaotic. The chaos bands are interspersed with "period windows" in which the iterates are periodic again. Most prominent is the window for period length 3 starting at $r \approx 3.83$. (B) An enlargement of the central portion of the period-3 window reveals another period-doubling cascade, followed by a second period-3 window, in which the period length is 9.

The reason for the bifurcation of chaos bands as the parameter r is decreased is easy enough to see and analogous to the bifurcation of periodic orbits. Consider the r value $\tilde{r}_1$, for which the third iterate x_3 of $x_0 = 0.5$ falls on the unstable fixed point $x^* = 1 - 1/\tilde{r}_1$. This yields the equation $\tilde{r}_1^3(4 - \tilde{r}_1) = 16$ for $\tilde{r}_1$, which yields $\tilde{r}_1 = 3.678573510 \ldots$. For r values slightly smaller than $\tilde{r}_1$, x_3 will fall just below x^* and x_4 slightly above, creating a gap around x^* into which no iterates can fall.

Similarly, for r values just below $\tilde{r}_2 = 3.5925721841 \ldots$, for which iterates of $x_0 = 0.5$ fall on the unstable period-2 orbit, bifurcation of the two chaos bands into four bands takes place. In fact, the dark sinusoidal contours visible in the Feigenbaum diagram (Figure 11), including upper and lower edges of chaos bands, are images of the stationary point $x = 0.5$. Chaos bands merge where these
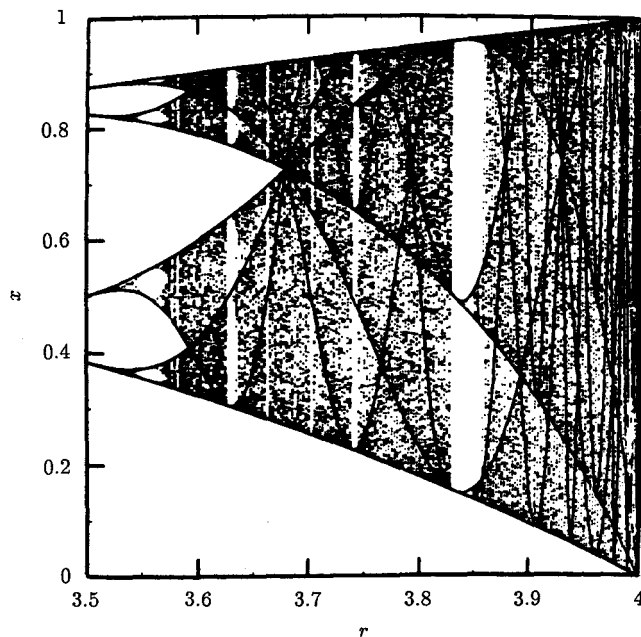
Figure 11   Reverse bifurcation of chaos bands as the growth parameter is decreased below 3.68 [JM 85].

contours cross each other, and periodic windows open where contours touch upper and lower edges [Lor 80, JM 85].

Interestingly, the parameter values $\tilde{r}_m$ at which 2^{m-1} chaos bands join 2^m chaos bands form a descending, asymptotically geometric progression with the same accumulation point, $r = 3.5699\ldots$, as that of the period-doubling bifurcations. And the scaling factor, too, is the same, namely, the Feigenbaum constant $\delta = 4.6692\ldots$.

The $\tilde{r}_m$ also correspond to the parameter values for the accumulation points of successive rows of the "Sharkovskii ordering" of orbits (see page 285), because the symbolic dynamics are the same. For example, the symbolic dynamics of the two chaos bands (starting with $x_0 = 0.5$), $CRLR^2(RL)^\infty$, are the same as those for the accumulation point of the orbits of period length $4 \cdot 3, 4 \cdot 5, 4 \cdot 9, \ldots$ (see pages 287−288).

The ordering with which the iterates x_n fall into the 2^m different chaos bands is also the same as the ordering of the iterates in a stable orbit of period length $P = 2^m$. For example, for both the period-4 orbit and the four chaos bands, the iterates, starting with the largest iterate x_1, are ordered as follows:

$x_1 > x_3 > x_4 > x_2$. This ordering is obtained from the ordering $x_1 > x_2$ for $P = 2$ by splitting each iterate x_i into the two iterates of x_i and x_{i+P} and inverting the order of every other pair of iterates. In this manner, the ordering for $P = 8$, $x_1 > x_5 > x_7 > x_3 > x_4 > x_8 > x_6 > x_2$, is easily derived from the ordering $x_1 > x_3 > x_4 > x_2$ for $P = 4$.

The ordering of the iterates x_n according to their values can also be obtained from a Gray code (which in turn is related to the Hilbert space-filling curve; see pages 10–13 in Chapter 1). For example, to deduce the correct order of the eight iterates x_n, $n = 1, 2, \ldots, 8$, interpret the three-digit Gray code (in which one digit at each step is changed, starting from the left) as ordinary binary numbers and add 1 to obtain the index n of x_n:

$$0 = 000 = x_1$$

$$1 = 100 = x_5$$

$$2 = 110 = x_7$$

$$3 = 010 = x_3$$

$$4 = 011 = x_4$$

$$5 = 111 = x_8$$

$$6 = 101 = x_6$$

$$7 = 001 = x_2$$

The close correspondence between period-doubling bifurcations and the reverse bifurcations of the chaos bands that we have glimpsed here is one of the many fascinating features of the quadratic map.

The Parenting of New Orbits

The algorithm described in pages 277–278 for constructing the symbolic dynamics of period-doubled superstable orbits can be considerably generalized. Thus the rule "append the string $CR \ldots$ to itself and change the second C to L (R) if the number of R's in the original string is odd (even)" applies not only to fundamental periods of length $P = 2^n$ but to any orbital length. In this manner the superstable orbit of length 3 with symbolic dynamics CRL is doubled via $CRLCRL$ to $CRLLRL$, which in turn is doubled to $CRL^2RLR^2L^2RL$ and so on to an infinite cascade of orbits of lengths $3 \cdot 2^n$.

More generally, one can derive orbits of lengths $k \cdot m^n$ from an orbit P of length k and an orbit Q of length m. For the orbit of length $k \cdot m$, one copies

the symbolic dynamics of P m times and replaces each of the $(m - 1)$ C's (except the first C) by one after another of the $(m - 1)$ symbols of Q, interchanging L and R of Q if the number of R's in P is odd. (The initial C in Q is ignored.) For example, the 2-orbit CR is tripled by means of the 3-orbit CRL by first copying the 2-orbit three times—$CRCRCR$—and then replacing the second and third C's by the complement of the second and third symbols of $C\underline{RL}$. This yields the orbit $CR\underline{LRR}R$ of period 6, distinct from the previously derived orbit CRL^2RL (The period-doubling algorithm described in the preceding paragraph is but a special case of this composition law with $Q = CR$.)

Which of the two orbits of period 6 just described "dominates" the other? By definition, this depends on the first symbol in which they differ. If the number of R's in the initial, equal portion of the strings is odd (even), then the orbit with L or C (R) as the first distinct symbol dominates the other. Thus, the orbit CRL^2RL dominates the orbit $CRLR^3$. The superstable r values of an orbit that dominates another orbit is the larger of the two superstable r values. Indeed, the approximate r values corresponding to CRL^2RL and $CRLR^3$ are 3.8445688 and 3.6275575, respectively.

A particularly attractive equivalent algorithm for "multiplying" two orbits of period lengths p_1 and p_2 to yield an orbit of length $p_1 p_2$ proceeds as follows. Write the symbolic dynamic of the orbit as a sequence of plus or minus signs, $+$ corresponding to either of the letters L or C and $-$ to the letter R. (This reflects the fact that the slope of unimodal maps is *negative* to the *right* of the maximum.) Thus, the superstable $p_1 = 3$ orbit CRL is written as $+ - +$. Next form the running product of these signs starting from the left. This transforms $+ - +$ into $+ - -$, which we shall call the σ *sequence* of the orbit. Similarly, the σ sequence of the superstable orbit CR with period length $p_2 = 2$ is $+ -$.

Using these σ sequences, period multiplication becomes a simple appending process. For example, period doubling of any orbit is realized by appending its σ sequence to itself with the opposite signs, corresponding to the period-doubling "operator" $+ -$, that is, the σ sequence of the period-2 orbit. Using this rule, the period-doubling cascade now looks as follows. Starting with the fixed point C, which has period length $p = 1$ and thus corresponds to a σ sequence consisting of a single sign, $\sigma_0 = +$, we have

$p = 1$	$+$	C
$p = 2$	$+ -$	CR
$p = 3$	$+ - - +$	$CRLR$
$p = 4$	$+ - - + - + + -$	$CRLR^3LR$

and so on. This iterative construction of the σ sequence of orbits with period lengths 2^n ("append the complement") leads, of course, to the previously discussed self-similar Morse-Thue sequence, which is generated by the same rule.

To recover the symbolic dynamics, one writes C followed by R's and L's, depending on whether the corresponding sign in the σ sequence is different from or the same as the preceding sign, as was just shown for $p = 2$, 3, and 4.

To triple the period length of an orbit, one appends its σ sequence two times to itself with opposite signs, corresponding to the σ sequence $+ - -$ of the period-3 orbit. The period-tripling cascade thus has the following σ sequences, starting with $\sigma_0 = +$:

$$
\begin{array}{lll}
p = 1 & + & C \\
p = 3 & + - - & CRL \\
p = 9 & + - - - + + - + + & CRL^2RLR^2L
\end{array}
$$

and so on. The resulting infinite self-similar σ sequence with $\sigma_{3k} = \sigma_k$ is a generalization of the Morse-Thue sequence.

The kth symbol s_k, $k > 0$, in the period-tripled symbolic dynamics is given by writing the index $k = 3^m \cdot q$, where $q = 1$ or 2 modulo 3: $s_k = L$ if m and $\langle q \rangle_3 := q$ modulo 3 are both even or both odd. If the parities of m and $\langle q \rangle_3$ are different, then $s_k = R$. Writing -1 for R and $+1$ for L, we have $s_k = (-1)^{m + \langle q \rangle_3}$. Thus, with $405 = 3^4 \cdot 5$, s_k equals $(-1)^{4+2} = 1 = L$.

The "second harmonic" of the period-3 orbit CRL, with the σ sequence $+ - -$, has the σ sequence $+ - - - + +$, which corresponds to a period-6 orbit with the symbolic dynamics CRL^2RL. It is obtained by appending $+ - -$ to itself with the opposite sign. Similarly, the σ sequence for the tripled period-2 orbit CR, the σ sequence $+ -$, is obtained by appending $+ -$ twice to itself with opposite signs, corresponding to the period-tripling operator $+ - -$. This gives $+ - - + - +$, which corresponds to $CRLR^3$, as we derived before by a less elegant rule.

We shall encounter the σ sequence again in the next section, where it is used to calculate the growth parameter of a linearized logistic map, the so-called tent map.

Another algorithm *interpolates* a new orbit between two known orbits P and Q by taking the intersection of *harmonic $H(P)$* of P and an *antiharmonic $A(Q)$* of Q [MSS 73]. The harmonic of an orbit P is formed, as before, by appending P to itself and changing the second C to L (or R) if the number of R's in P is odd (even). The antiharmonic of an orbit Q, which in general is not a possible periodic orbit, is defined just like the harmonic except that R and L are interchanged in the replacement of the second C. The σ sequence of an antiharmonic is obtained by appending the original σ sequence to itself without sign change. (The reader may wish to derive the rules that distinguish possible orbits from impossible strings.)

For example, the harmonic of $P = CR$ is the orbit $H(P) = CRLR$, and the antiharmonic of $Q = CRL^\infty$ is the string $A(Q) = CRL^\infty RRL^\infty$. The intersection of these two strings, meaning the string in which the initial symbols of the two

strings $H(P)$ and $A(Q)$ agree, is the "daughter" orbit CRL of period length 3. With this rule, the parameter value r for the daughter orbit always lies between those of the two parent orbits; hence the designation *orbit interpolation*. By repeated forming of harmonics and interpolation, *all* orbits of the map can be constructed from the "first" orbit (C) and the "last" orbit (CRL^∞).

The different period lengths p of stable periodic orbits of unimodal maps appear in a *universal order*. If r_p is the value of the growth parameter r at which a stable period of length p first appears as r is increased, then $r_p > r_q$ if $p \succ q$ (read p precedes q) in the following "Sharkovskii order":

$$3 \succ 5 \succ 7 \succ 9 \succ \cdots$$

$$2 \cdot 3 \succ 2 \cdot 5 \succ 2 \cdot 7 \succ \cdots$$

$$\cdots$$

$$2^n \cdot 3 \succ 2^n \cdot 5 \succ 2^n \cdot 7 \succ \cdots$$

$$\cdots$$

$$\cdots \succ 2^m \succ \cdots \succ 4 \succ 2 \succ 1$$

Thus, for example, the minimal r value for an orbit with $p = 10 = 2 \cdot 5$ is larger than the minimal r value for $p = 12 = 4 \cdot 3$ because $10 \succ 12$ in this witchcraft algebra.

Some of the consequences of this ordering are the following:

• The existence of period length $p = 3$ guarantees the existence of any other period length q for some $r_q < r_p$.

• If only a finite number of period lengths occur, their lengths must be powers of 2—that is, $p = 2^k, 2^{k-1}, \ldots, 4, 2, 1$, for some k.

• If a period length p exists that is not a power of 2, then there are infinitely many different periods.

The superstable orbits for the smallest parameter value r_p have the symbolic dynamics $CRLR^{p+3}$ for odd period lengths $p \geq 3$. The "last" superstable orbit of period length p, that is, the orbit with the *largest* value of r_p, has the symbolic dynamics CRL^{p-2} [CE 80]. For example, the last period-6 superstable orbit is CRL^4 and has $r \approx 3.9975831$.

These results are a consequence of Sharkovskii's theorem, which concerns the existence and ordering of orbits according to Sharkovskii's dominance definition given previously for a *fixed* value of r [Sha 64]. However, most of these orbits are unstable. (They are the remaining "ghosts" of orbits that were stable for smaller r values. The fixed point $x^* = 1 - 1/r$, for example, persists even for values of r exceeding 3, where it becomes unstable.) In fact, one-hump maps

$f(x)$ with a "Schwartzian derivative"

$$\frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2$$

that is negative were shown by D. Singer to have at most one stable periodic orbit [Sin 78]. (Here $f(x)$ is assumed to be three times continuously differentiable and to map the unit interval into itself.)

For example, for $r = 3.83187405529$ and $x_0 = 0.5$ we obtain the superstable orbit CRL of period length 3. But by choosing $x_0 = 1 - 1/r = 0.73903108882$, the unstable orbit of period length 1 is initiated. And for $x_0 = 0.89208905218$, the unstable orbit of period length 2 is obtained. Other initial values lead to an unstable period-4 orbit that is descended from the orbit $CRLR$. However, the period-4 orbit based on $CRLL$ cannot be realized for $r = 3.83 \ldots$, because $CRLL$ dominates CRL. The reader may want to find initial values for other period lengths, *all* of which are possible according to Sharkovskii's theorem for $r = 3.83 \ldots$ because the period length 3 is possible.

The Calculation of the Growth Parameters
for Different Orbits

Given a superstable orbit with the symbolic dynamics $Q = CRL \ldots$, what is the corresponding value of the growth parameter r? One method is to adjust r in $f(x) = rx(1 - x)$ iteratively until $f^{(p)}(0.5) = 0.5$, where p is the period length of the orbit and $f^{(p)}(x)$ is the pth iteration of $f(x)$. However, this method is likely to fail in regions where the r values for "similar" orbits are crowded. *Similar* here refers to orbits with equal parity (number of R's) and period lengths that divide the given period p. Also, of course, there must be a good initial guess of r.

The confusion with other orbits can be eliminated if the symbolic dynamics are actually used in the calculation of r—not just its parity and period length. For such a method it is advantageous to transform the variable x linearly to yield another, often-used form of the quadratic map: $g(x) = 1 - \mu x^2$, in which the growth parameter μ is related to r by the equation $\mu = r(r - 2)/4$ or $r = 1 + \sqrt{(1 + 4\mu)}$. For this form of the quadratic map, the maximum occurs for $x = 0$. Hence superstable orbits contain the value $x = 0$.

Let us take as an example the period-5 orbit with the smallest r value, which has symbolic dynamics $CRLR^2$. Set $x_0 = x_5 = 0$; then, because $g(0) = 1$, $g_R(g_R(g_L(g_R(1)))) = 0$. In this equation the subscripts (R or L) remind us which branch of $g(x)$ comes into play at each iteration; we need to know this in order to be able to invert the equation. Inverting yields

$$1 = g_R^{-1}(g_L^{-1}(g_R^{-1}(g_R^{-1}(0)))) \tag{2}$$

where

$$g_R^{-1}(x) = +\sqrt{\frac{1-x}{\mu}}$$

and

$$g_L^{-1}(x) = -\sqrt{\frac{1-x}{\mu}}$$

Multiplying equation 2 by μ gives

$$\mu = \sqrt{\mu + \sqrt{(\mu - \sqrt{(\mu - \sqrt{\mu})})}} \tag{3}$$

from which μ may be determined by iteration, as suggested by H. Kaplan. Starting with $\mu_0 = 2$, one obtains quickly and without ambiguity the correct parameter value for the orbit $CRLR^2$: $\mu = 1.625413725\ldots$, which corresponds to $r \approx 3.738914913$.

For an arbitrary allowed orbit $CRL\ldots$, the plus or minus signs appearing in equation 3 are determined by the symbolic dynamics with the letter R corresponding to a minus sign and L to a plus sign. The first sign under the square root $(+)$ corresponds to the first letter (L) after the initial CR in $CRL\ldots$.

Equation 3 is particularly useful for the calculation of the parameter values for accumulation points of certain orbits. For example, looking at the Sharkovskii order, we may want to know at which μ value the orbits with odd period lengths $p = 3, 5, 7, 9, \ldots$ accumulate. These orbits have the symbolic dynamics $CRLR^{p-3}$, as already noted. For $p \to \infty$, equation 3 therefore becomes

$$\mu = \sqrt{\mu + \sqrt{(\mu - \sqrt{(\mu - \sqrt{(\mu \ldots)})})}}$$

with an infinite sequence of minus signs. Setting $\mu - \sqrt{(\mu - \sqrt{(\mu \cdots)})} = x$, we have $x = \mu - \sqrt{x}$. Eliminating x leads to the cubic equation for μ, namely, $\mu^3 - 2\mu^2 + 2\mu - 2 = 0$, with the solution $\mu = 1.543689012\ldots$, which corresponds to $\tilde{r}_1 = 3.67857351\ldots (=\mu^3)$. This is also the parameter value at which the last two chaos bands merge, because they have the same symbolic dynamics (see pages 279–280).

Because the orbits of *even* period length of the form $p = 2 \cdot 3, 2 \cdot 5, 2 \cdot 7, \ldots$ have the same symbolic dynamics, $CRLR^{p-3}$, as the odd orbits, they accumulate at the same parameter value. But they approach it from below.

Similarly, one determines the parameter value $\tilde{r}_2$ of the accumulation point of the orbits with period lengths $p = 4 \cdot 3, 4 \cdot 5, 4 \cdot 7, 4 \cdot 9, \ldots$ with the symbolic dynamics $CRLR^3(LR)^{p/2-3}$. This gives $\tilde{r}_2 = 3.5925721841\ldots$, which is also the point at which the four chaos bands merge into two bands.

The orbits CL^n, $n \to \infty$, lead to $\mu = \sqrt{(\mu + \sqrt{(\mu \cdots)})}$, that is, $\mu = \sqrt{(\mu + \mu)}$, for the accumulation point, with the solution $\mu = 2$ (corresponding to $r = 4$).

Another "one-hump" map, the piecewise linear "tent" map

$$f(x) = \lambda(1 - |x - 1|) \qquad 0 \leq x \leq 2 \qquad (4)$$

is much easier to analyze than the quadratic map [DGP 78]. Yet the tent map in equation 4 shares many properties with the quadratic map, such as the ordering of orbits, and their symbolic dynamics as the parameter λ is increased from 1 to 2. (However, some orbits of the quadratic map are missing in the tent map, such as those that have resulted from period doubling. Also, because $|f'(x)| > 1$ for $\lambda > 1$, there are *no* stable orbits.)

The determination of the parameter value λ for a given orbit $CRL \ldots$ is particularly simple. The value of λ is given as the solution of the equation

$$\sum_{k=1}^{P} \frac{\sigma_k}{\lambda^k} = 0 \qquad (5)$$

For purposes of iteration, starting with $\lambda = 1.5$, say, the following form is more convenient (it also uses the fact that $\sigma_1 = 1$ and $\sigma_2 = -1$):

$$\lambda = 1 - \sum_{k=3}^{P} \frac{\sigma_k}{\lambda^{k-2}} \qquad 1 < \lambda < 2 \qquad (6)$$

Here the σ_k equal $+1$ (or -1), depending on whether the number of R's in the orbit $CRL \ldots$ to the left of and including the kth symbol is even (or odd). For example, for the orbit CRL of period length $P = 3$, equation 6 reads $\lambda = 1 + 1/\lambda$, which has the solution $\lambda = (\sqrt{5} + 1)/2 = 1.618 \ldots$. For the "lowest" of the three orbits of period length 5, $CRLRR$, equation 6 yields $\lambda = 1 + 1/\lambda - 1/\lambda^2 + 1/\lambda^3 \approx 1.5128763968$. (Note that σ_3 always equals -1. In fact, there are no solutions to equation 6 for $1 < \lambda < 2$ otherwise.)

The σ_k are the same as the "σ sequence" that we introduced on page 283 to facilitate the generating of new orbits form known orbits. In a sense, the σ sequences are the most useful form of the quantized orbital dynamics of unimodal maps.

The common accumulation point for orbits with odd period lengths p and orbits with period lengths p of the form $2 \cdot 3, 2 \cdot 5, 2 \cdot 7, \ldots$ is particularly easy to find with equation 6. As noted before, the common symbolic dynamics of these orbits are $CRLR^{p-3}$. The corresponding σ sequences are $+ - - + - + - + - \cdots$; that is, $\sigma_k = (-1)^k$ for $k \geq 3$. For $p \to \infty$, we therefore have $\lambda = 1 + 1/(\lambda + 1)$, with the solution $\lambda = \sqrt{2}$.

For finite p, the solution is given by

$$\lambda^2 = 2 - \frac{(-1)^p}{\lambda^{p-2}}$$

or, asymptotically,

$$\lambda \approx \sqrt{2}\left[1 - \frac{(-1)^p}{2\sqrt{2^p}}\right]$$

Thus, the differences between the λ values for successive odd (or even) orbits and their accumulation point $\lambda = \sqrt{2}$ form an asymptotically geometric progression with a factor of $\frac{1}{2}$ as the period length is increased by 2. This compares with the factor $1/4.669\ldots$ for the period-*doubling* sequence of the quadratic map. And while the period-doubled orbits have *adjacent* parameter intervals, many other orbits intervene between the parameter values for period lengths 3 and 5 or those of other adjacent odd period lengths.

Another helpful property of the σ sequence is that equation 5 can be factored for orbits generated by parent orbits, yielding interesting relationships between the λ values of parents and offspring. For example, for the period-6 orbit with σ sequence $+ - - + - +$ (the tripled period-2 orbit), $\lambda = \lambda_3^{1/2}$. Here λ_3 is the parameter for the period-3 orbit $+ - -$.

Tangent Bifurcations, Intermittency, and $1/f$ Noise

Something strange happens for $r = 1 + \sqrt{8}$: a so-called *tangent bifurcation*. For r just below $1 + \sqrt{8}$ (see Figure 12), the iterates become "trapped" for a long time between the logistic parabola and the straight line $x_{n+1} = x_n$. Figure 13 shows intermittent period-3 pulses for $r = 1 + \sqrt{8} - 10^{-4}$. The power spectrum of this process decreases as the reciprocal of frequency f. This phenomenon, called *intermittency*, is one of the main mechanisms of $1/f$ noise in nature.

For r slightly above $1 + \sqrt{8}$, the thrice-iterated quadratic map $f^{(3)}(x)$ (see Figure 14) acquires six additional fixed points: three with an absolute slope larger than 1, which belong to an unstable orbit of period length 3, and three with a slope less than 1, which are the three points belonging to the stable orbit with period length 3. This is the famous period-3 orbit, which guarantees that all *other period lengths exist, albeit as unstable orbits, at the same parameter value.* This coexistence of an infinite number of unstable orbits has been called *chaotic* by Li and Yorke. The fact that "period three implies choas," the title of their paper, was enunciated by these authors in 1975 [LY 75].

As r is decreased below $1 + \sqrt{8}$, all these orbits become stable at small but finite intervals of the growth parameter. All orbital periods except $p = 2$
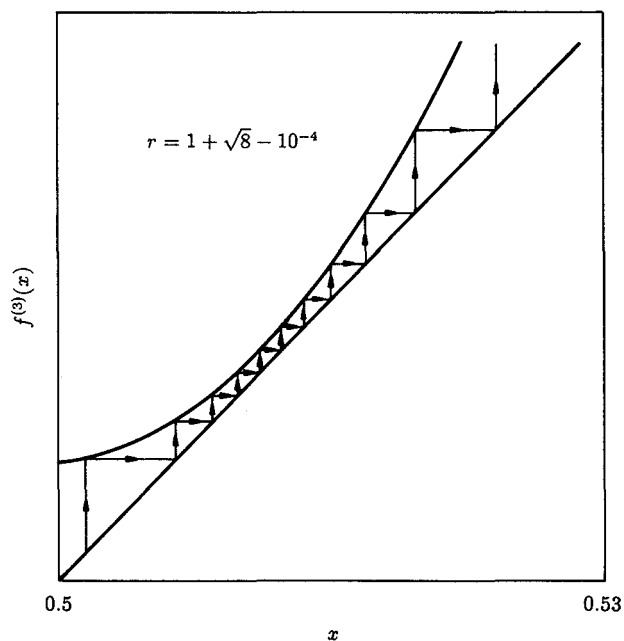
Figure 12 Trapped iterates near a tangent bifurcation for $r = 1 + \sqrt{8} - 10^{-4}$.

and 3 are stable at more than one r interval. For p an odd prime, the number of such intervals with different orbits equals $(2^{p-1} - 1)/p$. Together with the period-doubling pitchfork bifurcation, the tangent bifurcation is the main source of new orbits.

The r intervals for stable orbits are dense; that is, the parameter values for which no stable periodic orbits exist form no intervals. Nevertheless, they have
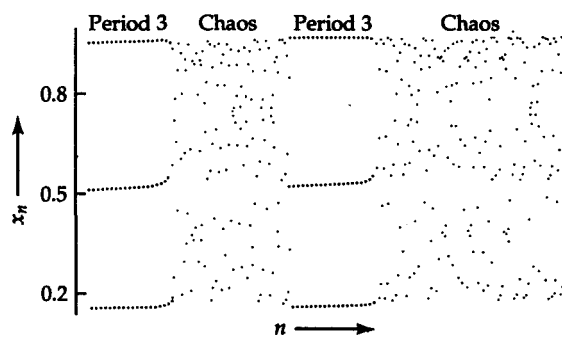

Figure 13 Intermittency for growth parameter just below tangent bifurcation: period-3 pulses alternate with random pulses.
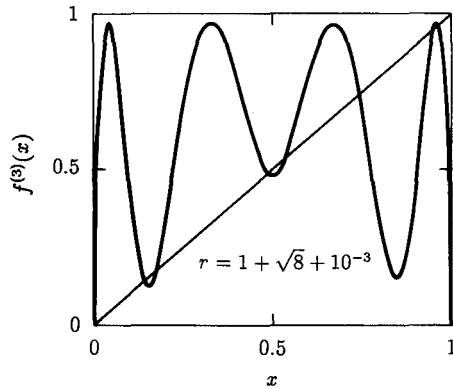
Figure 14   The thrice-iterated quadratic map for the growth parameter slightly above $1 + \sqrt{8}$. This iterated map has acquired six additional fixed points by "tangent bifurcation." Three of these six fixed points are stable (absolute slope smaller than 1) and are members of the stable period-3 orbit visible in Figures 10 and 11.

positive Lebesgue measure. This means that a random choice of the growth parameter has a nonvanishing probability of leading to an aperiodic orbit. This behavior is reminiscent of the irrational numbers, which, too, have positive Lebesgue measure although they form no intervals. Of course, with a finite-state automaton such as a digital computer—not to mention analog machines—an aperiodic orbit can never be proved as such.

Apart from the period-doubling bifurcations starting with the stable fixed point and ending at $r = 3.5699\ldots$, the range of r values for the period length 3 is larger than that for any other period length. The period-3 "window" in the Feigenbaum diagram (see Figure 10A) is thus the most prominent among all the periodic windows and one of the few that are actually visible without a "magnifying glass" (i.e., a computer program that enlarges a small interval of the growth parameter for better visibility). At sufficient magnification one can see period-doubling bifurcations, as in Figure 10B, in each of these periodic windows, each governed by the same Feigenbaum constant $\delta = 4.6692\ldots$.

A Case of Complete Chaos

A particularly interesting value of the growth parameter for the quadratic map is $r = 4$. The transformation

$$y = \frac{1}{2} - \frac{1}{\pi}\arcsin\left(1 - 2x\right) \qquad 0 \leq y \leq 1 \tag{7}$$
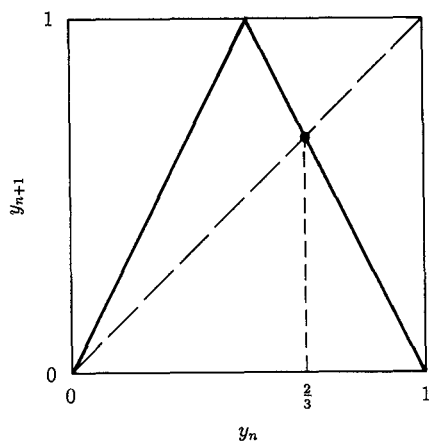
Figure 15   A chaotic tent map: $y_{n+1} = 1 - |2y_n - 1|$.

turns the quadratic map into another tent map:

$$y_{n+1} = 1 - |2y_n - 1|$$

which consists of two straight-line segments with slopes $+2$ and -2, and a maximum at $y_n = 0.5$ (see Figure 15).

By "flipping" the right half of the tent map ($y_{n+1} \rightarrow 1 - y_{n+1}$ for $y_n > 0.5$), we obtain the exceedingly simple *binary-shift map*:

$$y_{n+1} = 2y_n \bmod 1 \tag{8}$$

If we express y_n as a binary fraction, then this map is nothing but a left shift of the digits, with any 1s protruding to the left of the binary point dropped. As a consequence, a value of y that has a terminating binary fraction is mapped eventually into 0. In general, rational values of y, which have periodic binary fractions, lead to periodic orbits. By contrast, irrational y—that is, almost all y in the interval (0, 1)—give rise to nonperiodic orbits.

Although for $r = 4$ almost all initial values y in (0, 1) entail aperiodic orbits, such aperiodic values of y form no intervals. In fact, the initial values for periodic orbits are dense in (0, 1). To see this we truncate a given value of y after an arbitrarily large number of binary places and repeat the remaining bits periodically. Thus, for example, within less than 2^{-5} of $y = 0.10110001 \ldots$ we find an initial value with period length 5, namely, $y_0 = 0.\overline{10110}$ (or any longer period length).

Excepting "nonnormal" binary numbers (see pages 241–243 in Chapter 11) as starting values y_0, the iterates of irrational y_0 fill the unit interval with a uniform density. Such a distribution is called the *invariant distribution of the mapping,*

because a random variable starting out with it stays with it. As a consequence of the uniform distribution of y, the invariant distribution $p(x)$, which is related to y by equation 7, is the U-shaped distribution well known from the arcsine law of random walk theory:

$$p(x) = \left| \frac{dy}{dx} \right| = \frac{1}{\pi} [x(1 - x)]^{-1/2} \qquad 0 < x < 1$$

See Figure 16, which, after proper scaling and shifting, also approximates the invariant distributions in the chaos bands for other parameter values.

The map in equation 8 also illustrates very nicely what is meant by *deterministic chaos*. Suppose the initial condition of a physical dynamic system is represented by a value of y_0 with some *finite* precision. For example, for an eight-place binary precision, we would have

$$y_0 = 0.10110101^{000}_{111} \cdots$$

where the double entries reflect the fact that we do not know whether those digits beyond the eighth are 0 or 1. Because of this *unavoidable* lack of perfect precision, the ninth and all higher iterates of y_0 are

$$y_k = 0.^{000}_{111} \cdots$$

that is, they assume any possible values in the unit interval in a completely
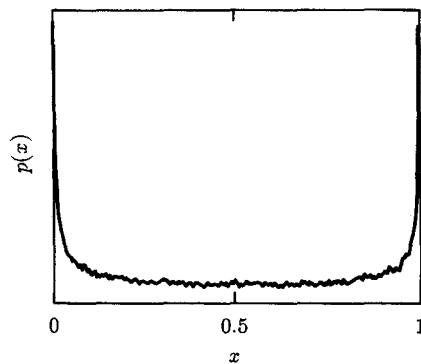

Figure 16   Invariant distribution of quadratic iterates for $r = 4$. Most iterates cluster near $x = 0$ and $x = 1$ [CE 80].

unpredictable succession. It is in this manner that a deterministic law, including the simple equation 8, produces chaotic results, called, appropriately, *deterministic chaos*. Even for completely deterministic dynamic laws, initial conditions with finite precision will—under proper magnifying circumstances—eventually produce totally unpredictable results.

For this to happen, the iteration law must be nonlinear and, in fact, not uniquely invertible. Maps that meet this requirement include those with a maximum (quadratic or otherwise) or with a remainder (mod) operation, as in equation 8. In addition, the nonlinearity must be strong enough that any initial uncertainty will grow exponentially. For example, for the binary-shift map (equation 8), the uncertainty grows by a factor of 2 with each iteration.

By contrast, for the logistic parabola with $r = 2$, for example, any initial value (other than 0 or 1) will lead to the fixed point $x^* = 0.5$ and the differences ε_n between successive iterates x_n and x^* will decrease (asymptotically) by a factor of $-2\varepsilon_n^2$ with each iteration. This is often referred to as *quadratic convergence*. For instance, starting with $x_0 = 0.45$ ($\varepsilon = -0.05$), successive differences will be approximately -0.0005, -0.00005, -0.000000005, and so on. In other words, the "error" becomes rapidly smaller and smaller; the distance of the 5 from the decimal point is doubled with every iteration—as opposed to the one-digit shift per iteration for the map in equation 8.

The convergence is not as rapid for values of r that do not correspond to superstable orbits (as $r = 2$ does, because the orbit includes the "flat top" of the parabola at $x = 0.5$). For example, for $r = 2.5$, the fixed point is $x^* = 1 - 1/2.5 = 0.6$, which differs from 0.5 and is therefore not superstable. Differentiation at the fixed point will show us how the iteration will converge. For $r = 2.5$, we obtain

$$f'(x^*) = 2 - r = -0.5$$

Thus, the differences between x_n and x^* will decrease (asymptotically) by a factor of -0.5. For example, $x_0 = 0.61$ will lead to the successive differences $x_n - x^* = 0.01000$, -0.00525, 0.00256, -0.00129, 0.00064, and so on. For this nonsuperstable orbit the convergence is much slower and is described as linear. In the example, it corresponds to an asymptotic left shift of $|x_n - x^*|$ by (an average) $\log_{10} 2 = 0.3$ decimal places—as opposed to the *doubling* of the left shift with each iteration for a superstable orbit.

Thus, superstable orbits have two major advantages:

1 They converge much faster and are more stable in the presence of small perturbations—hence the name *superstable*—and are therefore more easily measured experimentally.

2 They have simpler theoretical descriptions, such as the symbolic dynamics discussed in this chapter.

Many of the properties of the quadratic map are paradigmatic not only for other unimodal maps but for different nonlinear mappings as well. These laws in turn model a broad range of contemporary problems in which nonlinearities play an essential role.

The Mandelbrot Set

The shortest path between two truths in
the real domain passes through the
complex domain.
—JACQUES HADAMARD

As we saw in the preceding section, the quadratic map for $r = 4$ can be transformed into a simple tent map, which can be further simplified to the binary-shift map $y_{n+1} = 2y_n \bmod 1$. From this map the orbit of any initial point y_0 can be directly inferred by writing y_0 as a binary fraction: periodic binary fractions lead to periodic orbits, but irrational y_0 with normal aperiodic fractions lead to chaotic orbits.

Unfortunately, this simple mapping is not applicable for other values of the growth parameter of the real quadratic map that we have studied so far. However, if we "complexify" both the growth parameter and the variable, the quadratic map becomes considerably more transparent and amenable to analysis. As in many other branches of mathematics (number theory, for example) the introduction of *complex* variables makes many proofs and relations much simpler (as, for example, the proof of the prime number theorem, which dictates the distribution of primes). Hence the paradoxical mathematical motto "complexify to simplify."

. In its complexified version, the quadratic map is often rendered in the form

$$z_{n+1} = z_n^2 + c \tag{9}$$

where both the variable z and the growth parameter c are allowed to assume complex values, graphically represented by points in the complex z plane and the complex c plane. For real c, the relation with the previously used parameters is $c = -\mu = -r(r - 2)/4$.

One of the first questions that comes to mind when looking at equation 9 is, For what values of the parameter c do the z_n stay bounded as the iteration is continued indefinitely? Obviously, for $c = 0$ and z_0 in the unit disk $|z_0| \leq 1$, z_n stays within the unit disk for $n \to \infty$. However, even for $c = 0$, the initial value $z_0 = 2$, for example, gives $z_7 = 2^{2^7} > 10^{38}$; that is, the seventh iteration already exceeds the diameter of the universe measured in atomic units.

What about $c \neq 0$? For $c = -2$, for example, and $z_0 = 0$, equation 9 gives $z_1 = -2$, $z_2 = 2$, $z_3 = 2$, and so on. Thus, $c = -2$ leads to a preperiodic orbit, with period length 1—in other words, a fixed point, $z = 2$.

In general, the set of all points c for which the iteration $z_{n+1} = z_n^2 + c$, with $z_0 = 0$, stays bounded as $n \to \infty$ is called the *Mandelbrot set*, or **M** set for short, after Benoit Mandelbrot, who discovered it and analyzed many of its intricate details [Man 80]. The **M** set, shown in black in Color Plate 8A, consists of a large heart-shaped ("cardioid") area to which smaller disks are attached, to which even smaller disks are attached and so forth *ad infinitum* in a roughly self-similar progression. The same cardioid shape festooned with a proliferation of disks, also called *Apfelmännchen*, can be discovered in many other regions of the complex parameter plane if it is sufficiently magnified (see Color Plate 8B and C). But other characteristic shapes, too, are revealed by the computer "microscope": dendrites, whorls, and "sea horse" tails; see the color plates. In these illustrations, black areas belong to the **M** set, and the different colors signify different rates of escape to infinity of z_n for values of c outside the **M** set. (The individual colors were selected from a digital "palette" for distinctiveness and aesthetic appeal.)

Although the **M** set is not self-similar as a whole, it possesses many approximately self-similar substructures, such as cardioids and disks, and sea horse tails and whorls within whorls within whorls with infinitely fine filigree. To think that a resplendent structure such as the **M** set and its surround results from a simple quadratic equation is indeed astonishing; its mathematical gossamer continues to inspire awe even in the hardened professional. The complexity of the **M** set is also a vivid reminder that the complexity that we observe in many natural phenomena, including *Life Itself* (the title of a well-known book by Crick), can result from relatively simple laws [Cri 81]. Clearly, complex behavior does not necessitate complex laws.

Although parts of the **M** set look rather like isolated spots (in fact they are known to have been obliterated by some overzealous art editors), the set is actually a connected set, as proved by Douady and Hubbard [DH 82, 85]. However, it is not known whether the **M** set is everywhere *locally* connected. (A circle from which a single point has been removed is still connected, but is no longer everywhere locally connected: points on different sides of the gap, no matter how close, are connected only through a long circular arc.)

The large cardioid area of the **M** set and each circular disk correspond to a particular periodic orbit: the cardioid to period 1, the largest disk to period 2, and the other horizontally attached disks to periods 2, 4, 8, . . . , terminating in the Feigenbaum accumulation point of period-doubling bifurcations. The largest remaining cardioid on the real axis corresponds to orbits with period 3. Each of the infinitely many disks sprouting on the cardioids in the complex plane corresponds to a periodic orbit with a particular period related to that of its cardioid, and each of these disks has infinitely many smaller disks attached to it that all look similar to each other. In fact, the only deviation from self-similarity apparent to the naked eye in this succession of disks is the cleavage in the "rear" of the mother cardioid.

The Julia Sets of the Complex Quadratic Map

The fact that a given parameter value lies in the **M** set may, of course, not be all we want to know. We are eager to discern how the iterates z_n behave for different z_0. For which values of z_0 are the z_n bounded? For a given parameter c, the set of initial values z_0 for which the z_n are bounded form the so-called *filled-in Julia set* J_c. (The Julia set proper consists of the boundary points of J_c.)
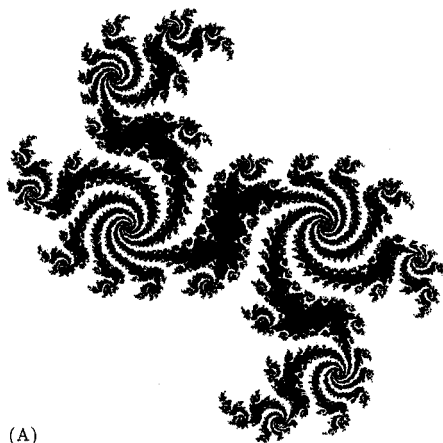
(A)
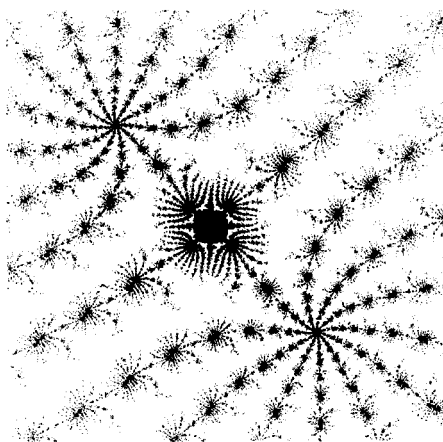
(B)

Figure 17   (A) Filled-in Julia set, defined as the set of all z_0 for which the iteration $z_{n+1} = z_n^2 + c$ is bounded. (B) Another, barely connected Julia set, illustrating the great variety of shapes obtained when the parameter c is changed [Man 83].

Computer experiments have shown that different values of c can lead to a stunning variety of Julia sets, minute changes in c often causing enormous metamorphoses in J_c (see Figure 17A and B).

Some Julia sets are connected; others are just "floating-dust" Cantor sets. Interestingly, those values of the parameter c for which J_c is connected are precisely all the members of the **M** set, so that the latter can also be defined as the set of all c values for which J_c is a connected set. This equivalence is a consequence of a theorem proved independently in 1918 by Gaston Julia and Pierre Fatou, a fact that was rediscovered jointly by Douady and Hubbard, who added many more insights to our store of—still sporadic—knowledge of the deceptively simple iteration $z \rightarrow z^2 + c$ [DH 82].

One of the most consequential discoveries of Douady and Hubbard is that the boundary of the Mandelbrot set can be mapped conformally to the unit circle and that the iteration $z_{n+1} = z_n^2 + c$ corresponds simply to doubling the angle on the unit circle. Thus, measuring angles α in multiples of 2π, the complex quadratic map corresponds to $\alpha_{n+1} = 2\alpha_n \bmod 1$. If the "external angle," as it is *called, is expressed by a binary fraction, then the iteration is a left shift of the* binary digits modulo 1. A c value with an external angle of $\alpha = \frac{13}{31} = 0.\overline{11001}$, for example, will lead to a periodic orbit of period length 5. The individual digits tell us which of the iterates z_n will fall into the upper (0) or lower (1) half plane.
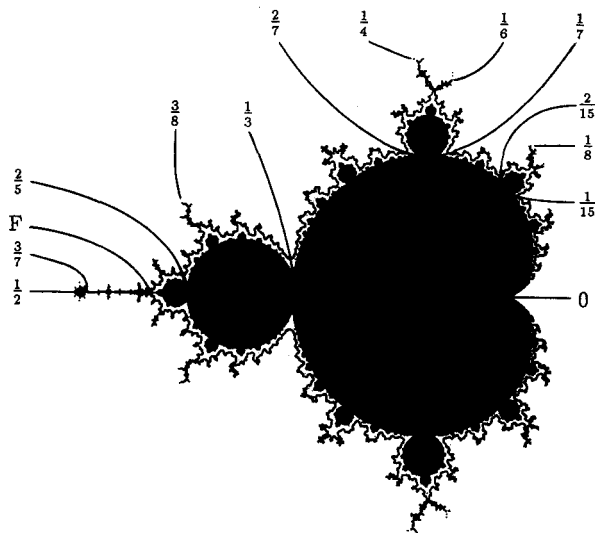

Figure 18 External angles for the Mandelbrot set. The fractions determine the period lengths of the iterates z_n for a given choice of the parameter c. The point "F" marks the accumulation point of the period-doubling cascade [Dou 86].

Figure 18 shows the **M** set and some rational external angles. The accumulation point of period-doubling bifurcations $c = -1.4011\ldots$ (marked "F," for Feigenbaum) has as its external angle the Morse-Thue constant $0.0110100110010110\ldots = 0.412\ldots$, whose binary digits are the Morse-Thue sequence.

The conformal mapping from the boundary of the **M** set to the unit circle may be visualized physically as a problem in electrostatics as follows. Consider an infinitely long conducting bar whose cross section is the **M** set, surround it with a distant electrode in the shape of a circular cylinder, and apply a voltage difference between bar and cylinder. Then the electric field lines from a point on the circle at an angle α with the real axis end on a point c on the boundary of the **M** set with external angle α. This is so because electric field lines obey the laws of conformal mapping. And the equipotential lines, which are orthogonal to the field lines, correspond to c values with equal rates of divergence to infinity for z_n with $z_0 = 0$. (Remember that values of c outside the **M** set lead to unbounded iterates z_n.)

The interested reader will find more fascinating details in H.-O. Peitgen and P. H. Richter's *The Beauty of Fractals*, which also includes an illuminating essay by Douady [PR 86a].