5

CHAPTER

Noises: White, Pink, Brown, and Black

And the rain was upon the earth
forty days and forty nights.

—GENESIS 7.12

Earnings momentum and visibility should continue to propel the market to new highs.

—E. F. HUTTON, the Wall Street brokerage firm, from a forecast issued on October 19, 1987, moments before the stock market plunged

Among the many domains where self-similar power laws flourish, statistics ranks very high. Especially, the *power spectra* (squared magnitude of the Fourier transform) of statistical time series, often known as *noises*, seem addicted to simple, homogeneous power laws in the form $f^{-\beta}$ as functions of frequency f. Prominent among these is *white noise*, with a spectral exponent $\beta = 0$. Thus, the power spectrum of white noise is independent of frequency. But white noise, that is, a noise with a constant or flat power spectrum, is a convenient fiction—a little white lie. Just like white light (whence the name *white* noise), the spectrum of white noise is flat only over some finite frequency range. Nevertheless, white spectra provide a supremely practical paradigm, modeling untold processes across a wide spectrum of disciplines. The increments of Brownian motion and numerous other *innovation processes*, the learned name for a succession of surprises, belong to this class. Electronic and photonic shot noises, thermal noise, and many a hiss from man or beast aspire to membership in the white noise "sonority."

If we integrate a white noise over time, we get a "brown" noise, such as the projection of a Brownian motion onto one spatial dimension. Brown noise has a power spectrum that is proportional to f^{-2} over an extended frequency

range. Some of the paradoxical consequences of such processes, such as the gambler's bad fortune, and the "dusty" consistency of their isosets (sets of constant capital), will be discussed in Chapter 6. However, white and brown noises are far from exhausting the spectral possibilities: between white and brown there is *pink* noise with an f^{-1} spectrum. And beyond brown, *black* noise lurks, with a power spectrum proportional to $f^{-\beta}$ with $\beta > 2$. Figure 5 on page 111 in Chapter 4 showed waveforms of white, pink, and brown noise; Figure 1 shows a waveform of black noise with $\beta = 3$.

As it turns out, both pink and black noises are widespread. Pink processes make their appearance in many physical situations and have surprising aesthetic implications in music and other arts.

Black-noise phenomena govern natural and unnatural catastrophes like floods, droughts, bear markets, and various outrageous outages, such as those of electrical power. Because of their black spectra, such disasters often come in clusters. Indeed, "Wyse men sayth . . . that one myshap fortuneth never alone"; so says A. Barclay in his translation of *The Ship of Fools* [Bar 1509].

All of these phenomena share an important trait: their power spectra are *homogeneous power* functions of the form $f^{-\beta}$ over some respectable range of frequencies, with the exponent β running the gamut from 0 to 4.

Such homogeneous spectra, and the space or time records from which they result, exhibit a simple scaling invariance: if such a process is compressed by a constant scale factor s, the corresponding Fourier spectrum is expanded[1] by the reciprocal factor $1/s$. However, changing the frequency scale by any constant factor does not change the frequency *dependence* for power-law spectra; they keep their form. This can be nicely demonstrated acoustically: when such processes (properly time-scaled to fall into the audio frequency range) are recorded on magnetic tape and played back at a higher or lower tape speed, they do not sound higher or lower in "pitch"; apart from a change in volume, they sound the same! Thus, such spectra are self-similar and the underlying processes are statistically self-similar or self-affine.

Pink Noise

Pink noise, also called $1/f$ noise, has equal power in octave frequency bands or any constant intervals on a *logarithmic* frequency scale. This is a desirable attribute in many applications. For example, pink noise is a favorite test signal in hearing research and acoustics in general because it approximates many naturally occurring noises. Pink noise also has the approximate property of exciting equal-length portions of the basilar membrane in our inner ears to equal-amplitude vibrations, thus stimulating a constant density of the acoustic nerve endings that report

1. This is a fundamental property of the Fourier transform, which underlies the uncertainty principle (see pages 112–114) and many other facts of physics that don't wear such a nice label.

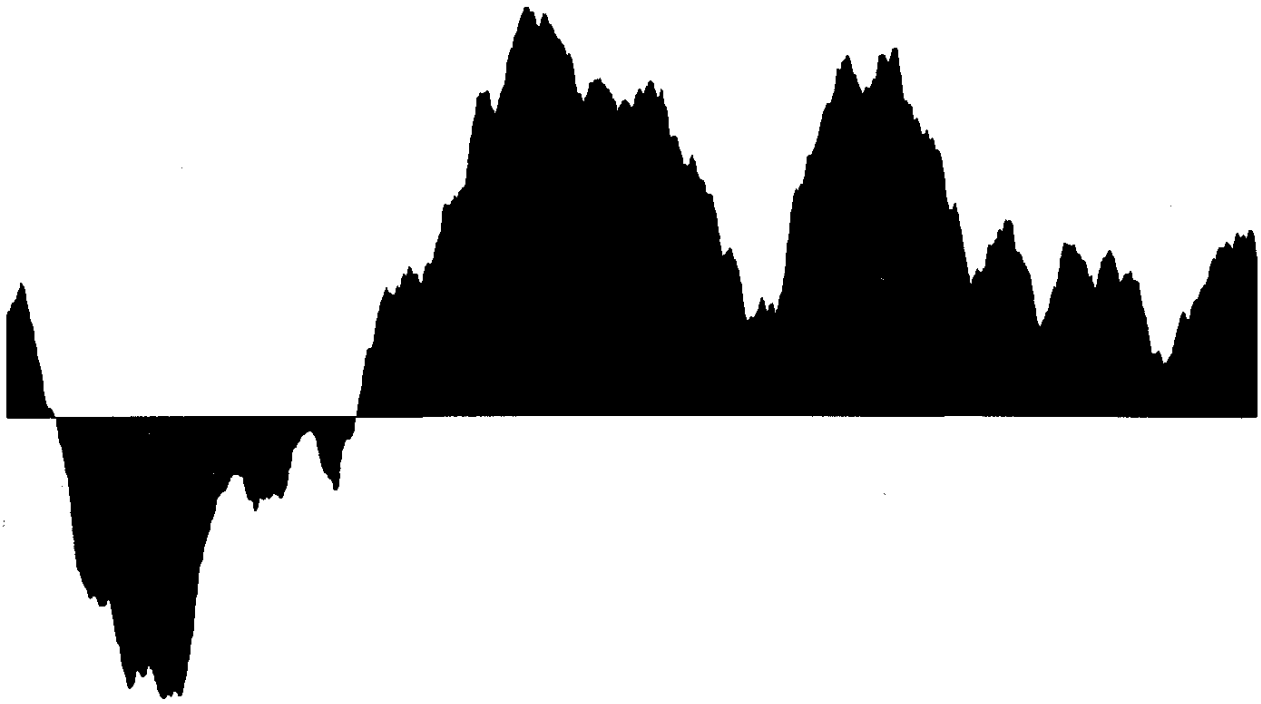


Figure 1 Waveform of "black" noise with $\beta = 3$.

sounds to the brain (see pages 85–86 in Chapter 3). Pink noise is therefore the *psychoacoustic* equivalent of white noise.

Pink noise is also encountered in a wide variety of physical systems, including semiconductor devices. One possible reason for the ubiquity of $1/f$ noises is their genesis through parallel relaxation processes, which abound in nature [Agu 76]. In a relaxation process (think of electrons trapped inside the walls of a potential well in a semiconductor), the trapped particle enters an excited state, where it remains for an exponentially distributed time interval with relaxation time τ. The power spectrum—that is, the squared magnitude of the Fourier transform—$P_\tau(f)$ of such a process is the familiar Lorentz resonance line centered at 0 frequency (a first-order lowpass filter response, to the electrical engineer):

$$P_\tau(f) = \frac{4\tau P_0}{1 + (2\pi f\tau)^2} \tag{1}$$

Here the total power P_0 of the relaxation process—that is, the integral over all positive frequencies $P_\tau(f)$—is independent of τ.

In many physical, chemical, or biological systems there is not just *one* relaxation time but there are a whole spectrum of relaxation times τ that depend on the energy barriers E which keep the structure temporarily trapped in the excited state. The relation between relaxation time τ and energy barrier E is the

famous law named after Svante August Arrhenius (1859–1927):

$$\tau = \tau_0 e^{E/kT} \tag{2}$$

where T is the absolute temperature and k is Boltzmann's constant. Suppose these energies are distributed uniformly in the interval $[E_1, E_2]$; then the distribution of the relaxation time $p(\tau)$ can be obtained from equation 2 by applying the elementary rules for transforming probabilities, yielding a hyperbolic distribution for τ

$$p(\tau) = \frac{kT}{E_2 - E_1} \cdot \frac{1}{\tau} \qquad \tau_1 \leq \tau \leq \tau_2 \tag{3}$$

where $\tau_{1,2} = \tau_0 \cdot \exp(E_{1,2}/kT)$

Superposition of many independent relaxation processes with power spectra given by equation 1 and relaxation times distributed according to equation 3 yields

$$P(f) = \frac{2kTP_0}{\pi(E_2 - E_1)f} [\arctan(2\pi f\tau_2) - \arctan(2\pi f\tau_1)]$$

where the difference inside the brackets, in spite of its somewhat forbidding appearance, is roughly constant in the frequency interval

$$\frac{1}{\pi\tau_2} < f < \frac{1}{4\pi\tau_1} \tag{4}$$

The main point now is that the frequency interval in expression (4) could be, and in numerous situations *is*, very wide. Suppose, for example, that the barrier energies span a range as narrow as $7kT$; then $\tau_2/\tau_1 \approx 10^3$. The corresponding frequencies for which the $1/f$ law $P(f) \sim f^{-1}$ holds within an accuracy of 3 decibels (dB) range over a factor greater than 1200.

Distributions of relaxation times over wide ranges of values have been observed in many physical and biological phenomena [Man 83]. Thus, the electrical voltage on a Leyden jar, an early storage capacitor for electricity, does not decay exponentially with time, with a single relaxation time. Rather, the charge decays *hyperbolically*, implying a wide range of relaxation times [Koh 1854]. The ubiquitous electret microphone, a kind of latter-day Leyden jar, shows a kindred decay of its internal charge [Ses 80].

Similarly, the recovery times of neurons after firing were found by Jerry Lettvin to stretch from fractions of milliseconds to hours and days. And when Wilhelm Weber, following a suggestion by his father-in-law Carl Friedrich Gauss, measured the lengthening of elastic silk threads used in his apparatus, he found that an applied load would give rise not only to an immediate stretching but to

a long-lasting aftereffect, a continual further lengthening that followed a hyperbolic law with elapsed time [Koh 1847].

Hyperbolic decay with time can even be observed in concert halls with insufficient sound diffusion. As a consequence, sound decay in such halls cannot be characterized by a single reverberation time even at a single frequency [Schr 70]. It seems that wherever we look or listen, we see or hear that exponential behavior is much less common than commonly supposed.

We shall later encounter still another mechanism for hyperbolic behavior that gives wide-range $1/f$ spectra: *intermittency*, stemming from *tangent bifurcation* in the logistic parabola and other iterated nonlinear mappings. A generator for generic $1/f$ noise in chaotic Hamiltonian systems, hostage to a self-similar hierarchy of "cantori," was recently proposed by Geisel [GZR 87].

While brown noise is easy to generate (just keep summing independent random numbers), pink noise is a bit more difficult to produce. A relatively simple method of generating pink or $1/f$ noise on a computer is to add several relaxation (first-order lowpass) processes with power spectra like equation 1 and relaxation times τ that form a self-similar progression with a similarity factor equal to 10 (or less for a better approximation). In this manner, just three relaxation times will cover a frequency range of nearly three powers of 10 (see Figure 2).

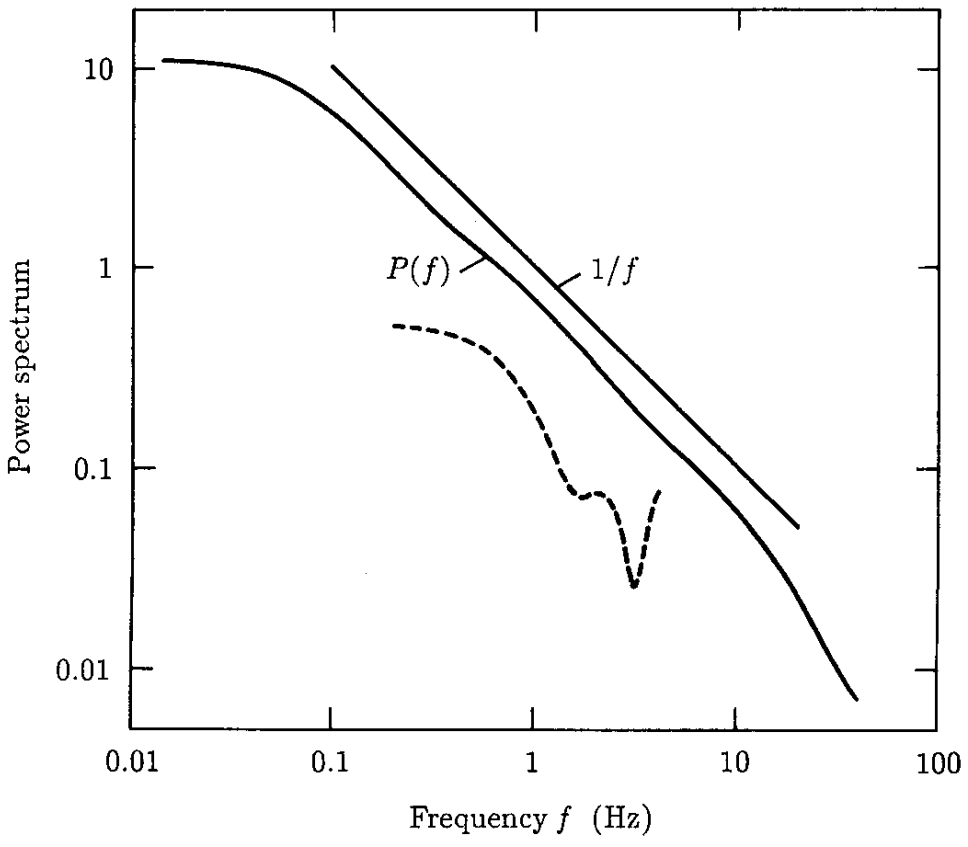


Figure 2 Pink noise from relaxation processes. Solid curve: superposition of three relaxation processes. Dashed curve: Superposition of values on three dice.

A relaxation process with discrete time samples x_n can be generated with the help of a computer's internal random number generator for producing independent random samples r_n and using these (zero mean) samples in the recursion relation

$$x_{x+1} = \rho x_n + \sqrt{1 - \rho^2}\, r_n \qquad x_0 = 0$$

Here ρ is the desired correlation coefficient between adjacent samples. It is related to the relaxation time τ by the equation $\rho = \exp(-1/\tau)$. Thus, for a set of relaxation times that increase by a factor of 10 ($\tau = 1, 10, 100, \ldots$), the correlation coefficients are obtained by taking successive tenth roots (e.g., $\rho = 0.37, 0.90, 0.99, \ldots$).

For less accuracy, three dice, instead of a computer, suffice as random number generators: the first die is rolled for every new sample of the pink noise, the second die is "updated" only every other time, and the third die is thrown only every fourth time. This ingeneously dicey idea is due to Richard Voss [Gar 78]. The sum of the dots on all three dice then forms a random variable with a mean of 10.5 and a variance (noise power) of 8.75 that is a rough approximation to a pink noise over a limited frequency range.

In this parlor-game approach to pink noise, the different relaxation times are mimicked by different persistence times of the dice (increasing by a factor of 2 in our example). However, the three-dice method (the dashed curve in Figure 2) does not approach a straight $1/f$ slope nearly as well as three relaxation processes (the solid curve in Figure 2).

Self-Similar Trends on the Stock Market

One of the neighborhoods where power-law noises dominate the scene, and chaos reigns the charts, is Wall Street, U.S.A. At stock and commodity exchanges, self-similarity weighs in on many scales. This is perhaps best illustrated by my once mistaking a chart of *minute-by-minute* stock averages (see Figure 3) for *day-by-day* fluctuations. I would not have been surprised by self-similarity between daily, weekly, and monthly prices, but I never suspected the same kind of fluctuations to prevail right down to 30-second intervals, which is what the "minute-by-minute" charts actually show.

Of course, once in a while there is an uncharacteristic jump in the data, as in October 1987 (see Figure 4), when the computers that handle programmed trading went wild (through no fault of theirs, of course). Technically, such price jumps are also known as "innovation processes"—some *innovation*, and certainly no consolation to the unlucky trader waiting impatiently for his broker to answer

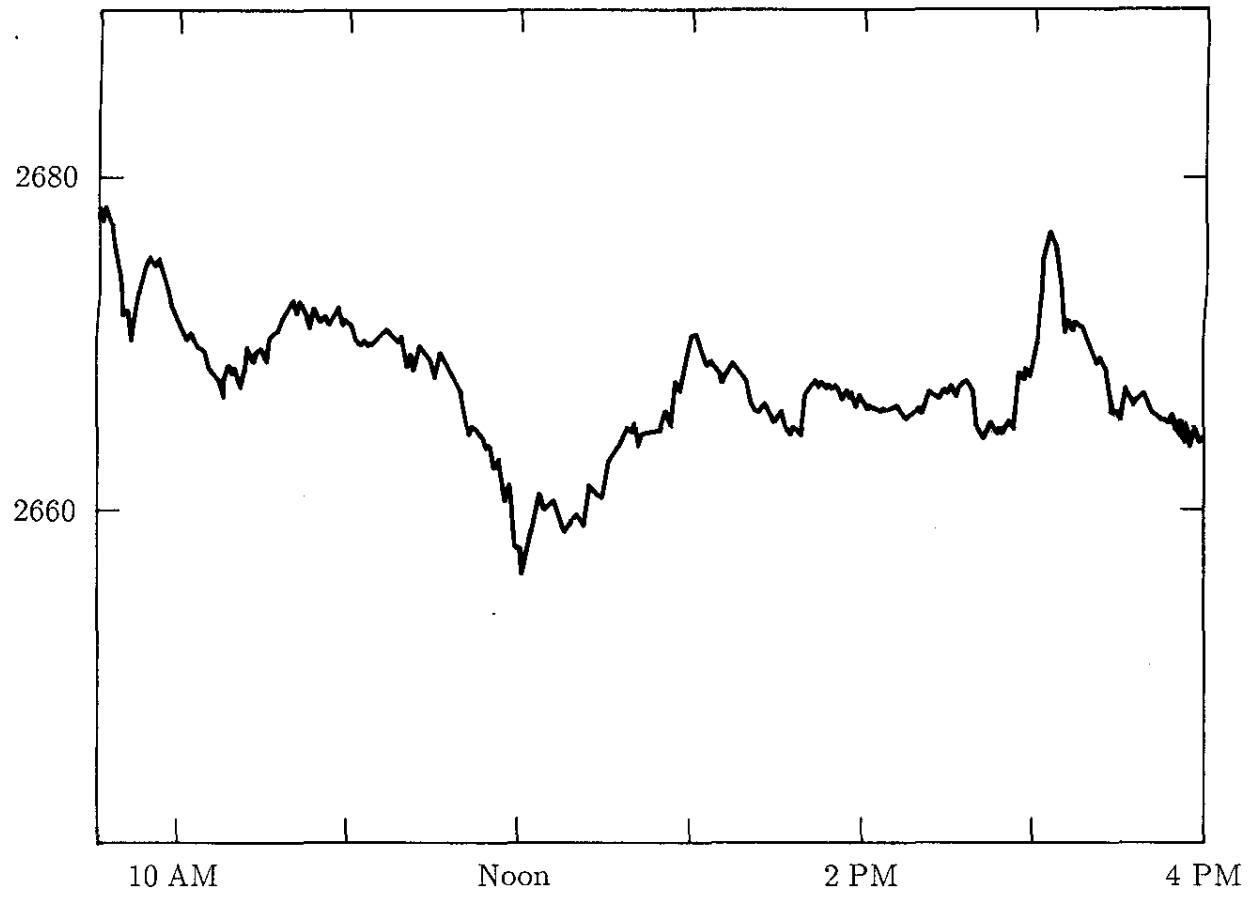


Figure 3 Minute-by-minute stock averages look much like daily averages because stock averages are a self-affine process.

the phone! But after the "novelty" has worn off, price fluctuations resume their habitual course.[2]

The trends and fluctuations of stock prices have been analyzed in great detail in terms of information-theoretic concepts, such as cross entropy and mutual information. In fact, Claude Shannon, the father of communication theory (as he called it), is reported to have become quite rich after he began to apply his theory to the stock market. Now, market analysis is a firmly entrenched branch of information theory, as are other economic applications of entropic principles. But the emergence of programmed trading, executed by fast and soulless machines governed by instant feedback, will necessitate much rethinking of the rules by all concerned: the exchange board, the numerical analyst, and the hapless investor.

2. Speech signals, which are highly redundant (no matter what the semantic content), can also be reduced to innovation sequences (by linear prediction from past sample values). These "prediction residuals" can then be quantized and encoded by a single binary information bit for every four samples of the speech signal, or 0.25 bit per sample, without loss of quality [SA 85]. Ordinarily, speech signals require 8 bits per sample, and compact disks use 16 bits per channel.

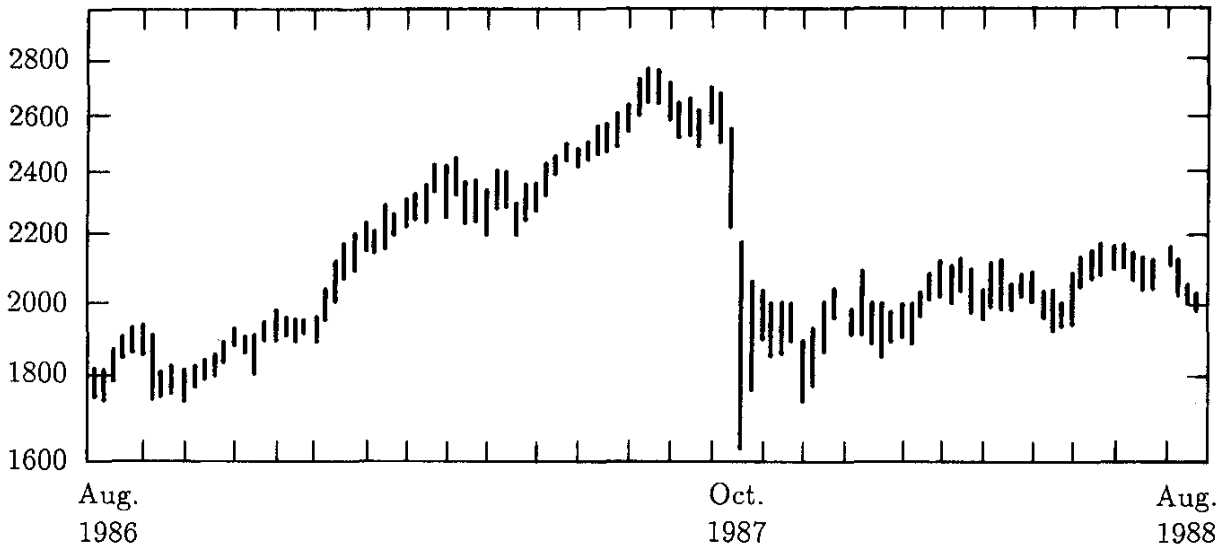


Figure 4 Drop of stock market in October 1987.

In a first approximation to stock averages, one assumes the actual prices to be generated by *independent increments*. The resulting price "noise" has a power spectrum that is proportional to the inverse square of the frequency. Such random runs are now often called brown noises, in an allusion to Brownian motion, that jittery dance of floating dust specks seen in a microscope by the Scottish botanist Brown. (In Brownian motion, the innovation process consists of the independent kicks given the suspended particles by the molecules of the liquid in which they float.)

Another, purer paradigm of brown noise is the fluctuating capital of a gambler, for which independent rolls of dice constitute the innovation process. Suppose the probability of winning a dollar is p for every roll (and that of losing a dollar is $1 - p$). What optimum strategy does information theory teach? Answer: Unless your chances of winning exceed 50 percent, don't play! (Another case of science most profound confirming common sense.)

But what to do if $p \geq 0.5$? This is not as unrealistic as it may seem. The gambler could have side information—legal or otherwise—of the mechanical statistics of the roulette wheel. Here information theory provides another useful answer: Bet, but don't bet *all* your money! To maximize your capital growth, bet the fraction $2p - 1$ of your present capital [Kel 56]. Then the logarithmic growth rate of your capital will attain its highest value, given by Shannon's *information capacity* $C(p) = 1 - H(p)$ of the binary symmetric channel with error probability p. Here $H(p)$ is the entropy function $H(p) = -[p \log p + (1 - p) \log (1 - p)]$. Thus, for $p = 0.6$, say, the gambler should invest 20 percent of his current capital at every roll. Taking logarithms to base 2, $H(p)$ equals 0.97 bits per roll and $2^C = 1.02$. Thus, the well-informed gambler can expect an average gain of 2 percent *per roll*—which is a lot better than the taxable 2 percent interest per *year* that a major European bank recently offered the author, without blushing.

The more timid player who bets only 5 percent of his capital will gain an average of less than 1 percent per roll. After 200 rolls, his gain will be only one-tenth that of the optimum player who properly exploits information theory.

By contrast, the greedy gambler who always bets half of his current capital will, on average, *lose* 3.5 percent of his money per roll, or practically all of it (99.9 percent) during a 200-roll evening. And the reckless player who bets all his current money on every roll will, of course, be cleaned out completely after surviving an average of two rolls. (See also pages 150–152 in Chapter 6.)

Brownian motion contains several subtle statistical self-similarities, and we shall return to the Brownian theme, including more on gambling and the topic of constructing interesting topographies from noise, in the following sections and in Chapter 6.

Black Noises and Nile Floods

If the gambler thinks brown noise is bad enough, expose him to processes with power spectra proportional to $f^{-\beta}$ with $\beta > 2$, which we have called *black* noises. A diffusion process with independent increments Δx diverges but does so only with the square root of elapsed time t; that is, the root mean square distance is proportional to the square root of time:

$$\langle x^2 \rangle^{1/2} \sim t^{1/2}$$

To characterize black processes, we need a new measure of divergence. This was provided by Harold Edwin Hurst (1900–1978) [HBS 65] and Mandelbrot [Man 83].

The quantity in question is the *rescaled range* R/S, which is essentially the range $R(\Delta t)$ of the data over a time interval Δt (after subtracting any linear trend) divided by the sample standard deviation $S(\Delta t)$. For a white Gaussian noise, the ratio R/S tends to a constant for large Δt. In a sense, both R and S measure the range of the data, but R "looks" at the data linearly and S is based on the squared data. For some processes this yields no new information and R/S is asymptotically constant, that is, proportional to Δt^0. But this is not so for numerous geophysical records such as floods and a host of other inhospitable data.

For a Brownian function (power spectrum proportional to f^{-2}), R/S is proportional to $\Delta t^{0.5}$, reflecting the long-range dependence, or *persistence*, hiding behind brown processes. The water-flow statistic of the river Rhine (at the Swiss-French-German triple point near Basel, anyhow) converges on a similar long-term behavior with $R/S \approx \Delta t^{0.55}$ (see Figure 5A).

But other rivers are not as mild and tame as the Rhine [MW 69]. The water-level minima of the river Nile, for instance, taken between the years 622 and 1469 (the drought-dreading Egyptians must have been very patient papyrus

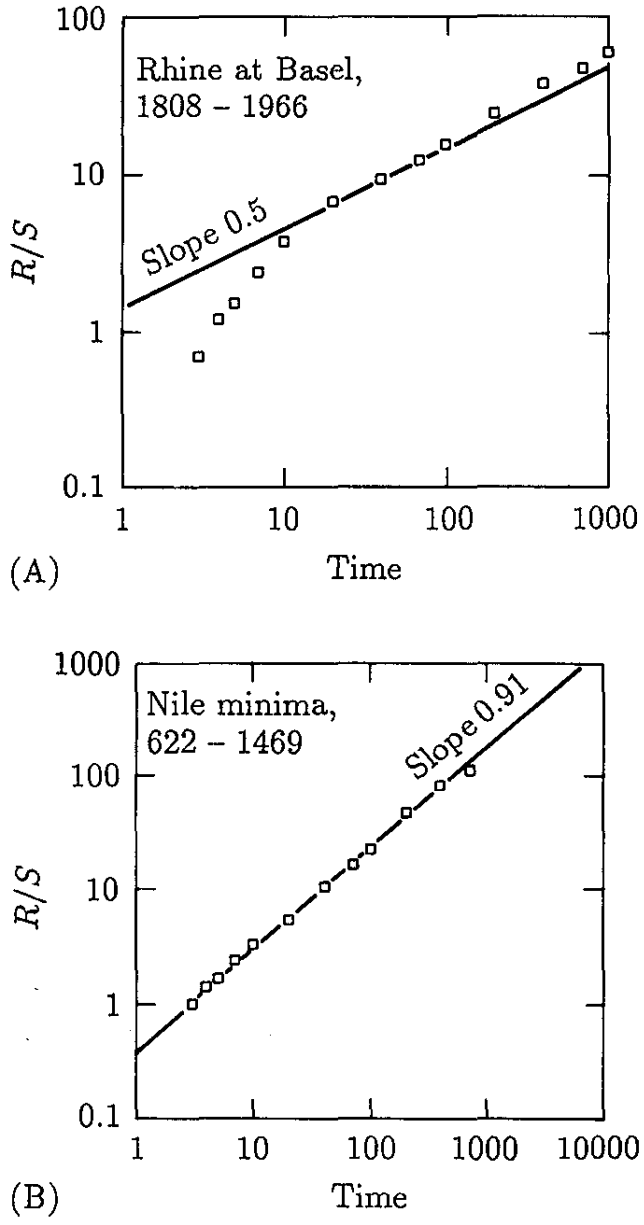


Figure 5 (A) Water-flow statistics of the Rhine; (B) water-level statistics of the Nile [MW 60].

keepers!) show a dependence $R/S \sim \Delta t^{0.9}$ (see Figure 5B), an exponent that reflects a high degree of persistence, of which the Bible has taken due note in the heart- (and coat-rending) Joseph story (Genesis 41).

The *Hurst exponent*, defined by $H := \log (R/S)/\log (\Delta t)$, is a convenient measure of the persistence of a statistical phenomenon. For white noise, which has no persistence, $H = -0.5$; for brown noise, which does have persistence, $H = 0.5$.

Interestingly, there is a simple relation between the Hurst exponent H and the spectral exponent β: $\beta = 2H + 1$. Thus, the Nile noise has a power spectrum proportional to $f^{-\beta} = f^{-2.8}$, implying, like the large Hurst exponent 0.9, a long-range persistence that requires unusually high barriers, such as the Aswân High Dam, to contain damage and rein in the floods.

Warning: World Warming

Processes with pronounced persistence pose perplexing puzzles—and are prone to frequent misinterpretation. Again and again one hears alarmists crying wolf when confronted with seemingly threatening data, but impartial analysis may reveal nothing more threatening than a statistical artifact.

Let us look at a Nile-like noise, with a power spectrum that decays as f^{-3} for large frequencies f. Of course, for small frequencies, the spectrum would *diverge, implying an infinite energy. But no matter how catastrophic, real* calamities are finite. If nothing else, finite observation times T would limit observable excesses. Thus, a realistic power spectrum $P(f)$ with asymptotic f^{-3} dependence, obtained from data collected over a time period T, might look like this:

$$P(f) = \frac{T^4 f}{1 + T^4 f^4} \qquad (f > 0)$$

This spectrum is plotted in Figure 6 for an observation period $T = 1$ (say, 1 year).

But now suppose observations are extended over *two* years. The "new" power spectrum, as seen by the extended observations, is shown by a dashed line in Figure 6. Thus, just extending the observations from 1 year to 2 years has added all the doom-ordaining power shown in gray in the plot.

For an actual example, let us look at a study of the annual population variation of a large number of terrestrial animals over a period of 50 years. Ecologists found that fluctuations over 20 years are roughly twice as large as those observed over 2-year spans—in spite of the fact that the populations appeared to be relatively stable over half a century [RP 88].

Thus, before drawing doomsday conclusions from the exceedingly warm 1988 summer in the continental United States, one should remember Hurst and his exponent, and the strong dependence of extremes on the length of observation. There may well be a "greenhouse" effect of global warming in the air, but confirmation requires a lot more patience (but perhaps not much more carbon dioxide [Fis 90, Wei 90].

On the other hand, minimum viable populations of endangered species must be considerably larger than current estimates, based on time-limited data, would indicate [Law 88]. (For further paradoxes resulting from power-law statistics, see Chapter 6.)

Fractional Integration: A Modern Tool

Brownian motion is obtained from summing independent increments. Summing (or integrating) the increments changes the spectrum from f^0, for innovation

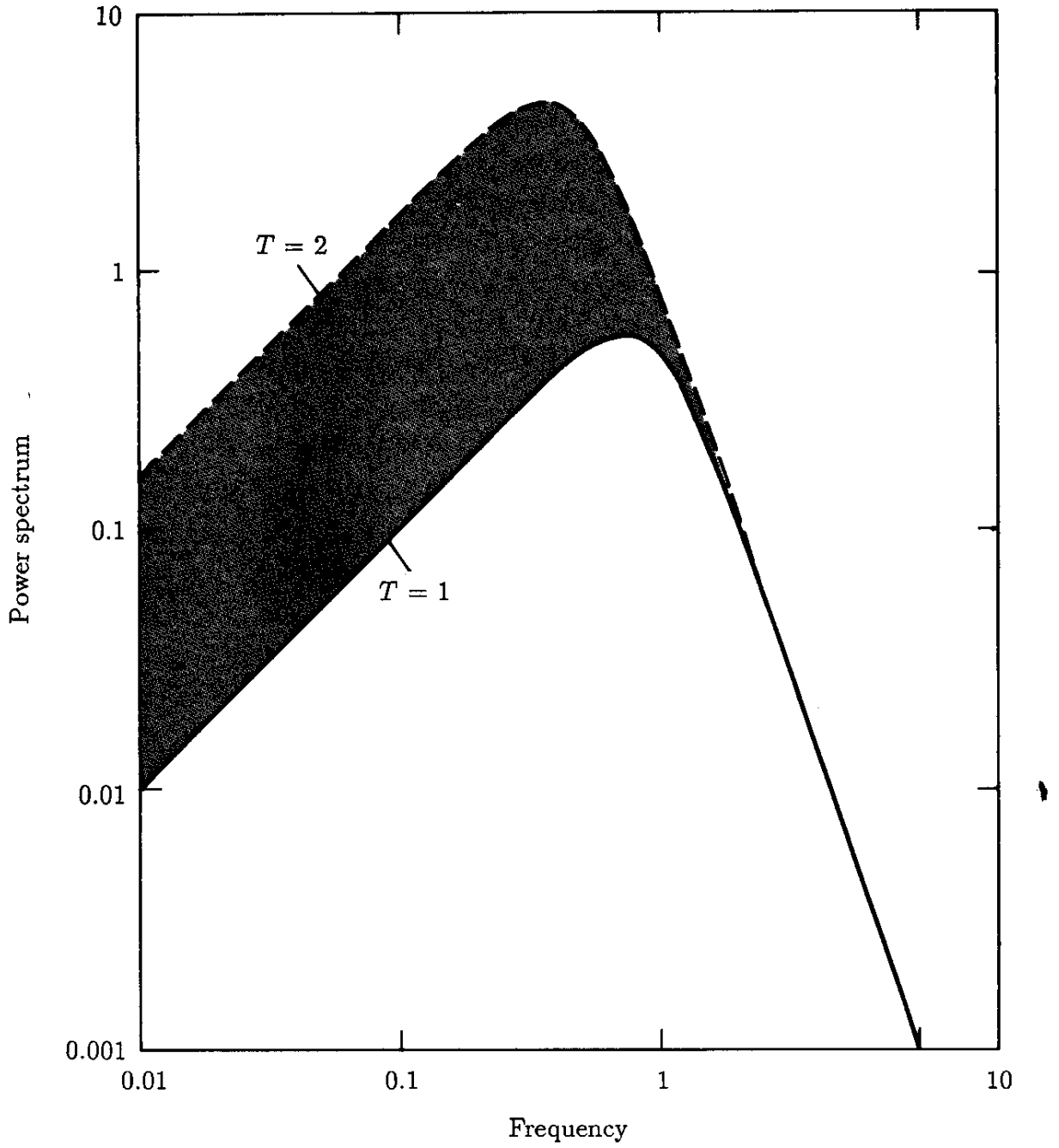


Figure 6 Dependence of measured spectrum on observation time T.

processes, to f^{-2} for the integrated process. Can we not get f^{-1} processes from integration? Yes, but we must first reinvent *fractional* integration, and before that, we must *define* it.[3]

Since integration multiplies a power spectrum by f^{-2}, let us define a *half integration* as an operator that multiplies the power spectrum by f^{-1}. In the conjugate Fourier variable (time or space, say), the operation is a convolution

3. Amazingly, as already noted, Leibniz thought of fractional differentiation and integration 300 years ago, right after the invention of calculus proper. (The incredible Leibniz also seems to have invented the grooves under the head of a nail to improve its cohesion with the material into which it has been hammered.)

integral whose kernel is the inverse Fourier transform of $|f|^{-1/2} \cdot \exp(i\phi(f))$ with appropriate phase $\phi(f)$.

More generally, we can define ν-fractional integration by an operator that multiplies the power spectrum by $f^{-2\nu}$. The corresponding convolution kernel is proportional to $t^{\nu-1}$ [Erd 54].

Fractional integration and differentiation have been useful tools in quantum mechanics and other fields for some time. Now they may also serve in the automated assembly of fractal landscapes and other self-similar structures. While the required numerical convolution in the time or space domain may consume much computer time and memory space, the alternative—Fourier synthesis from the prescribed spectrum—imposes potentially unrealistic periodicities on the resulting fractal.

Brownian Mountains

How can we generalize the Brownian function, $B(t)$, to a function of *two* variables, $B(t_1, t_2)$? In other words, we want to erect a Brownian *mountain*, $B(t_1, t_2)$, over the (t_1, t_2) plane in such a way that any cut of the mountain with a plane perpendicular to the (t_1, t_2) plane will be a typical Brownian function $B(t)$. Interestingly, as seen in the next section, the answer is connected with computer tomography—and, equally unlikely, with imaging by rotating cylinder lenses.

We begin by recalling that $B(t)$ has a power spectrum proportional to f^{-2}. Its amplitude spectrum is therefore proportional to $|f|^{-1}$. For the two-dimensional case, we need an amplitude spectrum which is proportional to $|\mathbf{f}|^{-1}$, where $\mathbf{f}$ is the frequency *vector* (f_1, f_2) with components f_1 (corresponding to the "time" variable t_1) and f_2 (corresponding to t_2). The length of the frequency vector is $|\mathbf{f}| = (f_1^2 + f_2^2)^{1/2}$.

Thus, one method of constructing a Brownian mountain $B(\mathbf{t})$, where $\mathbf{t}$ is the "time" vector (t_1, t_2), is to take sufficiently many independent, identically distributed random samples on a square lattice in the frequency plane, to multiply them by $|\mathbf{f}|^{-1}$, and to Fourier-transform the product into the time plane. For mountains in our world, the time plane will, of course, be a spatial plain, measured in square kilometers.

An alternative method of constructing $B(\mathbf{t})$ is to start with independent samples in the *time* plane and perform an operation on them that is equivalent to multiplying by $|\mathbf{f}|^{-1}$ in the frequency plane. This equivalent operation is a *convolution* with the inverse Fourier transform of the function $|\mathbf{f}|^{-1}$, which, in two dimensions, happens to be the function $|\mathbf{t}|^{-1}$ (within a constant factor that is of no interest). In other words, the two functions $|\mathbf{f}|^{-1}$ and $|\mathbf{t}|^{-1}$ from a *Fourier pair* for the *two*-dimensional Fourier transform. (Another, better known case where a function is similar to its Fourier transform is, of course, the Gaussian distribution

function. In fact, for the Gaussian function this similarity holds in any number of dimensions.)

Thus, if we use $W(\mathbf{t})$ to refer to a two-dimensional array of independent, identically distributed random samples in the time plane, a Brownian mountain is given by

$$B(\mathbf{t}) = W(\mathbf{t}) * |\mathbf{t}|^{-1} \tag{5}$$

where $*$ stands for a convolution integral. Figure 7 shows a mountainscape generated in this fashion. Other methods for generating fractal landscapes and other self-affine fractals are described by Richard Voss [Vos 88] and Dietmar Saupe in *The Science of Fractal Images* [PS 88].

Radon Transform and Computer Tomography

It can be shown that imaging an object in the $\mathbf{t}$ plane with a rotating cylinder lens, combined with time averaging, is equivalent to convolving with the function $|\mathbf{t}|^{-1}$ [SHJ 88]. The same is true for tomographic imaging (using x rays or some

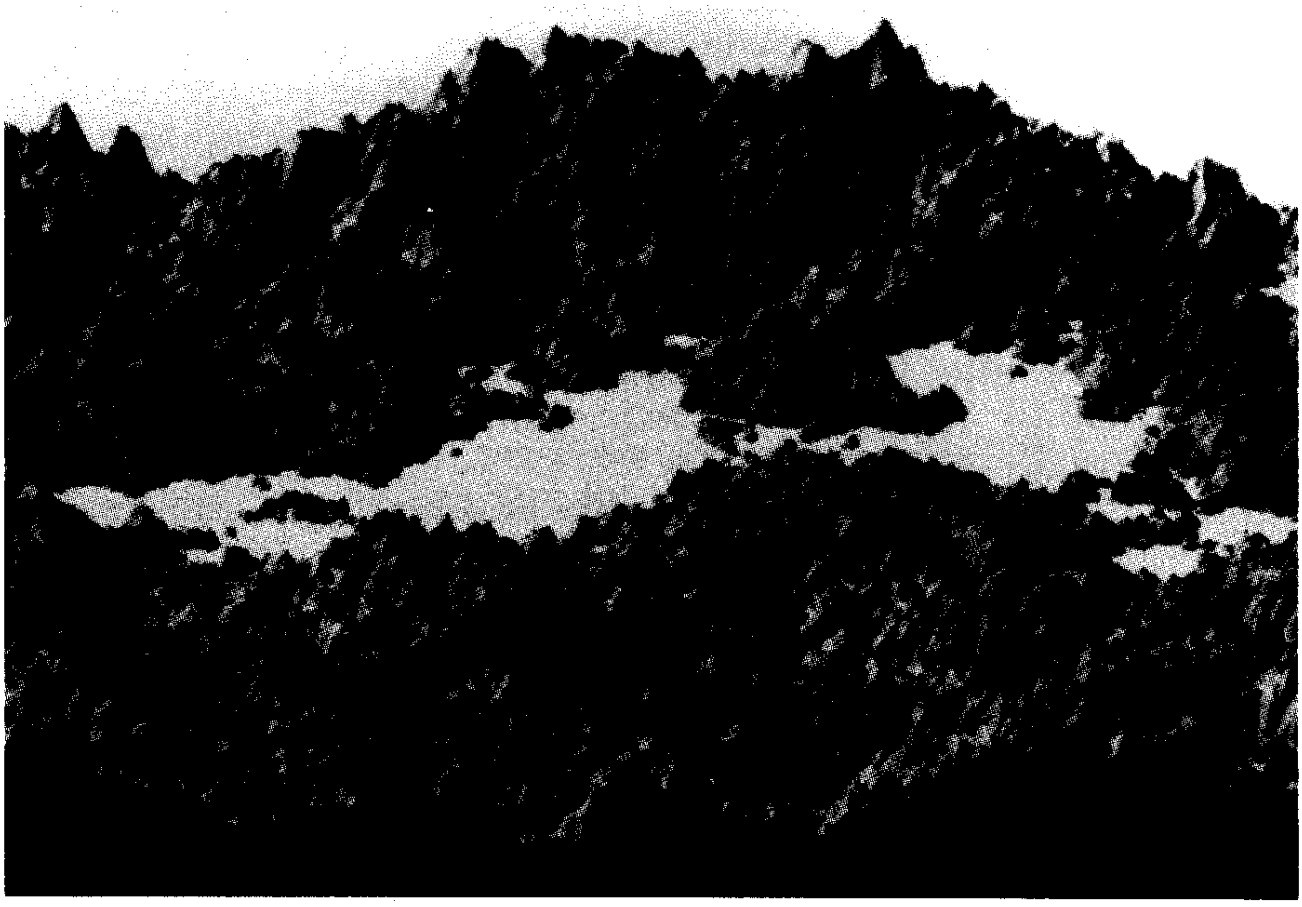


Figure 7 Computer-generated Brownian mountain [Man 83].

other shadow-casting radiation). The transformation in equation 5, which describes these"fuzzy" imaging methods, is also called the *Radon transform.* Figure 8 shows the image of a letter A obtained with a rotating cylinder lens and time averaging.

To recreate a sharp image $W(\mathbf{t})$ from the "blurred image" $B(\mathbf{t})$, we need to execute an *inverse* Radon transform on $B(\mathbf{t})$. And we already know how to do this! We Fourier-transform the $B(\mathbf{t})$ information (preferably on a computer; hence the term *computer* tomography), multiply the result by $|\mathbf{f}|$ to cancel the factor $|\mathbf{f}|^{-1}$, and Fourier-transform back into the $\mathbf{t}$ plane to yield the sharp image $W(\mathbf{t})$. Computer tomography is that simple—if we shun dull and deadly detail!

Fresh and Tired Mountains

A Brownian mountain with an $|\mathbf{f}|^{-2}$ power spectrum has a surface whose Hausdorff dimension is $D = 2.5$ [Vos 85], which makes it look rather rugged, like a geologically very young, ragged mountain (see Figure 9A) whose jagged peaks have had no time to erode.

If we want smoother mountains, like the North American Rockies (except the Grand Tetons), the power spectrum must fall off faster than $|\mathbf{f}|^{-2}$. A good earthlike mountainscape is obtained by mulitplying independent samples in the frequency plane by a factor $|\mathbf{f}|^{-\gamma}$ with $\gamma > 1$. Fourier transformation then yields

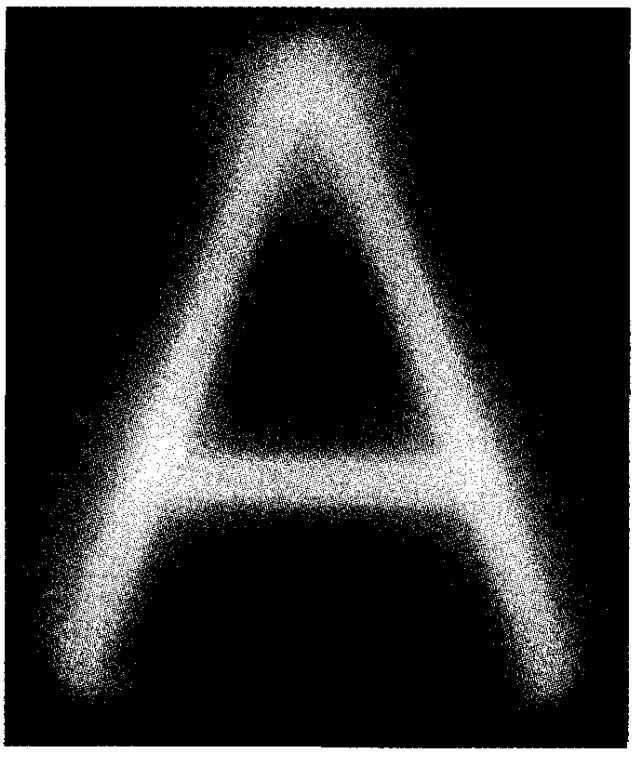
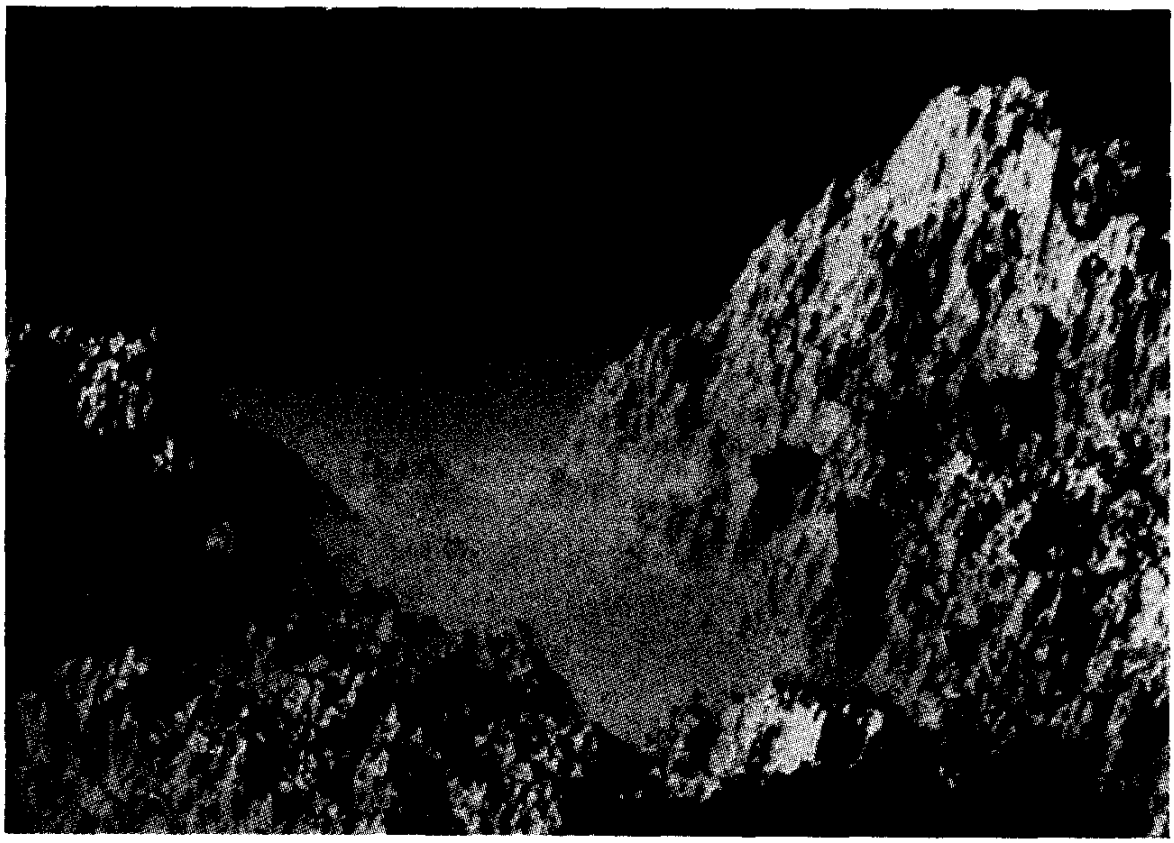


Figure 8 Image of letter A obtained with rotating cylinder lens.

(A)



(B)

Figure 9 (A) Brownian mountain with fractal surface dimension 2.5, (B) mountain with surface dimension 2.1 [Vos 88].

Figure 10 Coastline with fractal dimension 1.33 [Man 83].

mountains with a fractal surface dimension $D = 3.5 - \gamma$. For the spectral exponent $\beta = 2\gamma$, we thus have $\beta = 7 - 2D$.

Figure 9B shows a mountainscape with $D_M = 2.1$ having a power spectrum proportional to $|\mathbf{f}|^{-2.8}$.

Many mountain lakes (see the white areas in Figure 9A) have coastlines with Hausdorff dimensions equal to $D_c = D - 1$, which equals $D_c = 1.5$ in the Brownian case ($\beta = 2$) and $D_c = 1.1$ for $\beta = 2.8$ (Figure 9B). This is somewhat less than the fractal dimension of the west coast of Britain ($D_c \approx 1.25$).

A coastline with $D_c \approx 1.33$ generated in this manner is shown in Figure 10.

In general, for a fractal embedded in E Euclidean dimensions, the fractal dimension D is given by

$$D = E + \frac{3 - \beta}{2}$$

There is also a connection between D and the previously introduced Hurst exponent H (see pages 129–130). With $\beta = 2H + 1$, we have

$$D = E + 1 - H$$

Since for many shapes in nature (mountains and clouds, for instance) $D \approx E + 0.2$, Hurst exponents near 0.8 are an excellent general choice for generating such natural designs. Today, the use of computer-generated fractal landscapes in movie animation is pervasive.

6

CHAPTER

Brownian Motion, Gambling Losses, and Intergalactic Voids: Random Fractals par Excellence

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

—ALBERT EINSTEIN

Some of the most fertile fields for fractals are fluctuating phenomena. In fact, nature abounds with self-similar structures that are statistical in nature, covering many different disciplines: from the distribution of galaxies in astronomy to cloud formation, climate, and the weather in meteorology; from polymerization and rusting in chemistry to the design, in biology, of our lungs and vascular systems and the growth patterns of many plants; from "fingering" in oil exploration, the branching and drainage basins of river systems, and the occurrence of floods in geophysics to physics proper, where we encounter fractals and statistical self-similarity in Brownian motion, fracture surfaces, soap bubbles, coagulation, percolation, diffusion-limited aggregation, and dielectric breakdown—such as lightning and Lichtenberg figures—not to mention the energy valleys in *spin glasses* and, last but not least, *turbulence.*

Cosmic strings, too, those wispy threads thought to have been created during the birth of the universe and potentially responsible for the clumping of galaxies, are statistically self-similar as the universe expands [Vil 87].

In view of this encompassing scope of statistical self-similarity, it is perhaps not surprising that fractals have even invaded the art industry. Following Mandelbrot's pioneering work and his specific suggestions, mountains and other

"backdrops" in videos, movies, and images in general are increasingly being generated by computers, programmed to produce self-similar structures with the desired pleasing statistics—pleasing, it is said, because fractals have interesting features *on many size scales.*

In this chapter we shall first explore the random fractal par excellence: Brownian motion and certain games of chance.

The Brownian Beast Tamed

Brownian motion, the paradigm of random fractals, was first observed in the nineteenth century by the Scottish botanist Robert Brown (1773–1858) and properly described by him in 1827 as a *physical*[1] phenomenon.

Thus, the stage was set for mathematical physics to step in, and it was none other than Albert Einstein, in 1905, and Marian Smoluchowski (1872–1917), a bit later, who first shed light into a very murky situation [Pai 82]. Interestingly, when Einstein began thinking about random thermal motions of macroscopic objects, he was not even sure that there was such a thing as Brownian motion.[2] But he felt that there *should* be macroscopic manifestations of molecular motion and that their observation would confirm the molecular theory of heat, thereby proving the existence of finite-size atoms. This is what Jean Baptiste Perrin (1870–1942) proceeded to do, and in 1926 he was awarded the Nobel Prize in physics for his work on Brownian motion. In fact, proper observation of the motion under the microscope and an application of the law of large numbers allowed Perrin, following Einstein's suggestion, to "count" the number of molecules in a given volume.

But the headaches that Brownian motion had caused were far from over. Now the puzzling questions arose in *mathematics,* owing to the nondifferentiability of the motion, until laid to rest by Norbert Wiener (1874–1964) and others. So, to mathematical physicists, Brownian motion is now known as a Wiener process—really nothing new or novel, considering that Weierstrass functions had been available for some time as a good mathematical model for nondifferentiable continuous functions. In fact, it was one of the great minds of the nineteenth century, Ludwig Boltzmann (1844–1906), who felt that there were physical

1. This is not the place to retell the hilarious tales stimulated by that wiggly motion seen under the microscope, the fanciful interpretations ranging from living molecules—endowed with their own free will—to the outright supranatural. Suffice it to say that, when Brown had the liquid boiled, frozen, and reheated, the little specks still wiggled as madly as many a modern disco crowd.

2. In his first paper on the subject Einstein wrote: "It is possible that the motions discussed here are identical with the so-called Brownian molecular motion; the references accessible to me on the latter subject are so imprecise, however, that I could not form an opinion about this,"[Ein 05].

problems that are best described by nondifferentiable functions and that one could have invented such functions from the proper consideration of physical problems.[3]

Brownian Motion as a Fractal

Figure 1A shows a typical way of portraying the Brownian motion of a dust particle, say, as seen under the microscope. However, this portrayal is exceedingly misleading: Does the particle really move in straight lines between vertices? No! Does it move in curved lines then? No again! Then how *does* the particle move from point A to point B in Figure 1A?

Let us photograph the particle's motion at a shutter speed 100 times faster so that we see the particle 100 times between A and B. The result, magnified 10 diameters, is shown in Figure 1B: the straight line connecting A with B has metamorphosed into 100 straight-line segments, each, on average, about as long as the segments in Figure 1A (really 10 times shorter, because Figure 1B is magnified tenfold).

Does the particle move in a straight line from C to D in Figure 1B? Once more the answer is no. Looking again 100 times more often and magnifying 10 diameters will result again in a picture *statistically similar* to Figure 1B. That is why we call the Brownian motion *statistically self-similar*.[4] Every time we make our spatial resolution 10 times higher, we get 100 times as many pieces. In general, if we increase the spatial resolution by a factor of $1/r$, we get $N(r) \sim 1/r^2$ more pieces to cover. Hence, the Hausdorff dimension (see pages 9–10 in Chapter 1) of Brownian motion is given by

$$D_H = \frac{\log N(r)}{\log (1/r)} = 2 \qquad (1)$$

which happens to be an integer. With $D_H = 2$, Brownian motion in two dimensions could be plane-filling, but it is not; there is much self-overlap (see Figure 1A and B). In fact, for a Brownian motion in two dimensions (think of enzymes wandering around on the surface of a cell), the probability of returning to the neighborhood of a given location, no matter how narrowly defined, is 1. By contrast, for Brownian motion in three dimensions, the embedding dimension 3

3. In a letter dated Vienna, January 15, 1898, to Felix Klein (1849–1925).

4. More accurately, geometric figures whose parts can be brought into correspondence with the whole by scaling different directions by different factors are called *self-affine*. Thus, Brownian motion is statistically self-affine, with a scaling factor r, say, in the spatial dimensions and a scaling factor r^2 in the time dimension.

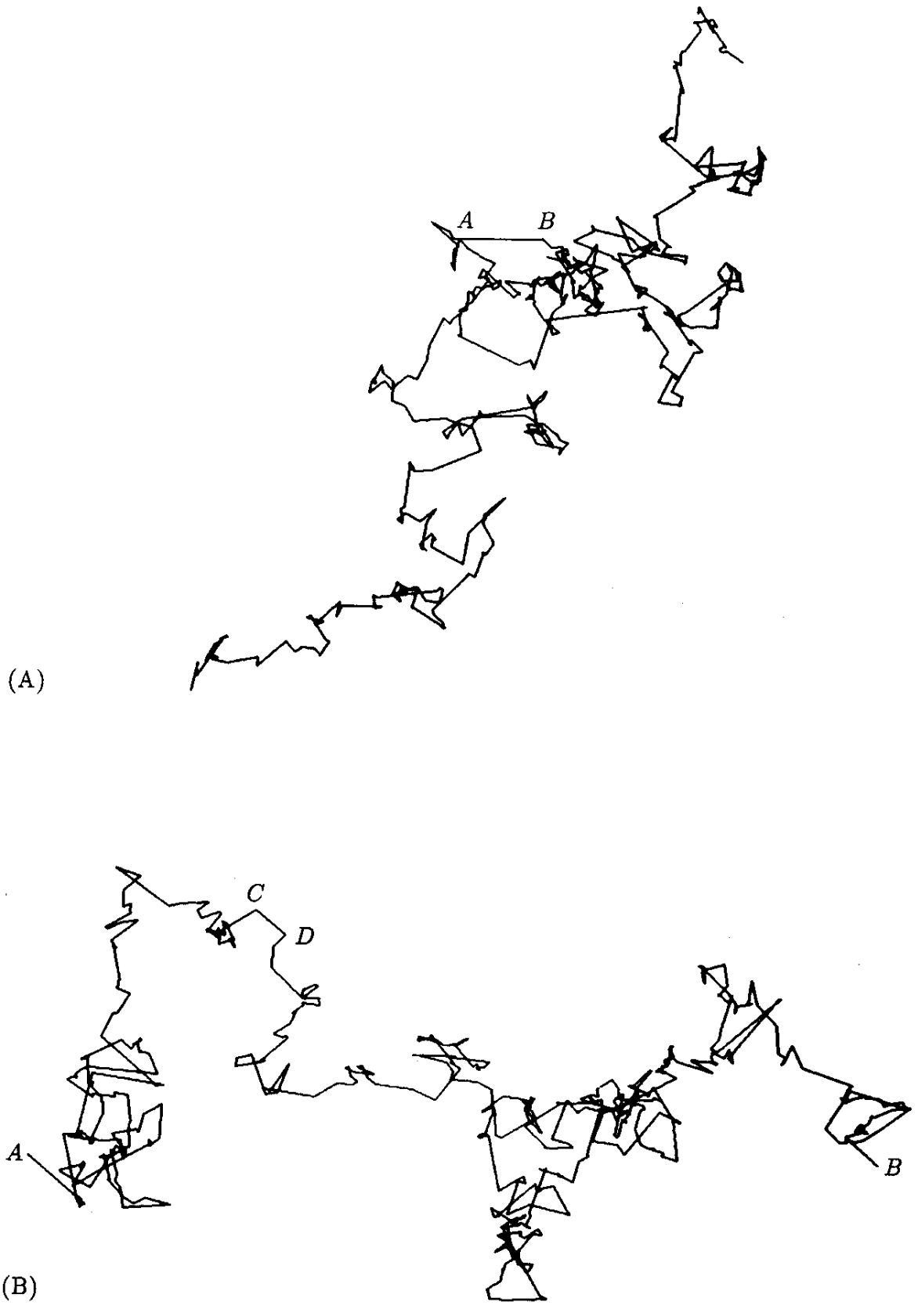


Figure 1 (A) Brownian motion. (B) The segment *AB* of the Brownian motion in part A sampled 100 times more frequently and magnified 10 times.

being larger than the Hausdorff dimension of the motion ($D_H = 2$), the return probability is smaller than 1.[5]

As we remarked earlier, every self-similar process in the real world has to have a largest and a smallest scale; scaling up or down cannot go on forever. But for Brownian motion, the range of lengths over which self-similarity prevails covers many powers of 10: from the size of the vessel containing the liquid, 0.1 m, say, to the mean free path between molecular impacts, which, for small test *par*ticles, could be as small as 10^{-9} m. In many situations we are willing to call an object self-similar if it scales over a range of just 10 to 1 or even less, in perhaps as few as three discrete steps. By contrast, Brownian motion scales over a range of 10^8 to 1, covered by a *continuum* of intermediate scales.

Brownian motion is as close as we get in physics to a *nondifferentiable* function. And, as Boltzmann remarked so soberly in his previously cited letter to Klein (of *Klein bottle* fame), if [Weierstrass] had not already thought of such functions (in his attempt to show the world how totally counterintuitive something innocently called a *function* could be) then physicists, or botanists, would have had to invent the strange mathematical beast themselves.

How Many Molecules?

The physical law behind the scaling behavior that leads to equation 1 is the *diffusion equation*

$$\overline{x^2} = 2Dt \tag{2}$$

where $\overline{x^2}$ is the mean square displacement of a Brownian particle in time t. Equation 2 is simply an expression of the mathematical fact that if independent random lengths (whose average is zero and whose distribution has a second moment) are added, then the total distance is obtained by adding up the *squared* individual lengths and then taking the square root. We have called the constant of proportionality $2D$ in equation 2, following Einstein's original nomenclature.

The so-called *diffusion constant* D must be related to the "microscopic" variables (really, the variables the botanist can*not* see under the microscope), namely, the mean free path λ and the average time between collisions, τ. A "little thinking" will reveal that the relationship is as follows:

$$2D = \frac{\lambda^2}{\tau} \tag{3}$$

5. This is believed to be an important reason why Mother Nature lets many crucial life-sustaining chemical reactions occur on *surfaces* rather than in three-dimensional space.

or, introducing the thermal velocity $v \approx \lambda/\tau$ of the observed particle,

$$2D = \overline{v^2}\tau \tag{4}$$

where $\overline{v^2}$ is the mean square velocity. Now, the thermodynamic equipartition of energy tells us that

$$\overline{v^2} = 3\frac{kT}{m} \tag{5}$$

where k is Boltzmann's constant, T is the absolute temperature, and m is the mass of the particle.

The average time between collisions is given by

$$\tau = \frac{4}{\bar{v}nF} \tag{6}$$

where n is the *number of buffeting molecules* per unit volume and F the cross section of the buffeted particle (which, like m, can be measured macroscopically).

Now, measuring $\overline{x^2}$ and putting equations 2 to 6 together, allows us to determine the number of molecules n in a given volume. If the result is finite, as it will be, then there must be a finite number of molecules in the liquid. And since the total weight is also known, measuring Brownian motion under the microscope allowed Perrin to measure the weight of an individual molecule. To have foreseen this possibility of establishing the finite, nonvanishing reality of atoms and molecules is one of Einstein's greatest contributions to our understanding of the physical world in which we live—almost on a par with his redefinitions of time and space in special and general relativity.

The Spectrum of Brownian Motion

What is the *power spectrum* of a Brownian function $B(t)$, which we define as the projection of Brownian motion onto one spatial dimension as a function of time? Brownian motion is generated by independent increments (the individual impacts of the buffeting molecules), which have a flat ("white") power spectrum. The sum or integral of the increments therefore has a power spectrum that is proportional to f^{-2}. Noises having such spectra are now called *brown* noises (see Chapter 5)—an allusion also to the fact that brown light has a stronger admixture of red light (low optical frequencies) than white light.

The Gambler's Ruin, Random Walks, and Information Theory

Other instances of brown noise are stock market prices and gains or losses from other *jeux d'hasard* and, generally, random walks subject to independent increments. Let us look at the infamous paradigm called the "gambler's ruin." Each time a coin is thrown and heads come up (with probability p), the player wins a dollar, and he loses a dollar when tails shows. His capital as a function of time (number of throws) is a Brownian process with fixed increments, also called a Markov-Wiener process, after A. A. Markov (1903–1922) and Wiener.

Let the player start with an initial capital K (K as in Karl Marx and Das Kapital). After the first trial (apt word!) his capital is either $K + 1$ (with probability p) or $K - 1$ (with probability $q = 1 - p$). Calling his probability of ultimate ruin q_K, we have the following difference equation:

$$q_K = pq_{K+1} + qq_{K-1} \qquad 0 < K < B \tag{7}$$

where B is the capital of the bank, $q_0 = 1$, and $q_B = 0$.

Such difference equations can be solved by the generating-function (or z-transform) *ansatz* $q_K = z^K$, which yields a quadratic equation in z: $z = pz^2 + q$, with the two solutions (for $p \neq q$) $z = 1$ and $z = q/p$. Thus, for $p \neq q$, with the particular solutions $q_K = 1$ and $q_K = (q/p)^K$, the general solution is

$$q_K = a + b\left(\frac{q}{p}\right)^K$$

or, with the "boundary conditions" $q_0 = 1$ and $q_B = 0$,

$$q_K = \frac{(q/p)^B - (q/p)^K}{(q/p)^B - 1} \qquad p \neq q \tag{8}$$

For $p = q$, we apply l'Hospital's rule to the limit as $p \to 0.5$, yielding for the probability of ultimate ruin the simple result

$$q_K = 1 - \frac{K}{B}$$

which satisfies the necessary symmetry for $p = 0.5$, namely, $q_K + q_{B-K} = 1$. Of

course, for the poor gambler whose capital K is small compared with that of the bank B, ultimate ruin is almost certain:[6] $q_K \approx 1$.

Counterintuition Runs Rampant in Random Runs

One of the several counterintuitive facts of the fair-coin tossing game is that the expected number of times equals 1 that a player will increase his capital by any given amount G before the first return to his initial capital—*independent of how large* the gain G is! More tangibly, in a \$1 per bet fair-coin tossing game, the player will reach \$1 million on the average once before he has incurred any loss (i.e., dropped below his initial capital). The only consolation for the bank is that the *expected* gain is still 0. (As we saw in the preceding section, for the game without a time limit, the probability that the player will lose his entire capital of $K < B$ dollars is $1 - K/B$, and the probability that the bank will go broke is K/B).

A similarly counterintuitive result concerns the *duration* D_K of the fair game, obtained by solving a difference equation similar to equation 7:

$$D_K = K(B - K)$$

In other words, if the player has just a single dollar and the bank \$1 million, the expected duration of the game is 999,999 tosses! As we shall see in what follows, these mind-boggling conclusions are related to the fact that returns to a given position in the unconstrained ($B = \infty$) fair game, while occurring with probability 1, are Cantor sets with fractal dimension $\frac{1}{2}$.

The counterintuitive results just described are symptomatic for unbiased random walks, expressed most forcefully by the so-called *arcsine law*. Couched in the language of a discrete random walk of a diffusing particle in one spatial dimension, the arcsine law says: The probability $p_{2n}(2k)$ that in the time interval from 0 to $2n$ the particle will spend $2k$ time units on the positive side is given by

$$p_{2n}(2k) = \binom{2k}{k}\binom{2n-2k}{n-k} 2^{-2n}$$

$$\approx \frac{1}{\pi[k(n-k)]^{1/2}} \tag{9}$$

6. Surprisingly, a certain statistics *professor*, who shall remain nameless, was not able to derive the formula for the probability of ruin for the finite-duration game which he had posed as an exercise for the *students* at the beginning of the summer semester in 1948. During the entire semester his recurrent refrain was "we'll tackle that one next week." But he never did tackle it.

[Fel 68], a distribution which has most of its weight near $k = 0$ and $k = n$ (just like the amplitude distribution of a sine wave oscillating between its extreme values, a fact on which time-averaging holography of vibrating bodies is based).

The arcsine law, so named because the integral of equation 9 is the arcsine function, has many curious consequences. For example, for a fair game, the probability that in 20 tosses each player will lead 10 times (which seems fair enough) is only about 6 percent. But the probability that one of the players will lead for *all* 20 tosses (how unfair!) is greater than 35 percent! In other words, in more than one-third of the cases the lead never changes!

More Food for Fair Thought

Another "unfair" result of fair-coin tossing is as follows. Consider $2n$ tosses, half of which turn out heads (and half tails). And let $2k$ again measure how often the accumulated number of occurrences of heads is greater than the accumulated number of occurrences of tails. Then the number of possibilities $N_{2n}(2k)$ for this outcome is given by the "Catalan number:"

$$N_{2n}(2k) = \binom{2n}{n} \cdot \frac{1}{n+1}$$

independent of k.[7]

This kind of counterintuitiveness has led to numerous false conclusions in the history of science, and statistics in particular. In 1876 Sir Francis Galton (1822–1911), inventor of the Galton board (a kind of fakir's bed for bouncing balls), tested some data on plants furnished him by the even more famous Charles Darwin (1809–1882). There were 15 treated plants and 15 untreated specimens (the control group). In rank-ordering the data, Galton saw that the treated plants were ahead of the untreated plants with the same rank in 13 out of 15 cases. Galton concluded, understandably, that the treatment was effective. But assuming perfect randomness in the data (30 measurements from the *same* pool of plants), the probability of Galton's observation is $\frac{3}{16}$. In other words, in 3 out of 16 cases a perfectly ineffectual treatment appears very effective. How many bad answers and bogus inferences have been drawn from this single source of statistical misdemeanor alone?

7. Catalan numbers, although perhaps not widely known, are truly ubiquitous. For example, N_{2n} is the number of ways $2n$ people seated around a table can shake hands in n pairs without their arms crossing. For many other applications of Catalan numbers, see the paper by Eggleton and Guy [EG 88].

The St. Petersburg Paradox

Games of chance have led to more than one paradox, most often related to the counterintuitive aspects of random walks and their inherent fractal nature.

Around 1700, Nicolas Bernoulli (1687–1759), nephew of Jakob (1654–1705) and Johann (1677–1748), introduced a curious game of chance with infinite mean winnings (*mean* certainly from the bank's point of view). The game was analyzed by still another Bernoulli,[8] Daniel (1700–1782), in the *Commentarii* of the St. Petersburg Academy—depository of much of Euler's writings.

Suppose a coin has a probability $p > \frac{1}{2}$ of coming up heads. The player flips the coin until heads comes up for the first time. If it takes n flips, the player wins 2^{n-1} dollars. What are his expected winnings W if the game can continue indefinitely? The answer is simple enough:

$$W = 2^0 p + 2^1(1-p)p + 2^2(1-p)^2 p + \cdots \tag{10}$$

or

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

For $p = 0.55$, for example, the expected gain is $W = 5.5$ dollars, and for $p = 0.51$ it is $W = 25.5$ dollars.

What happens for a "fair" coin? For $p = \frac{1}{2}$, the geometric series (equation 10) does not converge, and the expected winnings become infinite! Thus, a fair fee for the game would be an infinite ante, or so a bickering banker could reason.

But a prudent player, quite apart from being "temporarily out of" infinite capital, would consider the infinite *mean* winnings promised by equation 10 quite unfair and would prefer to base his ante on the finite *median* winnings, namely, a single dollar. Banker and player are at odds—much more so than even they would have surmised!

How can mean and median be so different? The answer is, of course, that for $p = \frac{1}{2}$, the mean does not even exist, and what does not exist lacks the ability of being different.

The divergent mean winnings may be reminiscent of the infinite length of a fractal curve, and indeed we can tame the St. Petersburg paradox by introducing fractals and Hausdorff dimensions, as we shall see in subsequent sections.

8. A parallel between the abundant Bernoullis and the shingle of a typical American law firm is drawn in Figure 2.

Figure 2 Bernoulli, Bernoulli, Bernoulli & Company [Har 77]. (© 1991 by Sidney Harris)

Shannon's Outguessing Machine

Not all games of chance are fair, perhaps least of all those which (who) proclaim fairness the loudest ("I am not a crook"). But some games make no pretensions of fair play; in fact, *un*fairness is their very reason for being—such as Claude Shannon's engaging "outguessing machine" [Sha 53].

Shannon's entrapping contraption initially makes random heads-tails choices against a human contender. But once the machine has experienced its first win, it begins to analyze the opponent's "strategy" to a depth of two throws. Does he or she change after losing a throw? Does the player keep on choosing tails if tails has brought two previous wins? Or does the gambler get chary and head for heads next? For most people such strategies are mostly subconscious, but the machine assumes the human to act like a second-order Markov process and uncovers the underlying transition probabilities without fail. Exploiting these, the machine always wins over the long haul, except against its creator. Shannon, keeping track of his machine's inner states, can beat it 6 times out of 10. Of course, anyone could win 5 out of 10 throws on average by playing random (perhaps by flipping a true coin). But this is precisely what people, deprived of proper props, are incapable of doing, as Shannon's machine has demonstrated again and again by beating a wide variety of human would-be winners. Specifically, man's mind appears to abhor long strings of like outcomes—as occur perfectly naturally in truly random sequences.

Of course, the machine can have bad luck too, especially in its initial guessing phase. I once wanted to show off the machine's prowess to a foreign friend (mathematician Fritz Hirzebruch) visiting Bell Laboratories. As luck would have it, Hirzebruch won 13 times in a row before his first loss. But thereafter the machine took off with a vengeance, overtaking the renowned mathematician on throw 31 (i.e., the machine won 16 out of the next 18 throws!) and never fell behind again—in spite of the fact that Hirzebruch had been told (in general terms) how the machine worked.

The Classical Mechanics of Roulette and Shannon's Channel Capacity

While Shannon's outguessing machine is intrinsically (and intentionally) unfair, some ostensibly fair games of chance can be subverted into unfairness. Great fortunes (and some unfortunate lives) have been lost through gambling at roulette while banking on a bewildering mélange of "infallible" stratagems. In the end the bank always wins, because, for an "even" bet, its odds are 19 to 18 (or even 20 to 18 at tables with a *double* zero, prevalent in American casinos).

Yet this inveterate observer of the gambling scene thought that simple mechanics—classical, not quantum—allied with fast, real-time computation could moderate the odds in his favor. Clad in a brand-new tuxedo and armed with an array of stopwatches hidden in his pockets, he set sail for some of the world's best known casinos: Baden-Baden in the Black Forest (parent to Monte Carlo), Baden near the Vienna Woods, and Évian-les-Bains on Lake Geneva. These field trips, undertaken in the early 1960s, confirmed what I had first found on a full-

size regulation roulette wheel lodged in my basement: at the time the average croupier cries *rien ne vas plus* (no more bets, *nichts geht mehr*), the final resting position of the ball is not totally unpredictable; its probability distribution is anything but exactly uniform. In fact, the modulation depth of the probability amplitude around the wheel is typically 10 percent. Thus, instead of losing on average 19 out of 37 even bets, one could *win* about 20 out of 37. For some casinos the winning chances are even greater, as, for example, in Évian, where an easy-going croupier at one table spun the wheel so slowly that the ball, once having left the upper rim, sailed down into its final slot without further hopping to and fro.

To reap the promised harvest, one first has to determine the friction coefficients of the ball and the wheel (usually quite small) and enter these parameters into a small portable computer. (A friend and I had built ourselves a special-purpose *analog* computer for this task. Now, a quarter century later, one would of course place one's trust in a *digital* chip to dislodge the bank's chips and steer them in the right direction [Bas 90].) The computer, through two pushbuttons, gets timing information signaling two successive passes of the ball and the zero of the wheel past a preselected mark on the rim. This fixes speeds and relative positions of wheel and ball. (Friction is determined in a prior measurement.) The most likely outcome computed from these data is communicated via a modified hearing aid to the prospective winner.

How does one maximize the expected rate of increase of one's capital knowing that the chances of winning, p, are better than even ($p > 0.5$)? How much of one's capital should one risk for each spin of the wheel? The answer comes from *information theory*, in fact, one of its earliest applications to gambling. In a landmark paper that we already mentioned on pages 128–129 in Chapter 5, John L. Kelly, Jr. [Kel 56], proved that, in order to maximize the rate of increase of one's capital, one should bet a fraction $2p - 1$ of the current capital. And of course, for $p \leq 0.5$ one should abstain completely and seek another pastime. The expected (exponential) growth of one's capital is then given by the factor $2^{C(p)}$, where $C(p)$ is Shannon's channel capacity of a *binary symmetric channel* with error probability p:

$$C(p) = 1 + p \log_2 p + (1 - p) \log_2 (1 - p)$$

Thus, for $p = 0.55$, for example, one should risk $2p - 1$ or 10 percent of one's capital for every spin and expect a rate of enrichment of $2^{C(0.55)} = 0.005$ or 0.5 percent *with each spin*. (This means that, during a weekend of 138 spins, one's holdings would be doubled—less tips and taxes, of course.)

This application of information theory was the first instance (and is still the only one, as far as I know) in which a benefit can be reaped without the elaborate coding that is necessary to realize the error-free transmission promised by Shannon's theory.

How much did I win? Well, once I knew that classical mechanics (and my analog computer) worked, I lost all interest in the project. Also, I knew that casinos can banish anyone who wins "too much" from their hallowed halls without having to state any reason. Finally, once the management learns the mechanics behind the winning method, it can simply instruct its croupiers to spin their wheels a bit faster and to call *rien ne va plus* a little earlier. End of dream, Eden lost.

The Clustering of Poverty and Galaxies

Certain perfectly simple statistical rules generate random collections of points, called *point processes*, that exhibit unexpectedly large voids with a statistical self-similar structure. The distribution of galaxies in the universe is a good case in point: the voids between the largest clusters of galaxies are only a few times smaller than the entire universe itself. And the voids between galaxies are as large as large clusters of galaxies, and so on. A similar structure also governs games of chance between successive ruins and numerous other events infested by holes. Let us look at a fair game of chance, with probability $p = 0.5$ of either winning or losing a yen. The current capital, $K(t)$, of the player as the game progresses may look somewhat like Figure 3A: the capital has a tendency to drift to large positive or negative values but eventually returns to 0. (The ruined player, who has lost all, is allowed to continue on credit.) Once the capital has reached 0, the probability that it will do so several times in a short time span

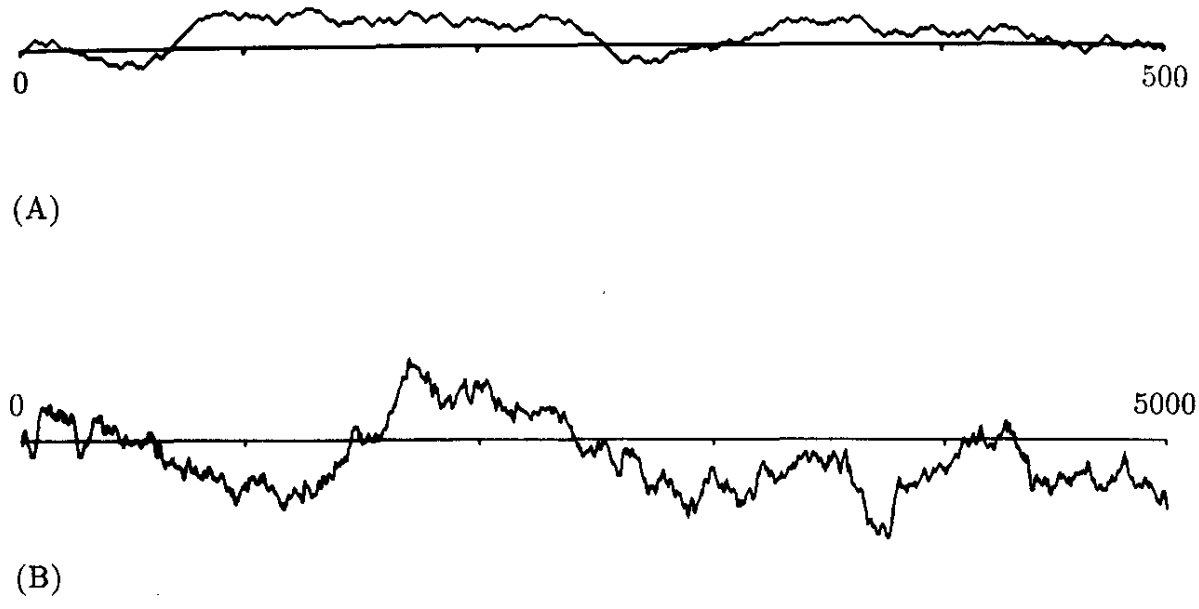


Figure 3 (A) Gambler's capital as a function of time in an honest game of chance. (B) Fluctuations of the gambler's capital over very long times.

is, for obvious reasons, very high. In other words, the zeros of the capital function $K(t)$ form clusters.

What else can we say about the function $K(t)$? Other than the sizes of the basic steps in time ($\Delta t = 1$) and money ($\Delta K = 1$), the problem has no further scales. We therefore expect the player's ruin to exhibit self-similarity and self-affinity. Indeed, if we plot $K(t)$ over a longer time span and scale t and K appropriately, the new plot (see Figure 3B) will look much like the old one. The proper scaling factor for K is the square root of the scaling factor for t, just as in Brownian motion.

The number of expected zeros, N_0, in the time interval t also scales with the square root of t [Fel 68]: $N_0 \sim t^{1/2}$. Waiting 4 times as long will increase the number of ruins only by a factor 2 (sounds good!).

What is the probability $p(z)$ of observing a given distance z between successive zeros? Since the problem, given infinite capital resources, contains no long-range cutoff, $p(z)$ must approach a self-similar power law, for long games:

$$p(z) \approx \text{const} \cdot z^{\alpha} \qquad 1 \ll z \leq t$$

Since N_0 is proportional to $t^{1/2}$, so is the mean gap length $z = t/N_0$. With $\bar{z} \approx \text{const} \cdot t^{\alpha+2} \sim t^{1/2}$, we have $\alpha = -3/2$ and, asymptotically,

$$p(z) \approx \text{const} \cdot z^{-3/2} \qquad 1 \ll z \leq t$$

From this we obtain the cumulative distribution of the lengths of the zero-free voids exceeding length z:

$$P(z) := \sum_{k=z}^{t} p(k) \approx \text{const} \cdot z^{-1/2} \tag{11}$$

This distribution has a very long tail that drops off to zero very slowly with increasing gap size. Figure 4 shows experimental results (obtained with a pocket calculator) that confirm equation 11 over five orders of magnitude.

Figure 5 shows the "voids within voids within voids" structure of the zero-free regions. Roughly one-half of every "cluster of zeros" is actually devoid of zeros! Going to the limit of a continuous time scale, the zeros form a rather thin dust indeed, a Cantor set with Hausdorff dimension equal to the negative of the exponent in the cumulative distribution equation (equation 11): $D_H = 0.5$. This value can also be derived from the value $D_H = 2.5$ for the surface of Brownian mountains (pages 134–136 in Chapter 5). A vertical cut through such a mountain produces a Brownian "profile" with $D_H = 1.5$, which corresponds to our $K(t)$. A further cut with the line $K = 0$ produces our zero set (i.e., the set of t values for which $K(t) = 0$) with $D_H = 0.5$. In general, lowering the *topological* dimension of a fractal by 1 by forming a zero set also reduces the *fractal* dimension by 1. For example, Brownian mountains with a surface $D_H = 2.2$ generate coastlines

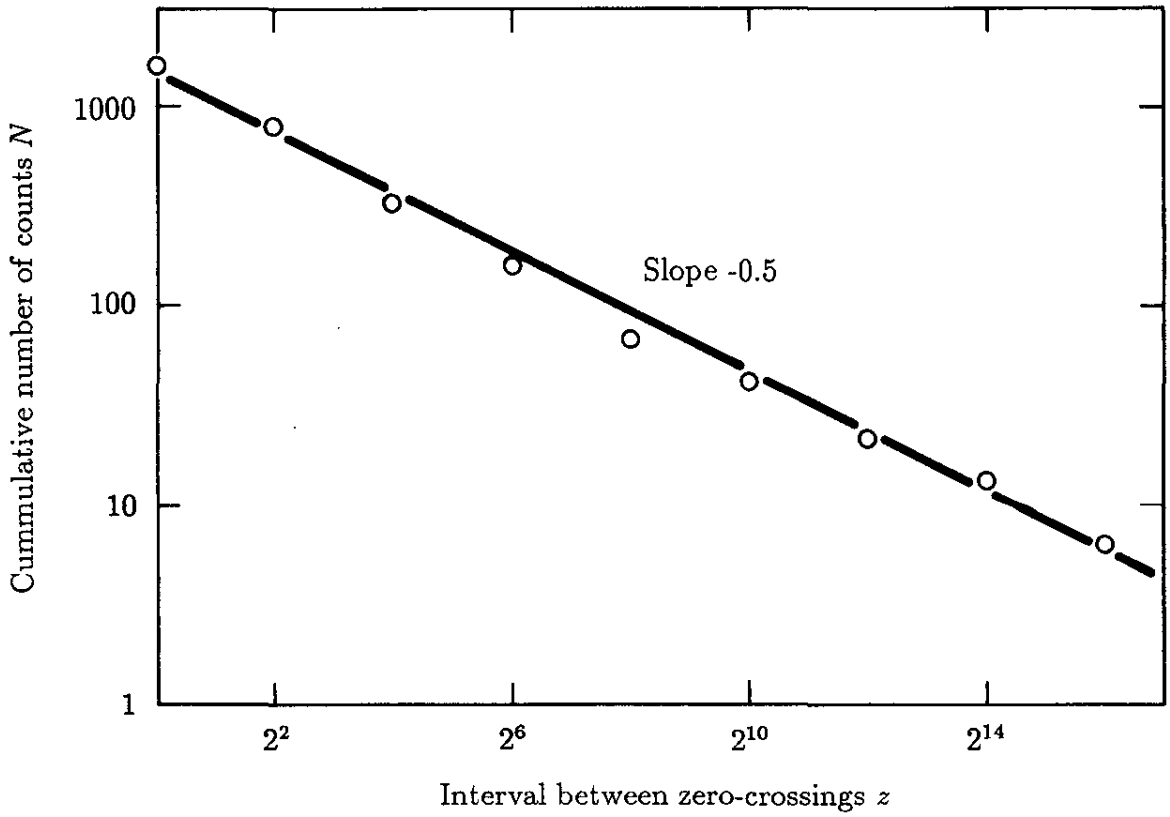


Figure 4 Distribution of time intervals without ruin.

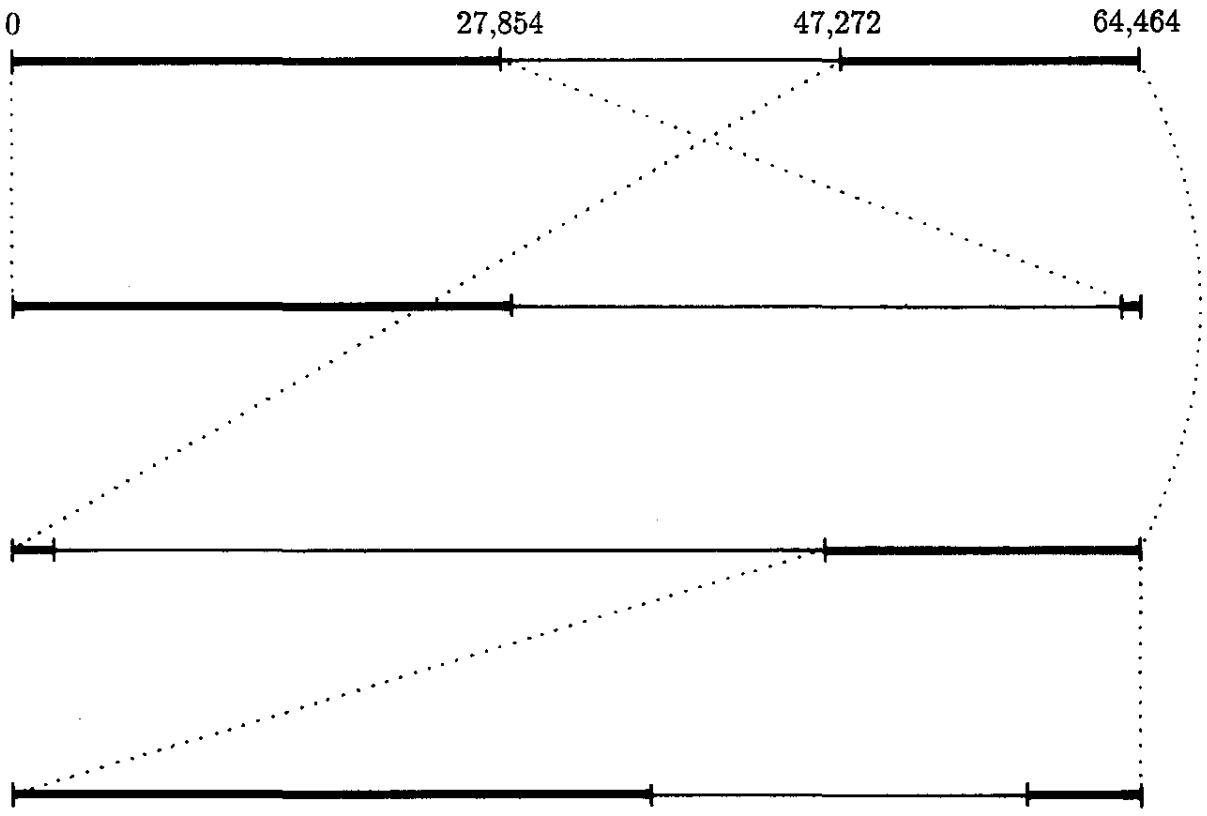


Figure 5 Ruin-free time intervals (shown by thin lines) occur on all scales.

(i.e., lines at zero elevation) with $D_H = 1.2$, similar to the Hausdorff dimension of the west coast of Britain.

Levy Flights through the Universe

Another way to generate such point processes as the zero set of the gambler and to generalize them to more dimensions is the so-called *Levy flight* [Man 83]. In a Levy flight, named after the French mathematician Paul Lévy (1886–1971), one strings together independent increments ("flight paths") whose lengths z are (cumulatively) distributed according to a homogeneous power law:

$$P(z) = \text{const} \cdot z^{-D} \tag{12}$$

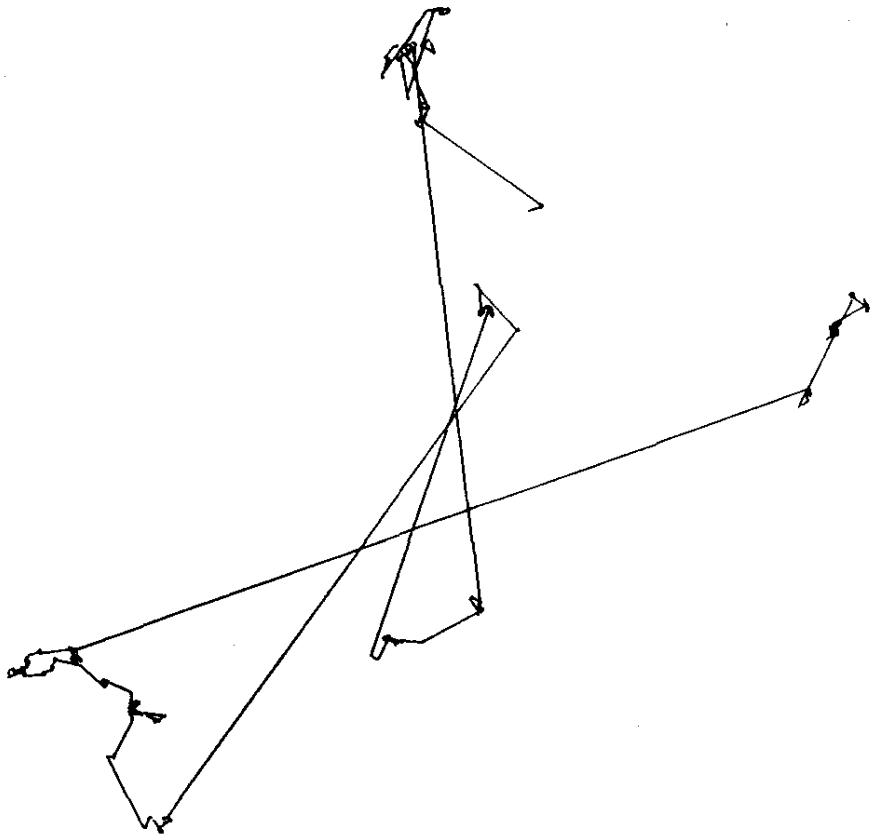
where D turns out to be the Hausdorff dimension of the resulting "dust." For $D = 0.5$ and one spatial dimension, equation 12 generates the voids in the gambler's ruin.

Figure 6A shows a *two*-dimensional isotropic Levy flight with exponent $D = 1.26$, making larger voids more probable than for $D = 0.5$. The turning points, that is, the "galaxies," generated by this process are shown in Figure 6B. For $D = 1.26$, the resemblance with the distribution of galaxies in the universe as seen from earth is astounding [Haw 88]. The implication, of course, is that the universe is a Cantor dust that has no natural length scales other than its own size. But the best-matching exponent, $D = 1.26$, for the galaxies and their soap-bubble-like aggregation still cries out for a proper explanation. One of the persistent puzzles in understanding the evolution—past, present, and future—of our universe is the mysterious role of dark matter, including black holes and, at the other end of the mass scale, the long-elusive but ubiquitous neutrino.[9]

Paradoxes from Probabilistic Power Laws

Probability distributions that follow a self-similar power law can have some rather paradoxical consequences. Consider a random variable $1 \leq x < \infty$ with a probability of exceeding a given value x given by x^{-D}. The *conditional* probability that the random variable exceeds the value x, given that $x > x_0$, equals $(x_0/x)^D$.

9. Enduring enigmas in the distribution of quasi-stellar objects ("quasars") at the far end of the universe have recently been interpreted as resulting from *gravitational lens* effects of dark matter surrounding "foreground" galaxies [BPS 87]. This optical effect, predicted by Einstein but believed by him to be unobservable, may yet turn out to be the most legible "fingerprint" of the dark universe. (The imaging properties of a gravitational lens are akin to those of the foot of a wine glass, producing multiple images.)



(A)

(B)

Figure 6 (A) Two-dimensional Levy Flight. (B) Resulting cluster of "galaxies."

For $D > 1$, the mean exists and equals $D/(D-1)$. The conditional mean, given that $x > x_0$, equals, innocently enough,

$$\bar{x}_{x_0} = \frac{x_0 D}{D-1} \tag{13}$$

As expected for a self-similar distribution, the conditional expectation depends linearly on x_0.

Now suppose that the completion times of a given human endeavor—writing a difficult report, for instance—are distributed according to a power law with an exponent $D = 1.5$, say. The expected completion time of the job is then $D/(D-1) = 3$ hours or days or whatever unit of time one chooses. Intuitively, one would expect, 5 days after starting the work and not having completed it, that the expected completion time would be considerably less than 3 days—after all, it was only 3 days at the *start* of the project. However, equation 13 tells us that the expected completion time is now 15 days. And after another 60 days without finishing the job, the expected completion time has moved on to 180 days! In other words, the longer one works on such a project without actually concluding it, the more remote the expected completion date becomes.

Is this really such a perplexing paradox? No, on the contrary: human experience, all-too-familiar human experience, suggests that in fact many tasks suffer from similar runaway completion times. In short, such jobs either get done soon or they never get done. It is surprising, though, that this common conundrum can be modeled so simply by a self-similar power law. In fact, it might not be a bad idea to classify such dragging jobs by their characteristic exponents D—and pay the laggard contractor accordingly. For $D = 2$, $\bar{x}_{x_0} - x_0$ equals x_0, which is characteristic of the proverbial undertaking whose completion is always promised for *mañana*.

Invariant Distributions: Gauss, Cauchy, and Beyond

The sum of two Gaussian random variables is another Gaussian variable. The Gaussian distribution is therefore said to be *invariant under addition*. The variance σ^2 of the sum variable equals the sum of the individual variances σ_1^2 and σ_2^2:

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \tag{14}$$

The invariance of the Gaussian distribution is intimately connected with the *central limit theorem* of probability theory, which states that the suitably normalized sum of many independent random variables with *finite* variances converges on a Gaussian distribution.

For the summation rule in equation 14, the Gaussian distribution is the only one that is invariant under addition. However, if we introduce a *general* exponent D and replace the (possibly diverging) standard deviations σ by some other measure s of the width of a distribution that is guaranteed to exist (such as the interquartile range), we have instead of equation 14 the more general rule

$$s^D = s_1^D + s_2^D \tag{15}$$

Are there distributions that are invariant under addition with exponents D other than 2? There are indeed, and the solutions are related to the self-similar power-law distributions that we encountered in the preceding sections.

For the exponent D equal to 1, the invariant distribution is the bell-shaped *Cauchy* density

$$p(x) = \frac{1}{\pi(1 + x^2)} \tag{16}$$

which has the same functional form as the intensity resonance (the "resonance line") of a linear oscillator as a function of frequency.

The Cauchy distribution, named after the French mathematician Augustin Louis Cauchy (1789–1857), is the source of several noteworthy paradoxes. Its mean and variance do not exist, because the corresponding integrals diverge. One therefore characterizes the Cauchy distribution by its *median* and its *interquartile range*. The median is the value for which the distribution integrated from x to ∞:

$$P(x) := \int_x^\infty \frac{dy}{\pi(1 + y^2)} = \frac{1}{2} - \frac{1}{\pi} \arctan x \tag{17}$$

equals $\frac{1}{2}$. Here the median equals 0. The interquartile range is the difference of the two x-values for which $P(x)$ equals $\frac{3}{4}$ and $\frac{1}{4}$, respectively. For equation 16 it equals 2.

The Fourier transform ("characteristic function") of the Cauchy distribution (equation 16) is the symmetric exponential function $\exp(-|t|)$. (Once we recognize the Cauchy distribution as having the shape of a resonance line, the exponential is, of course, to be expected, because the amplitude of the oscillation of a linear resonator decays exponentially.)

Since adding two random variables means convolving their probability distributions or multiplying their Fourier transforms, we see that the sum of two variables distributed according to equation 16 has the Fourier transform $\exp(-2|t|)$. The corresponding probability density is thus the same as equation 16, but with the abscissa x stretched by a factor 2. More generally, the width s of the distribution of the sum of two Cauchy variables is the sum of their individual widths:

$$s = s_1 + s_2$$

Thus, for the Cauchy distribution, the exponent D in equation 15 is indeed equal to 1. The underlying reason for this linear scaling behavior is that the Cauchy distribution has an exponentially decreasing Fourier transform and that, upon multiplying two exponentials, their arguments *add linearly*.

As a result of this linear scaling, the distribution of the *average* of N identically distributed Cauchy variables is the same as the original distribution. Thus, averaging Cauchy variables does not improve the estimate; averaging begets no benefit. This is in stark contrast to all probability distributions with a *finite variance* σ^2, for which averaging over N variables reduces the uncertainties by a factor $1/\sqrt{N}$. This nonstandard behavior of the Cauchy distribution is a consequence of its weakly decaying "tails" that produce too many "outliers" to lead to stable averages.

The reader is encouraged to sample this paradoxical behavior by simulating the averaging of Cauchy variables on a pocket calculator or home computer. In general, a random variable with an integrated distribution $P(x)$ is obtained from a *uniform* variable u in the interval [0, 1], which is available on many calculators, by inverting the equation $u = P(x)$. With equation 17, this gives the "recipe"

$$x = \tan [\pi(0.5 - u)]$$

to produce a Cauchy variable x from a uniform variable u.

The Cauchy distribution occurs in numerous practical applications. For example, the *ratio* of two independent identically distributed zero-mean random variables is Cauchy-distributed. In other words, any two-dimensional isotropic distribution $p(x, y)$ centered on the origin has a Cauchy-distributed ratio x/y or y/x. This statement also implies that if a variable is Cauchy-distributed, so is its reciprocal. Thus, the logarithm z of a Cauchy variable also has a symmetric distribution: namely, $1/(\pi \cosh z)$, an important function in several branches of physics. (Light beams with a $1/\cosh z$ profile in time or space lead to *solitons* in optical fibers, for example. Sound velocity profiles in the ocean that vary according to $1/\cosh z$ with depth result in self-focusing and thereby engender low-loss transmission of acoustic energy over intercontinental distances.)

As we have seen, the exponent $D = 1$ in equation 15 for the Cauchy distribution follows directly from the fact that its Fourier transform is an *exponential* function, $\exp(-|t|)$, with a *linear* dependence of its argument on the Fourier variable t. Similarly, the exponent $D = 2$ for the Gaussian distribution results from the fact that its Fourier transform is an exponential function, $\exp(-t^2)$, with a *quadratic* dependence on the Fourier variable t. Following Cauchy, we therefore surmise that the inverse Fourier transform of $\exp(-|t|^D)$ will lead to a random variable that has an invariant distribution under addition with an exponent D. This is indeed the case in the range $0 < D \leq 2$. (For $D > 2$, Cauchy's prescription gives *verboten* negative probability values.)

Thus, the cherished Gaussian distribution with $D = 2$ stands revealed as but an extreme, albeit ubiquitous, member of an entire clan of distributions. And

instead of just *one* central limit theorem, there are many. Depending on the scaling exponent D in equation 15, a properly normalized sum of random variables will converge on a specific limiting distribution that is invariant under summation of random variables.

For example, for $D = \frac{1}{2}$, the invariant distribution is

$$p(x) = \frac{1}{\sqrt{2\pi}}\, x^{-3/2} \exp\left(-\frac{1}{2x}\right) \tag{18}$$

which is the probability density that a Brownian noise function $B(t)$ starting out at the value 0, that is, $B(0) = 0$, equals 0 again in the interval $x \leq t \leq x + dx$.

The fact that a variable distributed according to equation 18 scales with an exponent $D = \frac{1}{2}$ is immediately obvious from the distribution integrated from x to ∞, which yields $P(x) = \text{erf}\,(1/\sqrt{2x})$. The interquartile range $s = 9.12$ of this distribution can be obtained from tables of the error integral [JE 51]. Because the Fourier transform of $p(x)$, equation 18, is of the form $\exp(-|t|^{1/2})$, adding two independent random variables changes the Fourier variable t to $4t$ and therefore the random variable x to $x/4$. The interquartile range for the sum of two variables is therefore equal to $4s = 36.5$. Thus the *average* of two independent random variables of this sort has an interquartile range that is *twice as large* as that of an individual variable. In general, the average of N independent variables distributed according to equation 18 has a width that is *increased* by a factor N– instead of being reduced by a factor $1/\sqrt{N}$ as in standard cases. No wonder that this kind of statistics is sometime called *non*standard. Yet the world harbors a lot more nonstandard statistics than many experts innocently expected.

For large x, the distribution $p(x)$ according to equation 18 is proportional to $x^{-3/2}$ and the integrated distribution is proportional to $x^{-1/2}$, that is, x^{-D}. More generally, for all invariant distributions other than the Gaussian, the integrated distribution is asymptotically proportional to x^{-D}, where D is the exponent in equation 15. Conversely, since we already know (see pages 155–156) that Levy flights, with integrated distributions following a power law x^{-D}, generate fractal sets with Hausdorff dimension D, we recognize the exponent in the equation $s^D = s_1^D + s_2^D$ as a bona fide fractal dimension for this geometric realization of the distribution. However, such exponents do not generally have the significance of a fractal dimension.

7

CHAPTER

Cantor Sets: Self-Similarity and Arithmetic Dust

Eine Menge stelle ich mir
vor wie einen Abgrund.
(I imagine a set to be an abyss)

—GEORG CANTOR, ca. 1888,
as related by Emmy Noether

In this chapter we shall further pursue one of the most important sources of self-similarity: Cantor sets. Originally constructed for purely abstract purposes, Cantor sets have of late turned into near perfect models for a host of phenomena in the real world—from strange attractors of nonlinear dynamic systems to the distribution of galaxies in the universe.

A Corner of Cantor's Paradise

Most numbers in the continuum cannot be
defined by a finite set of words.
—MARK KAC

In the midst of the animated debates during the nineteenth century on the foundations of mathematics—and the very meaning of the concept of *number*—George Cantor (1845–1918) wanted to present his colleagues with a set of numbers between 0 and 1 that has measure zero (i.e., a randomly thrown "dart" would be very unlikely to hit a member in the set) and, at the same time, has *so* many members that the set is in fact uncountable, just like *all* the real numbers between 0 and 1.

Many mathematicians, and even Cantor himself for a while, doubted that such a "crazy" set existed[1]—but it does exist, and its construction is in fact quite straightforward. Imagine the real line between 0 and 1 (drawn with chalk on a blackboard, if you will) and wipe out the open middle third, that is, the interval from $\frac{1}{3}$ to $\frac{2}{3}$, excluding the endpoints $\frac{1}{3}$ and $\frac{2}{3}$. Next erase the open middle third of each remaining third, and so forth *ad infinitum*. The result of the first seven erasures was illustrated in Figure 10 in Chapter 1 (page 16), but there is no way to draw the final result, aptly called Cantor *dust* by Mandelbrot [Man 83]. In fact, the Cantor dust has holes *on all scales*[2]: No matter how powerful the "microscope" that we use to inspect the set, we always see holes; there is not a single continuous interval in the entire unit interval, but only isolated points. No Cantor number has another Cantor number as an immediate neighbor. The Cantor dust is totally discontinuous—yet infinitely divisible, just like a continuum. As much as it would have amazed the ancient Greeks, there is no fundamental antinomy or philosophical contradiction between the discontinuous (like matter composed of atoms) and the infinitely divisible; the Cantor dust is both.

Formally, a Cantor set is defined as a set that is *totally disconnected, closed,* and *perfect*. A totally disconnected set is a set that contains no intervals and therefore has no interior points. A closed set is one that contains all its boundary elements. (A boundary element is an element that contains elements both inside and outside the set in arbitrarily small neighborhoods.) A perfect set is a nonempty set that is equal to the set of its accumulation points. All three conditions are met by our middle-third–erasing construction, the original Cantor set.

In spite of its counterintuitive nature, there is a neat number-theoretic way to represent the Cantor dust, namely, by *ternary* fractions employing the digits 0, 1, and 2. For example, $0.5 = \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots$ corresponds to $0.111\ldots$ in the ternary notation. Representations in the ternary system, like those in the decimal system, are not unique. For example, $\frac{1}{3}$ can be written either as 0.1 or $0.0\overline{2}$, where the bar over the 2 stands for an infinite sequence of 2s. One way to make such number representations unique is to outlaw terminating fractions. Thus, writing 0.1 for $\frac{1}{3}$ is illegal—one *has* to write $0.0\overline{2}$ for $\frac{1}{3}$.

Here we shall adopt the following convention for making ternary fractions unique: we forbid a 1 followed by all 0s or all 2s. Thus, $\frac{1}{3} = 0.1$ is represented by $0.0\overline{2}$, and $\frac{2}{3}$ becomes 0.2 (*not* $0.1\overline{2}$).

1. See the excellent new biography *Georg Cantor* [PI 87] for Cantor's arduous career and the genesis and eventual acceptance of set theory.

2. The expression *on all scales*, as we said before, reflects one of the central facts of self-similar structures. For example, in percolation at the critical point (see Chapter 15) there are clusters on all size scales; in spin glasses there are magnetic domains on all length scales; Of course, for real physical systems—as opposed to simple mathematical models—the "all" in the expression *on all scales* has to be taken with a grain of salt. For example, in percolation, cluster sizes lie, of course, somewhere between the size of an individual "atom" and the size of the entire sample.

With this convention, the numbers in the open interval $(\frac{1}{3}, \frac{2}{3})$ are precisely all those numbers that in the ternary system have the digit 1 in the first position to the right of the ternary point. Wiping these numbers away on our route to the Cantor dust, we are left with numbers that begin with 0.0 or 0.2.

Similarly, the second wiping (third line in Figure 10 of chapter 1) eliminates all numbers with a 1 in the *second* place to the right of the ternary point. In the end, having arrived at the Cantor dust, we are left with all those proper ternary fractions that have no 1s in *any* place, such as $0.0\overline{2}$, 0.2, 0.2002, and $0.\overline{2002}$.

The members of the Cantor set form a *self-similar* set in the following sense: take any line in Figure 10 of Chapter 1, leave out the right half, and magnify the remainder threefold. This results in the line immediately above it. More precisely, the Cantor set is invariant, modulo 1, to scaling by a factor of 3. In fact, in the ternary notation, this scaling is nothing but shifting all digits one place to the left and dropping any 2s that protrude to the left of the ternary point. For example, the Cantor number 0.202202 maps into 0.02202, another Cantor number.

With the ternary notation, it is easy to see why the Cantor set has measure zero: the probability that a random digit in the interval [0, 1] has not a single 1 in its ternary expansion is, of course, zero. More precisely, the expression for the probability that there is no 1 in n ternary places equals $(\frac{2}{3})^n$, which goes to zero when n goes to infinity.

The ternary number system is also useful to show that no two Cantor numbers could be adjacent to each other. For example, a Cantor number close to the Cantor number 0.2 would be 0.20 . . . 002. By making the sequence of 0s in this "neighbor" longer and longer, we can approach 0.2 arbitrarily closely, but there are always non-Cantor numbers between 0.2 and, say 0.20002—for example, 0.200012.

But how can we prove that the members of this extremely sparse set are so numerous that they are not even countable (Note that the integers and the rational numbers—and even the algebraic irrational numbers—are countable.) The reason is that we can bring the members of the Cantor set, although it is very spotty, into a one-to-one correspondence with *all* the real numbers in the interval between 0 and 1. To accomplish this feat, we simply identify with each Cantor number the *binary* number obtained by changing all 2s to 1s. Thus, for example, 0.020222 corresponds to 0.010111 $(= \frac{23}{64})$. In this manner, each member of the Cantor set can be mapped into a real number and, conversely, *all* real numbers between 0 and 1 can be mapped into Cantor numbers, which have thus the same cardinality as the real numbers.

The well-known fact that the real numbers form an *uncountable* set was proved by Cantor using the "diagonal method" already known to Galileo. The diagonal method is used in an indirect proof that proceeds as follows.

Assume that all the real numbers between 0 and 1 form a countable set; they could then be written down, one after another, in a counting sequence. In

decimal notation, for example, the list might look as follows:

$$\begin{array}{l} 0.91971\ldots \\ 0.29216\ldots \\ 0.36638\ldots \\ 0.55389\ldots \\ \vdots \end{array}$$

To make the notation unique, we use only nonterminating fractions. Thus, 0.5 is written as 0.49999 . . . , for example.

Now write down a number whose first digit to the right of the decimal point is different from the corresponding digit in the first number in the preceding list and whose second digit is different from the second digit in the second number, and so forth. To avoid ambiguities, do not use the digits 0 and 9 in the replacements. The resulting number, say,

$$0.88578\ldots$$

cannot be found anywhere in the list because it differs from each number in at least one place. Thus, the list cannot be complete and our assumption that the real numbers form a countable set was false.

There are several ways to prove that *rational* numbers *are* countable. I find the following proof particularly appealing [Sag 89]. Write the rational number as m/n, where m and n are integers that are relatively prime. Let $m = p_1^{e_1} \cdot p_2^{e_2} \cdots p_k^{e_k}$ and $n = q_1^{f_1} \cdot q_2^{f_2} \cdots q_t^{f_t}$ be the prime-number decompositions of m and n. Then the desired counting function for the rational numbers is $f(1) = 1$ and

$$f\left(\frac{m}{n}\right) = p_1^{2e_1} \cdot p_2^{2e_2} \cdots p_k^{2e_k} \cdot q_1^{2f_1-1} \cdot q_2^{2f_2-1} \cdots q_t^{2f_t-1}$$

This function is uniquely invertible. For example, the rational number $\frac{2}{3}$ is the twelfth number on the list; and the eighteenth number on the list is $\frac{3}{2}$.

Another, completely counterintuitive consequence of Cantor's set theory is the *equivalence* of two-dimensional areas and one-dimensional lines. Two sets are said to be equivalent if there is a one-to-one mapping between them. Thus, the unit square (an area) and the unit interval (a line) are equivalent: each point in the unit square corresponds uniquely to a point on the unit interval, and vice versa. In communicating this discovery to his friend Dedekind in Brunswick (on June 20, 1877), Cantor wrote, "I see it, but I don't believe it."

Actually, Cantor's mapping from the unit square to the unit interval is almost trivial. For example, the point with the rectangular coordinates $x = 0.123$ and $y = 0.456$ is mapped into the point 0.142536 on the unit interval. Can you see what is going on and do you believe it?

Perhaps equally astounding is the fact that *any* number in the interval [0, 2] can be represented by the sum of two Cantor numbers, their extreme sparseness

in the unit interval not withstanding. The reader may find it instructive (and amusing) to prove this counterintuitive property either arithmetically or geometrically.

Cantor Sets as Invariant Sets

One of the arenas in which Cantor sets cavort is that of the so-called *invariant sets* of iterated mappings. (A member of an invariant set maps into a member of the set. An invariant set is the set of all such elements.) Let us consider the simple "tent map" (so called because of its tentlike shape)

$$x_{n+1} = 1.5 - 3|x_n - 0.5|$$

(see Figure 1). Because the slope of this mapping exceeds unity everywhere, there are no attractors, except $x = -\infty$. The two fixed points $x = 0$ and $x = 0.75$ are *repellors*: points near them diverge to infinity.

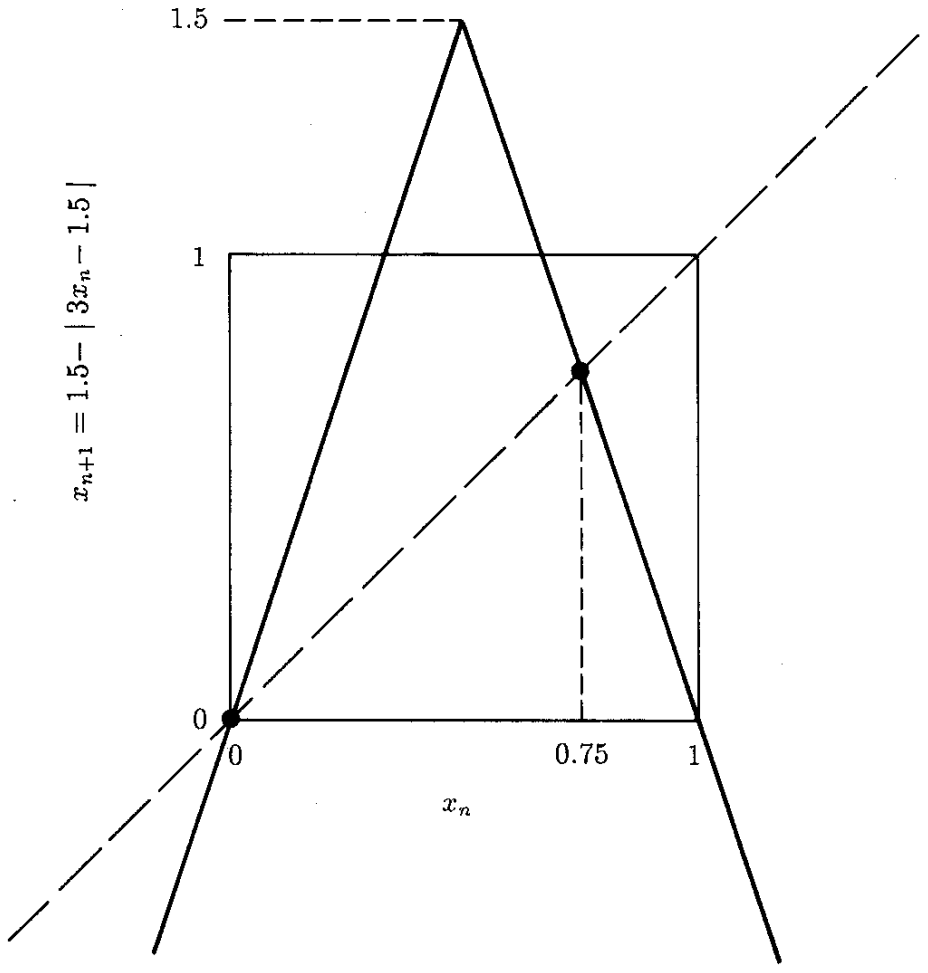


Figure 1 A tent map whose invariant set is the Cantor set of Figure 10 in Chapter 1.

Are there any points that do not diverge, that is, points whose iterates stay forever in the unit interval? To find the answer we write x_n in ternary notation. Then the mapping in Figure 1 says that, for $x_n \leq 0.5$, the next iterate, x_{n+1}, is given by a simple left shift of the digits of x_n. For $x_n > 0.5$, before left-shifting, we must first complement each digit, that is, replace 0 by 2 and 2 by 0, while digit 1 remains 1.

Now, if the initial value x_0 has a 1 anywhere in its expansion, this 1 will *eventually be shifted to the left of the ternary point so that thereafter the iterates* x_n will exceed 1 and diverge to infinity (unless the fraction terminates with a 1).

But suppose there is not a single 1 in the ternary expansion of x_0, that is, that x_0 is a Cantor number; then all the iterates will also be Cantor numbers (remember $0 \rightarrow 0$ or 2, and $2 \rightarrow 2$ or 0). In fact, the iterates will remain forever bounded in the unit interval. To see this, suppose that $x_n = 0.2022\ldots$, which is greater than 0.5. Thus, x_n has first to be complemented (0.0200 . . .) and then left-shifted to yield 0.200 . . . , which is smaller than 1. Of course, if x_n has 0 as a first digit, x_{n+1} is likewise smaller than 1. Thus, Cantor numbers—and *only* these—remain forever bounded, and they remain invariably Cantor numbers, which is why they are called the *invariant set* of the tent map $x_{n+1} = 1.5 - 3|x_n - 0.5|$.

A mapping similar to the one discussed here, but restricted to the unit interval, was described by Bau-Sen Du [Du 84].

Symbolic Dynamics and Deterministic Chaos

There is another important point that we can illustrate with the mapping described in the preceding section. Instead of characterizing a succession of iterates (called an "orbit") by the precise values of the x_n, they are often quantized to just two values. For example, if x_n is smaller than 0.5 that is, if it lies to the *left* of 0.5, we represent the values of x_n by a single symbol: L (for left). If x_n exceeds 0.5, then we write R.

Now, curiously, given any initial value x_0 from the invariant set, we can immediately predict the succession of L's and R's, called the *symbolic dynamics* of its iterates. For example, $x_0 = 0.022020002\ldots$ has the symbolic dynamics $LRLRRRLLR\ldots$, which is obtained by writing L if, in x_0, a digit and its left neighbor are the same and R otherwise. The symbolic dynamics describes the evolution in time of a *dynamic system* such as a playground swing—or the voting pattern of a political population.

Even curiouser, from the (discrete!) symbolic dynamics we can *uniquely* determine the exact value of x_0 (if it belongs to the invariant set). For example, the orbit with the symbolic dynamics $RLRLLLRRL\ldots$ has started at $x_0 = 0.220000200\ldots$. Can the reader see why?

Understandably, invariant sets (and their complements) play a crucial role in dynamic systems in general because they tell the most important fact about any initial condition, namely, its eventual fate: will the iterates be bounded, or will they be unstable and diverge? Or will the orbit be periodic or aperiodic?

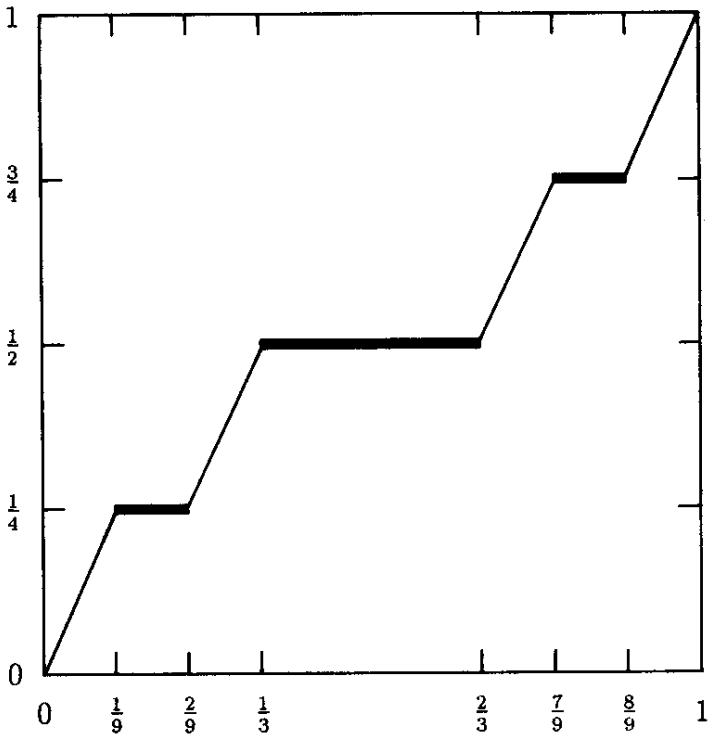
As we can see from these examples, invariant sets can be (and often are) self-similar Cantor sets, that is, uncountable sets with measure 0 and a scaling property. For the Cantor set, the similarity factor is 3, corresponding to a left shift by one digit in the ternary number system.

Representing the evolution in time of a dynamic system by left shifts of the digits in a suitable number system brings out another important property of systems for which such a representation is possible. No matter how accurately the initial condition x_0 of a coordinate is known, the accuracy will always be finite, that is, the digits of x_0 to the right of the last known digit will be unknown. As the dynamic system evolves in time, these unknown digits will be shifted to the left; that is, they will grow in significance, and will sooner or later arrive at the "decimal" point and thus *dominate* the behavior of the system. And because the digits are unknown, this behavior will be completely unpredictable. The resulting motion is called *chaotic*. To emphasize the fact that this chaos is caused by strictly causal, deterministic rules, it is called *deterministic chaos*. Thus we see that there is no contradiction between complete determinism and chaos. In fact, deterministic chaos can be found almost anywhere in nature from turbulence to population dynamics.

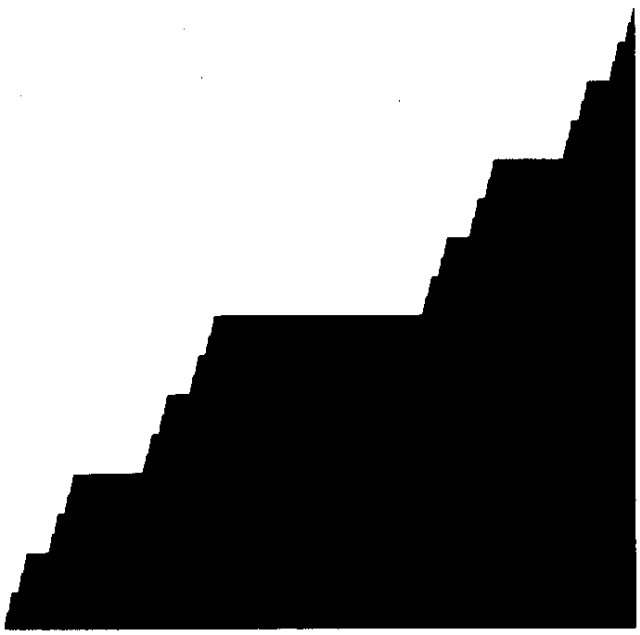
This analysis also tells us why the weather is so unpredictable. The reason is that the mathematical equations governing it are those of a chaotic system. To add a single day to reliable weather forecasting, the initial conditions of temperature, air pressure, wind velocity, and other variables at a large number of points on the earth would have to be known with a much greater accuracy than is presently feasible—not to mention the logistic and computational effort to deal with this mass of data. However, supercomputers now on the horizon, employing large-scale parallel processing, promise more reliable forecasting over a slightly extended time span.

Devil's Staircases and a Pinball Machine

One of the more interesting constructions based on Cantor sets is the devil's staircase. Take the original Cantor set, the middle-third–erasing set, and plot, as a function of x in the unit interval, the relative weight y of the set that lies to the left of x. In the first stage of construction, y rises from 0 to $\frac{1}{2}$ as x goes from 0 to $\frac{1}{3}$. Then y stays constant up to $x = \frac{2}{3}$. Beyond this plateau at $y = \frac{1}{2}$, y rises again from $\frac{1}{2}$ to 1 as x goes from $\frac{2}{3}$ to 1. The second stage of construction has two more plateaus, at $y = \frac{1}{4}$ and $y = \frac{3}{4}$ (see Figure 2A).



(A)



(B)

Figure 2 (A) Second stage and (B) a more advanced stage in construction of a devil's staircase. Plateaus from all earlier stages remain visible.

In the limit, the staircase function $y(x)$ has plateaus almost everywhere, yet it manages to rise from 0 to 1 at uncountably many values x; see Figure 2B, which shows an advanced stage of the resulting devil's staircase.

In order to know how high the staircase is for any given value of x, we have to write x as a ternary number and convert it into a binary fraction, replacing any digits 2 up to the first 1 (reading from left to right) by 1s. Keep the first 1 (if any) and write 0s for all following digits to the right. For example,

$$x = 0.20210012\ldots = \frac{1652}{2187}$$

is mapped into

$$y = 0.1011000\ldots = \frac{11}{16}$$

Thus, for every value of x, there is a unique y value.

To go from a given y value to the corresponding value(s) of x, we have to write y as a binary fraction and convert all 1s to 2s, except the last (if any) 1. After that, we replace each 0 by all combinations of 0, 1, and 2, thereby creating an interval. For example,

$$y = 0.1011 = \frac{11}{16}$$

goes into

$$x = 0.2021111\ldots \begin{matrix} 000 \\ 222 \end{matrix}$$

which represents all the numbers in the open interval $(0.2021\overline{0}, 0.2021\overline{2})$ or, in decimal notation, the interval $(\frac{61}{81}, \frac{62}{81})$. In other words, $y = \frac{11}{16}$ corresponds to one of the plateaus of the devil's staircase with a step width of $\frac{1}{81}$.

In general, any y value whose denominator (in lowest terms) is an nth power of 2 lies on a plateau with width 3^{-n}. All other y values (namely, all nonterminating binary fractions) have unique x coordinates. This is a devilish function indeed: it is constant (doesn't rise) almost everywhere, but it has uncountably many infinitely small discontinuities that allow it to "sneak up" all the way from 0 to 1.

Devil's staircases are excellent models for numerous complex situations in the real world—and the not so real world of mathematical physics. Michel Hénon once invented a kind of pinball machine (see Figure 3), whose symbolic dynamics for a given initial position x_0, $0 < x_0 < 1$, were obtained by computing the devil's function $y(x_0)$ as just described [Hen 88]. If $y(x_0)$ lies on a plateau of the devil's staircase, the ball will escape to plus or minus infinity. The devil's plateaus are, in effect, locked-in intervals for the attractor at infinity. However, if x_0 is a Cantor number, then $y(x_0)$ will not lie on a plateau and the orbit of the ball will forever remain confined to the interval [0, 1]. Its symbolic dynamics are then given by

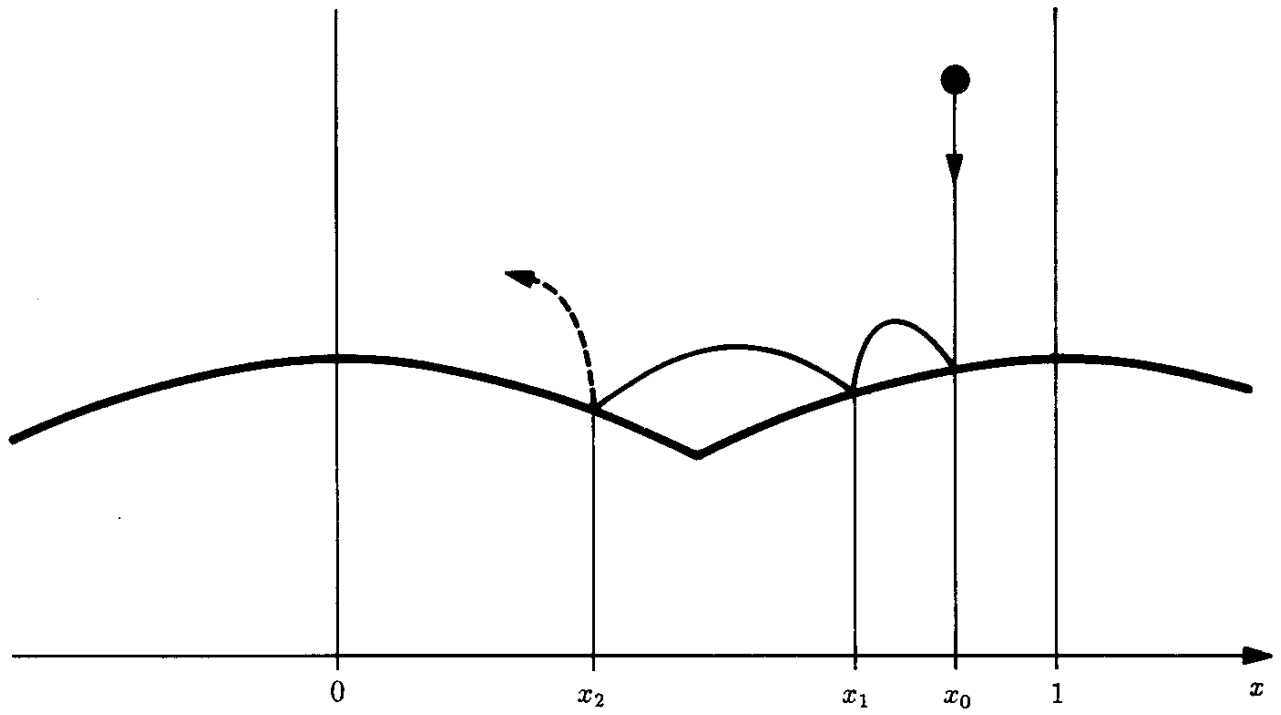


Figure 3 Hénon's pinball machine.

$y(x_0)$ written as a binary fraction with 0 interpreted as L (for left, i.e., $x_n < \frac{1}{2}$) and 1 interpreted as R (for right, i.e., $x_n > \frac{1}{2}$). This is much like the tent map illustrated in Figure 1. (However, Hénon assures me that there are also substantial differences between the tent map and his "inclined billiard," as he called his pinball machine.)

If Hénon's pinball machine were a game played on a computer and the aim were to keep the ball in constant confined motion, then the winning strategy would be to pick a Cantor number for the ball's initial position. This is reminiscent of *Sir Pinski's* game, discussed in Chapter 1 on pages 20–25, for which the Sierpinski gasket (a two-dimensional Cantor set) contains the winning points.

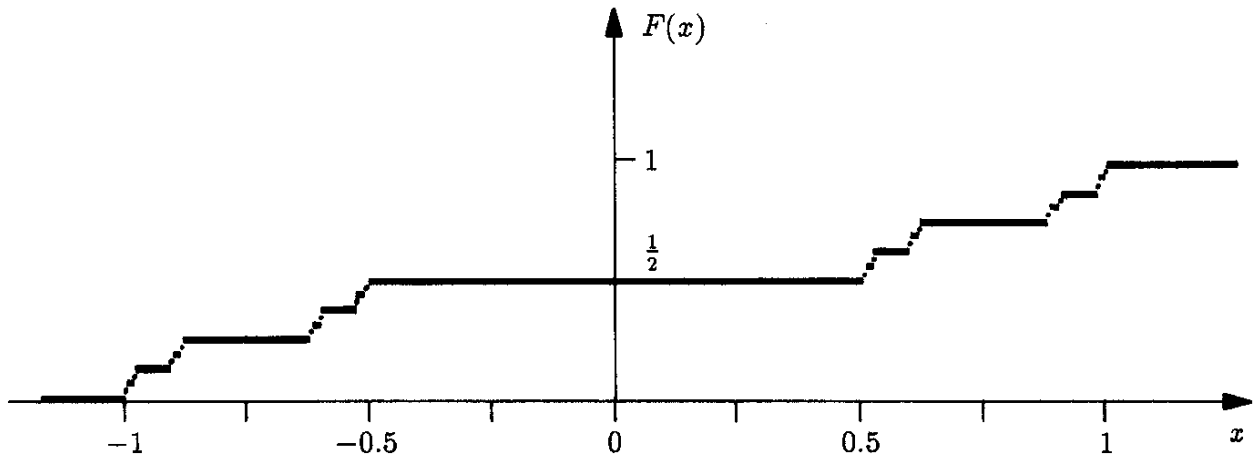


Figure 4 Distribution function for a random variable with a Cantor–set density.

Devil's staircases also appear in probability theory. Figure 4 shows the cumulative distribution of a random variable x given by

$$x = 3\sum_{j=1}^{\infty} \sigma_j \beta^j \qquad \beta = \frac{1}{4}$$

where the σ_j are independent equiprobable $+1$s or -1s. It is easy to see that x can never fall inside the ranges $(-\frac{1}{2}, \frac{1}{2})$, $(-\frac{7}{8}, -\frac{5}{8})$, $(\frac{5}{8}, \frac{7}{8})$, and so forth. This entails the large plateaus in the cumulative distribution seen in Figure 4. Similar staircase distributions that are constant almost everywhere and rise only at a set of x values forming a Cantor set are obtained for any positive $\beta < \frac{1}{2}$.

Mode Locking in Swings and Clocks

One of the more pervasive oscillatory phenomena in nature goes by the names of *mode locking, frequency pulling, phase locking,* or simply *synchronization of two oscillators.* Here, too, devil's staircases play an illuminating role. In mode-locking applications, the height of the devil's staircase corresponds to the frequency ratio of the two oscillators and the plateaus represent locked-in frequency ratios, locked in at any rational number (not just terminating binary fractions, as in the staircase derived from the original Cantor set). Also, the self-similarity may be only asymptotic and not exact.

Think of a playground swing. It has a natural frequency with which it oscillates when driven by a child moving his or her center of gravity up or down at *twice* the natural frequency, as the child soon discovers—long before ever having heard of parametric amplifiers.

But a swing can also be driven externally by a patient parent who pushes and pulls with a frequency not necessarily equal to the swing's natural frequency. Provided that the external force is coupled strongly enough to the swing, the latter will follow the external force; that is, it will be synchronized with the external frequency over a certain range of driving frequencies.

The first scientist to describe such a synchronizing phenomenon was the Dutch mathematical physicist and astronomer Christian Huygens (1629–1695), the discoverer of the Huygens principle of wave propagation. In a letter from Paris to his father in Holland he described how two pendulum clocks hanging back to back on the same wall separating two rooms would synchronize their motions and tick away in perfect lockstep (see Huygens's book *Horologium Oscillatorium* [Huy 1673]). As this example shows, even the tiniest coupling force can "enslave" one oscillator to another if the ratio of their natural frequencies is close to a small-integer rational fraction, such as $\frac{1}{1}$.

Another early observation of synchronization, this time in outer space, came in 1812 when Gauss discovered that the orbit of the asteroid Pallas was locked to the orbital period of Jupiter in the precise integer ratio 7 to 18, two Lucas numbers L_n, which obey the same recursion as the Fibonacci numbers, namely

$L_n = L_{n-1} + L_{n-2}$, but with the initial condition $L_1 = 1$ and $L_2 = 3$. Gauss never published this epochal finding, except for a brief note in the *Gelehrte Anzeigen*, the rapid-communications bulletin of the Göttingen Academy—in encrypted form! But he did communicate his arresting result on May 5, 1812, to his close friend the astronomer Friedrich Wilhelm Bessel (1784–1846), imploring him to keep it completely secret "for the duration." It seems that Bessel functioned exactly as requested, because Gauss never got his due credit. In fact, the "prince of mathematicians" was later not a little miffed by this turn of events, apparently having forgotten his own secretiveness. (Why was Gauss so loath to let the world know about Pallas and Jupiter? Was he afraid to have uncovered some divine interference in the planetary clockwork? No, Gauss knew full well that it was pure and simple nonlinear mechanics. Perhaps he felt that the news would be too upsetting, as in the case of his non-Euclidean geometry, which he kept encased in his desk for decades.)

A similar synchronization can be observed in some radio (and television) receivers with automatic frequency control (AFC): the dial (if the set still has one) can be detuned manually over a certain frequency range, yet the chosen channel will remain locked in. Still another example is the synchronization by an external signal of the horizontal deflection (the "time base") of an oscilloscope or television set. The internal time-base generator will lock in to an external frequency over a certain frequency range and then jump discontinuously to another *rational* frequency ratio, preferably a ratio involving small integers in the numerator and denominator, such as $\frac{1}{2}$ or $\frac{3}{4}$. In fact, the frequency ranges over which the two frequencies are locked into a rational ratio depend, in many applications, on the magnitude of these integers and particularly the denominator. Thus, in a given mode-locking situation, the locked frequency range for the frequency ratio $\frac{1}{1}$ will be larger than that for frequency ratios such as $\frac{1}{2}$, $\frac{1}{3}$, or $\frac{2}{3}$.

Interestingly enough, these locked frequency ranges show a high degree of universality covering innumerable, seemingly unrelated phenomena such as the oscillations of superionic conductors [MM 86] or the heartbeat of periodically stimulated chicken embryos. In fact, these phenomena can be modeled by asymptotically self-similar fractals, which in many cases have identical Hausdorff dimensions: $D = 0.86\ldots$. We will resume the topic of periodic, aperiodic, and chaotic oscillations in Chapter 14.

The Frustrated Manhattan Pedestrian

One of the more "charming" mode-locking situations is regularly suffered by this devoted pedestrian on his occasional forays to the Big Apple (also known as New York City). Walking along one of Manhattan's avenues, he is invariably caught, at every intersection, by a traffic light changing to red as he approaches.

Suppose the pedestrian's speed is just under two-thirds of the "speed" of the traffic lights (i.e., the distance between cross streets divided by the period

of one complete green-yellow-red cycle). The red lights will force him to wait at every intersection and slow him down to a speed of one-half.

Assuming for simplicity that the green cycle, during which the walker can safely traverse the cross street, lasts exactly half a period and that all lights are perfectly synchronized (as they certainly would *not* be on a one-way avenue), then, in the walker's speed range $\frac{1}{2} \leq s < \frac{2}{3}$, he is locked into an effective velocity $v = \frac{1}{2}$. In general, in the speed range

$$\frac{1}{n+1} \leq s < \frac{2}{2n+1} \tag{1}$$

where $n = 1, 2, 3, \ldots$, he is locked into an effective velocity $v = 1/(n+1)$. But the walker can be locked into many more rational speeds (although they may not appear "rational" to him). In fact, for

$$\frac{2(k-1)}{2(k-1)n+1} \leq s < \frac{2k}{2kn+1} \tag{2}$$

$k = 2, 3, 4, \ldots$, the walker's effective velocity is locked into the lower limit of s.

The staircase function corresponding to these locked intervals is illustrated in Figure 5. Although the graph of v versus s is not exactly self-similar, the locking pattern in the interval $\frac{1}{2} \leq s < 1$ is approximately rescaled and repeated

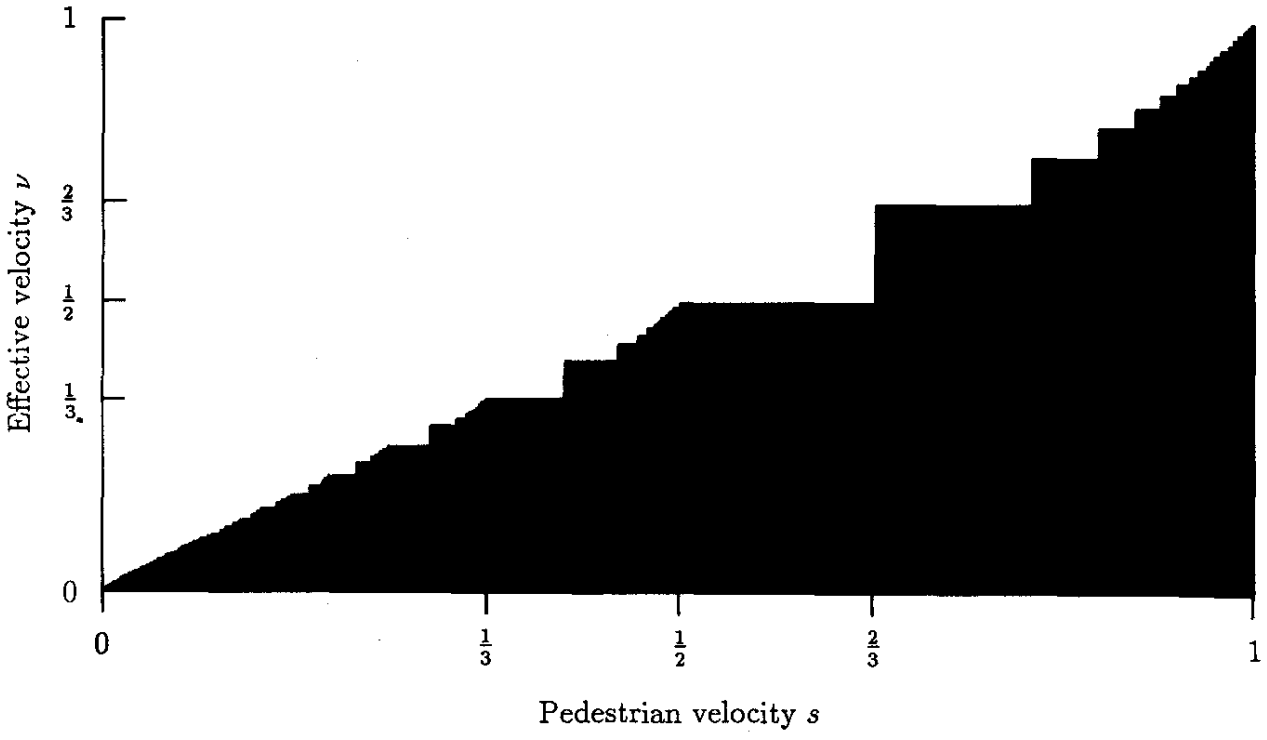


Figure 5 Progress of the frustrated Manhattan pedestrian as an illustration of mode locking.

in the intervals $1/(n + 1) \leq s < 1/n$. One also notices that the locked-in plateaus become smaller and smaller for increasingly larger denominators in inequalities 1 and 2 in the preceding paragraph. In fact, the locked-in speed intervals equal 2 divided by the product of the two denominators.

This scenario of locked intervals, being reciprocally related to the denominators of certain reduced fractions, goes far beyond the hapless Manhattan walker. In fact, we shall encounter staircase functions that, in contrast to Figure 5, have an *uncountable* number of steps.

Arnold Tongues

The plateaus of the devil's staircase encountered previously occur at all heights $y = (2k - 1)/2^n$, with $k, n = 1, 2, 3, \ldots$. However, there are even more satanic staircases that have plateaus at *every* rational number in the interval [0, 1]. While the staircase based on the Cantor set is exactly self-affine (with scaling factors of 3 in the x direction and 2 in the y direction), this is no longer true for satanic staircases, such as the one shown in Figure 6, obtained from the so-called *circle map*:

$$\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi} \sin(2\pi\theta_n)$$

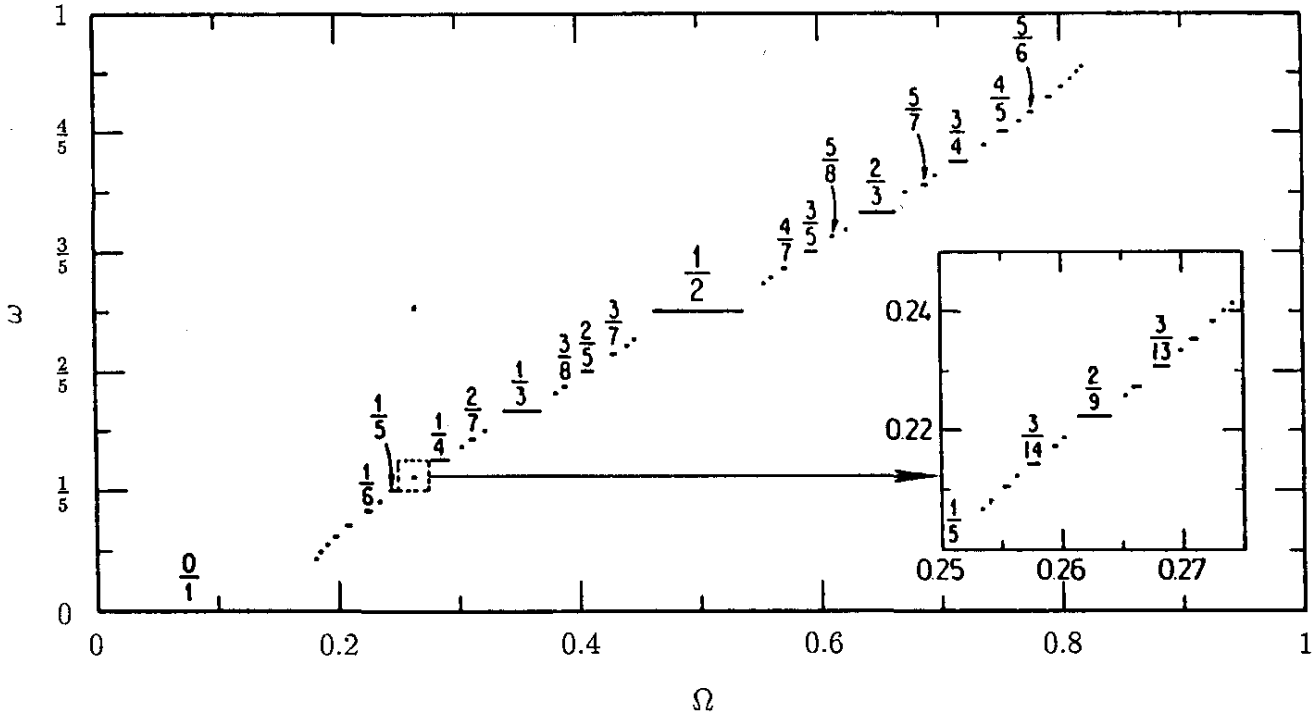


Figure 6 Satanic staircase with plateaus at every rational number.

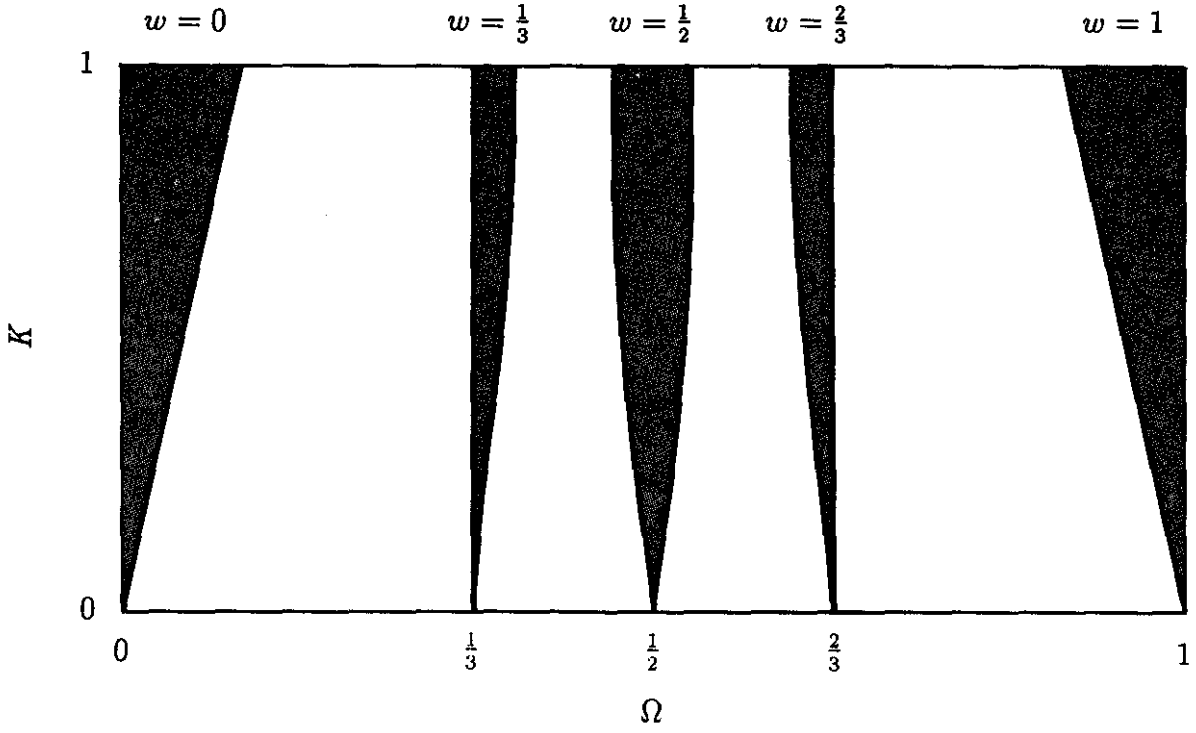


Figure 7 Arnold tongues: locked frequency intervals.

which models many mode-locking phenomena (see Chapter 14). Here K is a coupling strength parameter that controls the degree of nonlinearity and Ω is a frequency ratio, called the *bare winding number*. This frequency ratio may represent the ratio of a driving-force frequency and the resonance frequency of an oscillator (think of the swing, including the one "executed" on the dance floor; or contemplate the frequency ratios of planetary or lunar orbits and spins).

Without coupling ($K = 0$), the so-called *dressed* winding number w, defined as the limit as $n \to \infty$ of $(\theta_n - \theta_0)/n$, equals the bare winding number Ω. But for $K > 0$, w "locks" into rational (frequency) ratios, preferably ratios with small denominators.

Figure 7 shows some of the frequency-locked regions in the Ω-K plane. The shaded regions are called *Arnold tongues*, after their discoverer, the Russian mathematician V. I. Arnol'd. (There never seems to be a lack of suggestive terms in fractal heaven or hell.)

In other applications, the dressed winding number w may represent, for example, the relative number of up spins in an Ising model of an antiferromagnetic material or the relative abundance of a given element (or molecular structure) in a crystal or quasicrystal (see Chapter 13).

For the critical value $K = 1$ of the coupling parameter, the infinitely many locked frequency intervals corresponding to *all* the rational dressed winding numbers w between 0 and 1 actually cover the entire Ω range of bare winding numbers. Irrational values w correspond to an uncountably infinite set of zero measure of Ω values—in other words, a Cantor dust.

8

CHAPTER

Fractals in Higher Dimensions and a Digital Sundial

Jetzt Mengenlehre (Now set theory)

—KURT GÖDEL, in 1937,
when he decided to turn his mental scalpel loose
on set theory, after completing his proof of
undecidability in number theory

Self-similar or self-affine sets in higher dimensions are models of strange attractors and their basins of attraction; of porous materials, dendritic crystal growth, and quasicrystals; of mountainscapes, Brownian motion, and related stochastic processes that describe an assortment of catastrophes (plus a few happier happenings). Some of these "practical fractals" will be visited in Chapter 10. Here we examine some of their foundations and design a *digital* sundial based on a Cantor set.

Cartesian Products of Cantor Sets

The original one-dimensional Cantor set can be generalized to dusty sets in two or more dimensions in several different ways. Consider the set of all points in the unit square for which both the abscissa x and the ordinate y belong to the Cantor set C. The resulting *Cartesian product* of the Cantor set with itself, usually written $C \times C$, is a *Cantor* dust embedded in two dimensions (see Figure 1).

What is the Hausdorff dimension D of this dust? The set $C \times C$ can obviously be covered by $N(r) = 4^n$ squares of side length $r = 1/3^n$. Thus

$$D(C \times C) = \lim_{n \to \infty} \frac{\log (4^n)}{\log (3^n)} = \frac{\log 4}{\log 3} \approx 1.26 \ldots$$

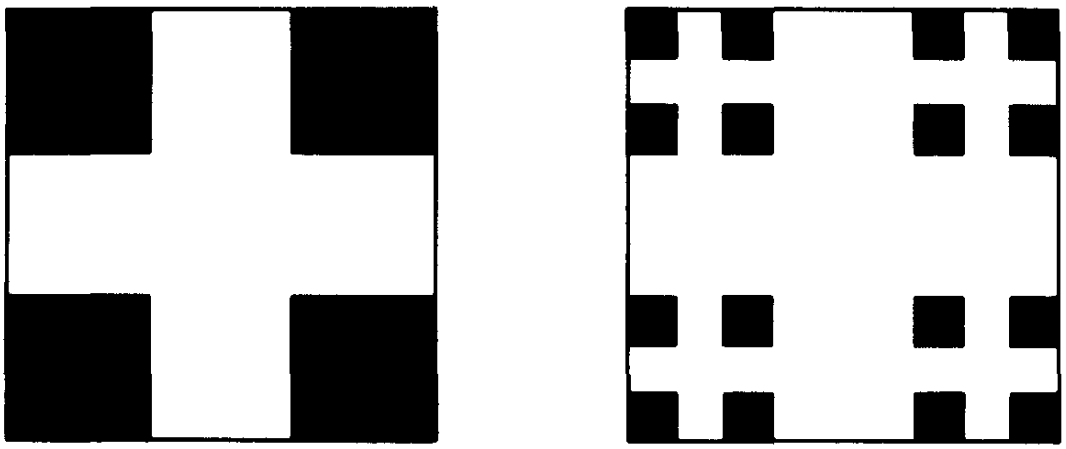


Figure 1 Toward a two-dimensional Cantor dust: the Cartesian product of the middle-third Cantor set with itself. (A) The first step and (B) the second step of construction. Fractal dimension = 1.26. . . .

or twice the value or the original one-dimensional Cantor set. In fact, it is easy to guess (and not difficult to prove) that for the three-dimensional Cartesian product $C \times C \times C$, the Hausdorff dimension equals

$$D(C \times C \times C) = \frac{\log 8}{\log 3} = 3\,\frac{\log 2}{\log 3} \approx 1.89\ldots$$

and so forth: forming a k-fold Cartesian product multiplies the Hausdorff dimension by the factor k, just as for ordinary Euclidean dimensions. (Note that $C \times C \times C$, a dust floating around in three-dimensional space, is so thin that its Hausdorff dimension doesn't even reach the value 2.)

A Leaky Gasket, Soft Sponges, and Swiss Cheeses

Consider another Cartesian product, that of the *complement* C' of the Cantor set C with itself: $C' \times C'$. The complement of $C' \times C'$, that is, $(C' \times C')'$, can be constructed recursively in the following way. The initiator is the unit square, and the generator is the unit square with the central square of side length $\frac{1}{3}$ deleted. In the next iteration, the central squares of side length $\frac{1}{9}$ are removed from the eight remaining squares of side length $\frac{1}{3}$ (see Figure 2A–C). Infinite iteration produces the *Cantor gasket*, approximated by the dark area in Figure 2D.

What is *its* Hausdorff dimension? Figure 2 suggests that $(C' \times C')'$ has a *larger* dimension than $C \times C$. In fact, since the Cantor gasket is strictly self-similar, we need to consider only the generator. To cover it, we need eight

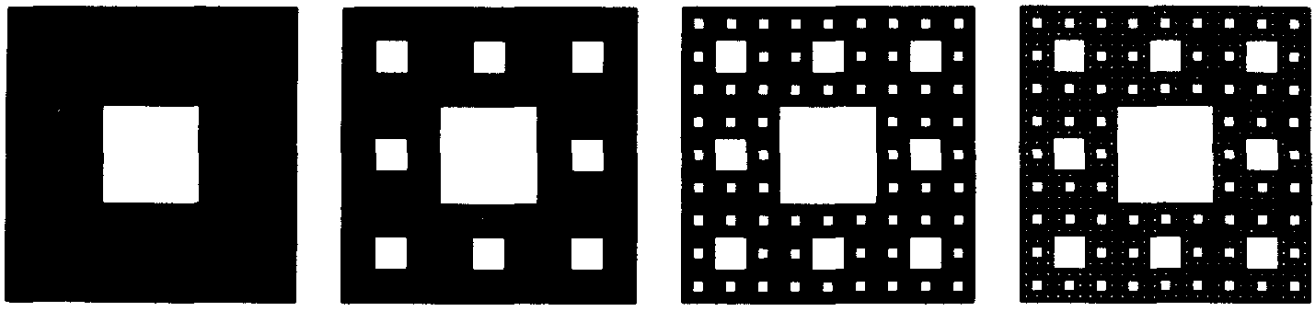


Figure 2 The first three steps in the construction of the Cantor gasket. Fractal dimension = 1.89 . . . .

squares of side length $\frac{1}{3}$. Thus,

$$D((C' \times C')') = \frac{\log 8}{\log 3} \approx 1.89$$

which is the same value we found for the three-dimensional dust $C \times C \times C$.

What about $(C' \times C' \times C')'$, nicknamed *Cantor cheese*? The generator can be covered by $27 - 1 = 26$ cubes of side length $\frac{1}{3}$. Thus,

$$D = \frac{\log 26}{\log 3} \approx 2.97$$

a value close to 3 because the Cantor cheese is quite solid and has only isolated holes.

The generalization of the Cantor cheese to k Euclidean dimensions, the set $(C' \times C' \times \cdots \times C')'$, has Hausdorff dimension $D = \log (3^k - 1)/\log 3 \approx k - 1/(3^k \log_e 3)$, a value just below the embedding dimension k.

There exists still another *symmetric* fractal set in three dimensions based on C (or any other fractal set in one dimension). It is called the *Menger sponge*, after its architect, Karl Menger, and is depicted in Figure 3 [Men 79]. It has no two-dimensional analogue. Its "holes" are open channels that penetrate the unit cube. Applying the inclusion-exclusion principle, one sees that the generator leaves $27 - 9 + 3 - 1 = 20$ cubes of side length $\frac{1}{3}$, giving a Hausdorff dimension $D = \log 20/\log 3 \approx 2.73$. Thus, the Menger sponge is intermediate between Cantor dust and cheese, but closer to the latter, as one might expect. Mandelbrot suggested the Menger sponge independently as a model of turbulent intermittency [Man 74].

A good set-theoretic description, which brings out the symmetry of the Menger sponge in the three Cartesian coordinates x, y, and z, can be constructed as follows. Call the set C in the x direction X, the set in the y direction Y, and

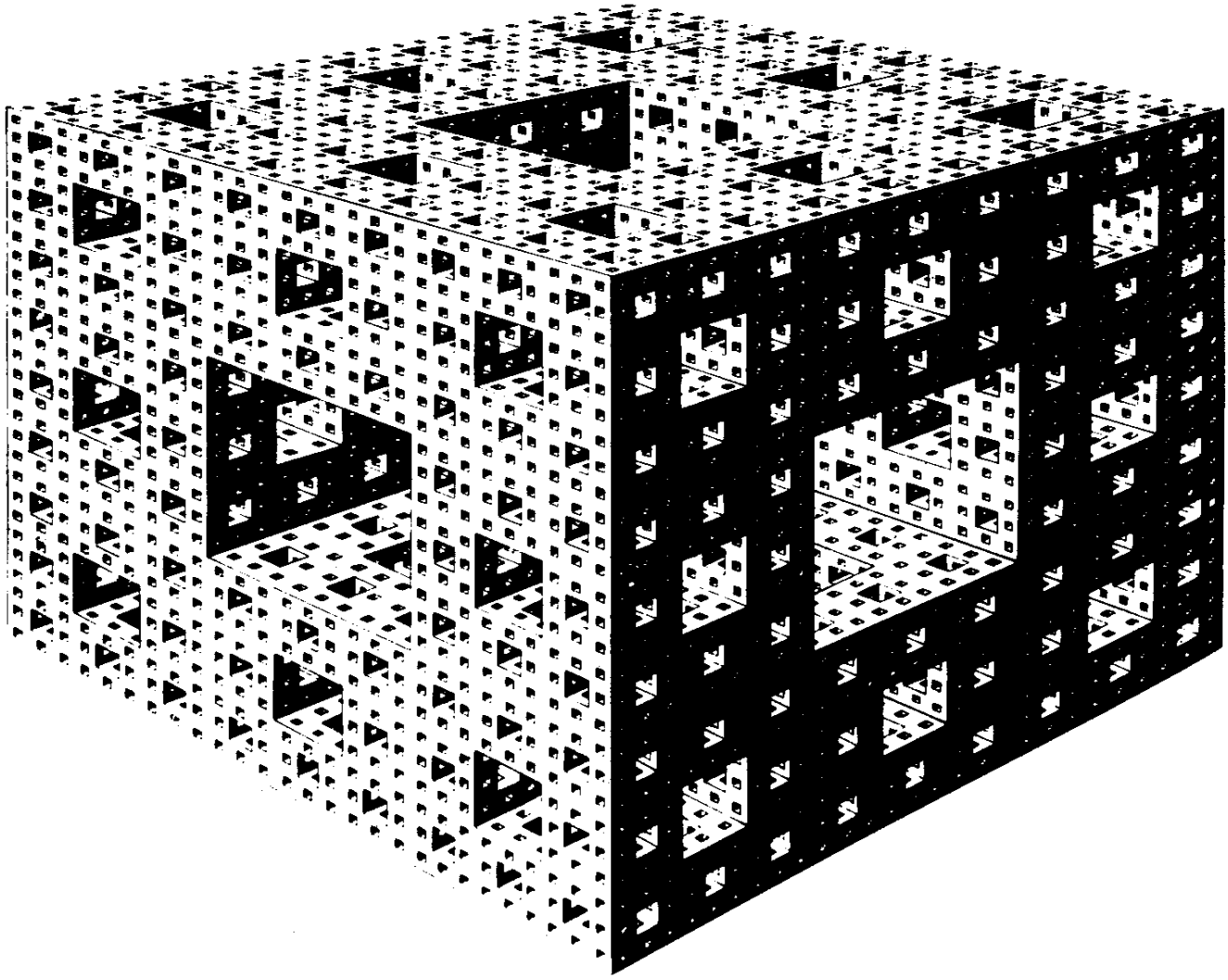


Figure 3 Menger sponge. Fractal dimension = 2.73 . . . .

that in the z direction Z. Then, just considering generators, $X' \cap Y'$ is the central square "hole" in the x-y plane (see Figure 2) and

$$(X' \cap Y') \cup (Y' \cap Z') \cup (Z' \cap X')$$

represents all the holes of the generator. The complement of this set is the generator of the Menger sponge. By De Morgan's rule [Hal 74], it is

$$(X \cup Y) \cap (Y \cup Z) \cap (Z \cup X) \tag{1}$$

Is the set whose *generator* is the complement of the sponge generator also fractal, with a Hausdorff dimension D between 0 and 3? (The complement of the sponge itself has, of course, a finite Lebesgue measure and $D = 3$.) The complementary generator has $27 - 20 = 7$ cubes of side length $\frac{1}{3}$: a cube bordered by 6 cubes. Hence, $D = \log 7/\log 3 = 1.77 \ldots$. It is a rather spidery set.

The original Cantor set generalized to higher embedding dimensions and similar sets produce an increasingly varied zoo of dusty sets.

A Cantor-Set Sundial

By applying the Cantor construction in two dimensions, we obtain, as we have seen, a Cantor "dust" dispersed in the plane. It can be generated from the unit square by eliminating center thirds in both the x and y directions and repeating the process *ad infinitum* (see Figure 1). Mathematicians call the resulting set also the *direct product* of the Cantor set C with itself and denote it by $C \times C$. As we have seen, its Hausdorff dimension D follows directly from the generator, which consists of 4 remaining squares out of 9. Since $C \times C$ can be covered by 4^n (but no fewer) squares of side length 3^{-n}, we obtain

$$D = \frac{\log 4}{\log 3} = \frac{2 \log 2}{\log 3} = 1.26 \ldots$$

which equals twice the value for C itself. In general, we will find that the Hausdorff dimensions of Cartesian product sets are the *sums* of the dimensions of the individual sets. Thus, for example, the Cantor dust $C \times C \times C$ floating around in three-space has $D = 3 \log 2/\log 3 = 1.89 \ldots$.

Arithmetically, such n-dimensional Cantor sets are described by n-tuples of Cantor numbers $(x_1, x_2, \ldots, x_n)$, where each x_k is a Cantor number, that is, a ternary fraction using only 0s and 2s an no 1s, as described in pages 162–163 in Chapter 7.

It is interesting to note that fractal sets embedded in higher-dimensional Euclidean spaces, when projected into spaces with fewer Euclidean dimensions, generate fractal sets whose Hausdorff dimensions depend on the direction of the projection. Consider, for example, the Cantor-like one-dimensional set C_4, defined by eliminating central *quarters* of the unit interval. The dust in three-space constructed from the triple Cartesian product of this set has Hausdorff dimension $D = 3 \log 2/\log \frac{8}{3} = 2.12 \ldots$, which exceeds 2. This set, $C_4 \times C_4 \times C_4$, when projected along one of its three coordinate axes, generates the set $C_4 \times C_4$ with Hausdorff dimension $2 \log 2/\log \frac{8}{3} = 1.41 \ldots$, which is smaller than 2. But other projection directions can produce sets with dimensions equal to 2 having *connected* pieces. In other words, the Cantor-like dust $C_4 \times C_4 \times C_4$ will cast "weightless" shadows in some directions and shadows with visible patterns in other directions.

On the basis of this observations, K. J. Falconer [Fal 87] has proposed one of the most paradoxical sets ever conceived: a *digital* sundial (see Figure 4). Depending on the position of the sun in the sky, the set will cast a shadow that changes every minute to correspond to the local time. If desired, the set can

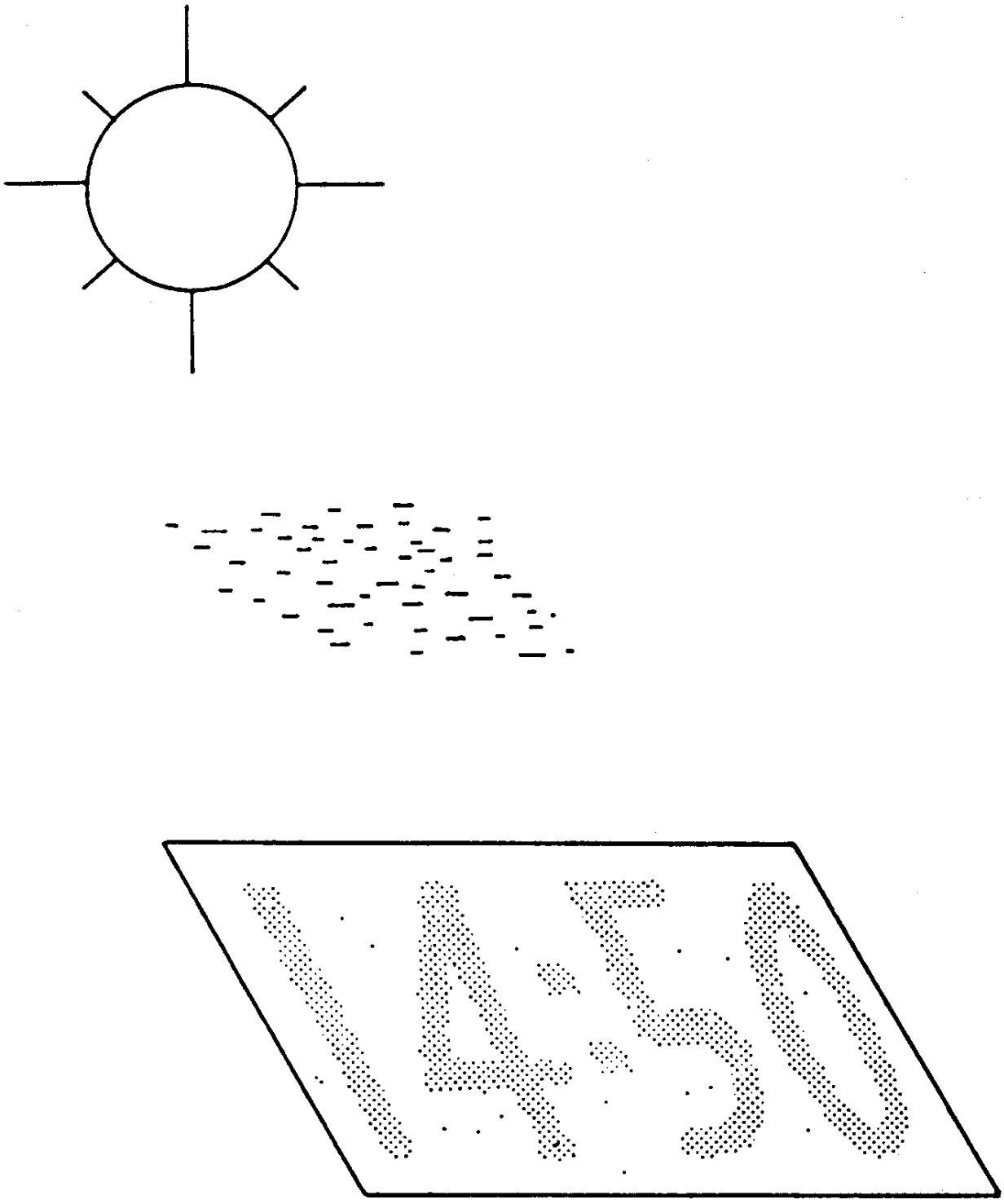


Figure 4 A digital sundial based on a Cantor set [Fal 87].

even be "enlarged" to show the correct date between winter and summer solstices. Here we have the ultimate timepiece driven by sun power.

Of course, the shadow-casting set is likely to be rather complicated, and the inventor understandably refrains from detailed instructions for its construction (presumably while patents are pending and diffraction limits are being circumvented). However, an inkling of how to set out constructing sets with projections of varying sizes is illustrated in Figure 5.

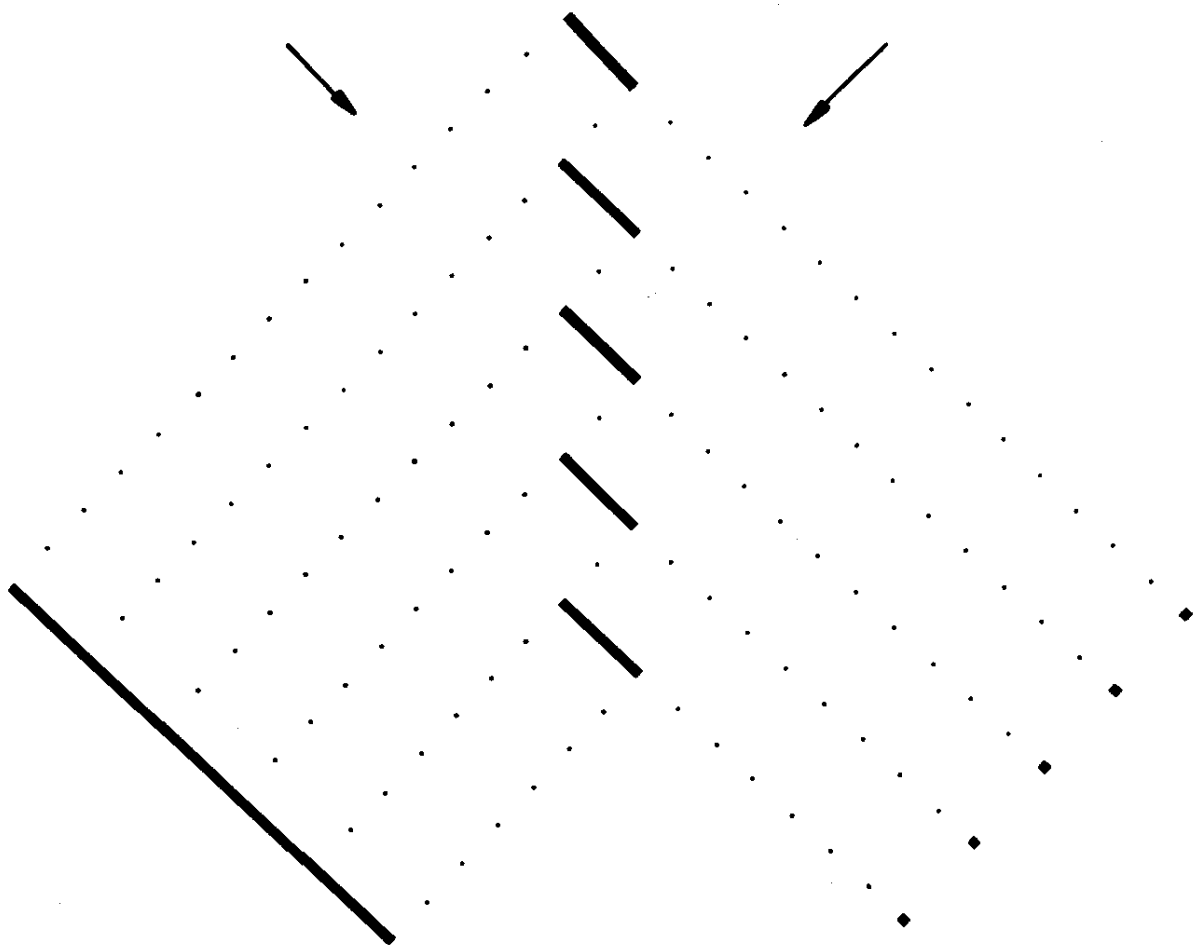


Figure 5 Idea underlying digital sundial: a set of bars that casts very different shadows, depending on the direction of projection.

Fat Fractals

The sundial fractal discussed in the previous section is an example of a fractal set that for certain projection direction has nonzero measure. There are numerous serious applications that are well characterized by fractal sets of nonzero measure. This is particularly true for nonlinear dynamic systems and their basins of attraction. For example, the parameter values for which the prototype of such systems, the logistic parabola (see Chapter 12), shows aperiodic behavior is such a set [Jak 81].

Corresponding to each periodic orbit is a finite interval of parameter values, called a periodic window. The union of periodic windows does not exhaust all parameter values. Thus, the parameter values for aperiodic orbits have a nonzero Lebesgue measure. On the other hand, the distribution of such parameter values has a fractal structure: it has holes (periodic windows) on all scales. Such fractal sets, with nonzero measure, have been called *fat fractals* [EU 86]. Another example of a fat fractal is the set of parameter values for which the subcritical circle maps show aperiodic behavior, that is, parameter values that do not lead to mode locking.

Clearly, such sets cannot be usefully characterized by their Hausdorff dimension, which would simply be equal to the embedding Euclidean dimension and would therefore not provide any additional information. Rather, fat fractals are distinguished by *scaling exponents.*

A simple example of a fat fractal is obtained by starting with the unit interval and removing the central $\frac{1}{3}$ in the first generation, the central $\frac{1}{9}$ of the two remaining thirds in the second generation, the central $\frac{1}{81}$ of the four resulting pieces, and so on, always cutting out central pieces of *relative* length $1/3^{2^k}$ (see Figure 6). After n iterations, we obtain 2^n pieces with a total length

$$\mu_n = \prod_{k=0}^{n-1} (1 - 3^{-2^k}) \tag{2}$$

which for $n \to \infty$ converges on a nonzero value, $\mu_\infty = 0.5851874\ldots$.

A somewhat leaner fat fractal is obtained by excising central pieces of relative length 3^{-k} at each iteration, resulting in a remaining length

$$\mu_\infty = \prod_{k=1}^{\infty} (1 - 3^{-k}) = 0.560\ldots \tag{3}$$

Fat fractals are distinguished by one or another of several scaling exponents. The most useful scaling exponent is defined as follows: Fill all holes of length up to ε and approximate the measure $\mu(\varepsilon)$ of the resulting set for $\varepsilon \to 0$ by the power law

$$\mu(\varepsilon) = \mu(0) + c\varepsilon^{\beta} \tag{4}$$

where c is a constant and β is the scaling exponent; β lies in the range $0 \leq \beta \leq \infty$.

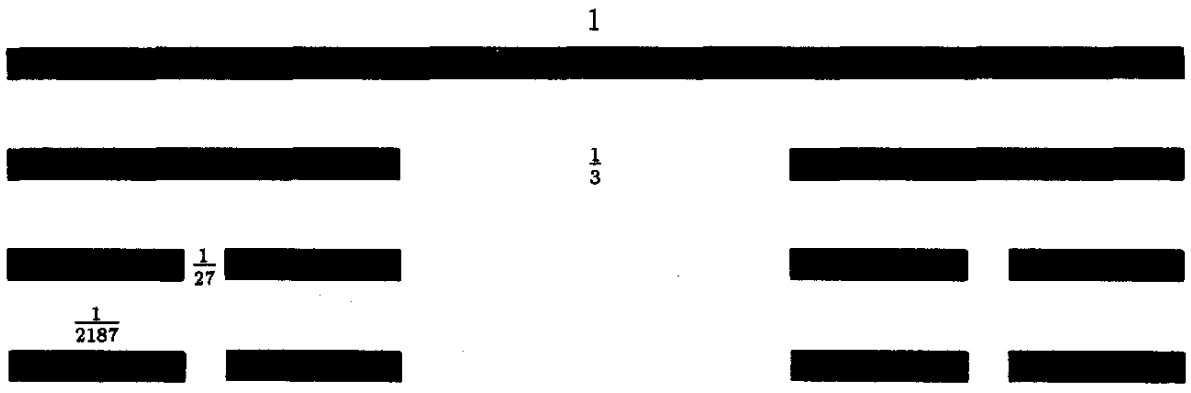


Figure 6 Construction of a fat fractal. It has inifinitely many holes, but the remainders keep a total length that is greater than zero.

In the present example, with the excising rule 3^{-k}, setting $\varepsilon = 3^{-n}$ gives

$$\mu(\varepsilon) = \prod_{k=1}^{n-1} (1 - 3^{-k}) = \frac{\mu(0)}{\prod_{k=n}^{\infty} (1 - 3^{-k})}$$

or

$$\mu(\varepsilon) = \mu(0)(1 + 3^{-n} + 3^{-n-1} \cdots)$$

and, asymptotically for $n \to \infty$,

$$\mu(\varepsilon) - \mu(0) = \mu(0)(3^{-n}) = \mu(0)\varepsilon$$

Thus, for the fat fractal defined by equation 3, the scaling exponent β equals 1.

Denoting the measure of the holes smaller than ε by $F(\varepsilon)$, the scaling exponent β is also given by

$$\beta = \lim_{\varepsilon \to 0} \frac{\log F(\varepsilon)}{\log \varepsilon} \tag{5}$$

The exponent β is determined by the rate with which the measure of the small holes vanishes.

For the quadratic map, J. Doyne Farmer determined $\mu(0)$ numerically as 0.89795 ± 0.00005. Both for the quadratic map and for a trigonometric map with a quadratic maximum, $x_{n+1} = \gamma \sin(\pi x_n)$. Farmer found $\beta = 0.45 \pm 0.04$. This is a hint that the exponent β is universal, that is, the same for all maps with a quadratic maximum [Far 85].

Another scaling exponent, α, is obtained by fattening all holes by ε. This not only fills in the small holes but also reduces the size of the large holes. Let $G(\varepsilon)$ be the additional contribution to the measure. The exponent α is then defined by

$$\alpha = \lim_{\varepsilon \to 0} \frac{\log |F(\varepsilon) + G(\varepsilon)|}{\log \varepsilon} \tag{6}$$

It can be shown that $\alpha \leq \beta$. If $\alpha < \beta$, then α is determined by the large holes. Since, in most applications, the fine-grain fractal structure is more important than the coarse-grain structure, β is the more useful exponent. However, the exponent α can still contain useful information in the case of fat fractals, describing, for example, parameter values for which a nonlinear system shows chaotic motion. For $\alpha = \beta$, such systems exhibit sensitive dependence on the *parameter*; that is, arbitrarily close to some parameter value resulting in chaotic motion, there are other values for which the motion is periodic, while for $\alpha < \beta$ this is not the case. Thus, the equality $\alpha = \beta$ signals an important property of nonlinear dynamic systems called *parameter sensitivity*.

9

CHAPTER

Multifractals: Intimately *Intertwined Fractals*

Cantorism is a disease from which mathematics would have to recover.
—HENRI POINCARÉ

Fractals have immeasurably enlarged our ability to describe nature. The abstract constructions going back to Bernard Bolzano (1781–1848), Cantor, and Giuseppe Peano (1858–1932) have furnished us with models of reality much more realistic than the Euclidean empire of integer exponents and smooth shapes. Yet there are many phenomena in physics, chemistry, geology, and crystal growth, in particular, that require a generalization of the fractal concept to include intricate structures with more than one scaling exponent. Many of these matters are in fact characterized by an entire *spectrum* of exponents, of which the Hausdorff dimension is only one. The generalized fractals fashioned to cope with these cases are called *multifractals*. Their applications range from the distributions of people or minerals on the earth to energy dissipation in turbulence or fractal resistor networks. Diffusion-limited aggregation, viscous fingering, and the distribution of faults in computer networks, or that of impurities in semiconductors, are likewise well modeled by multifractals, as are certain games of chance. And many strange attractors of nonlinear dynamic systems are also clearly multifractals. In fact, it was mainly with *multi*fractals that the "fractal geometry of nature" overtook pure geometry to conquer the natural sciences.

The Distributions of People and Ore

Gold is everywhere on earth. Not only can it be found in a few rich veins; there are thousands of lesser deposits where gold can be profitably produced. And

there are millions of sites on earth where gold is known to exist but not worth mining. In fact, gold is all around us and even *in* us. The total amount of gold in the oceans is estimated at billions of tons, but its concentration is less than 6 parts in 1 trillion parts of seawater. Thus, there seem to be no total voids in the distribution of gold on earth.

And what is true of gold is manifest for many minerals too. As the Dutch geologist H. J. de Wijs once wisely observed, a mineral is typically not concentrated exclusively in ore veins, but can be found between the veins, too, albeit in lesser concentrations. And the veins themselves show characteristic variations of the concentration [deW 51]. In fact, every time a volume of ore is bisected along the vein, the relative amounts of the mineral in the two half volumes are p and $1 - p$, respectively. Interestingly, the maximum value of the parameter p, which measures the variability of the mineral's concentration, stays roughly constant from bisection to bisection. But does this law live for ore alone?

How are *people* distributed over a large connected landmass, say, Eurasia or the Americas? If we cut the total land area into two equal-area pieces, we may find that perhaps 70 percent of the people live on one side of the dividing line, while only 30 percent live on the other side. In general, the proportion of people on one side may be p, with $p > 0.5$, whereas it is only $1 - p$ on the other side. In the present example, $p = 0.7$. The excess proportion p depends, of course on the direction of the cut. Let us therefore assume that the cut direction is chosen to maximize p.

If we now proceed to cut the denser half area into quarter areas, again positioning the cut to maximize the proportion in one of the quarters, we may find that the population percentages are approximately p^2 and $p(1 - p)$, respectively. Similarly, cutting the sparser half into two equal areas will result in quarter areas with densities near $(1 - p)p$ and $(1 - p)^2$.

Iterating this bisecting process results in an asymptotically self-affine distribution (see Figure 1). We begin with a uniform probability distribution over the unit interval (Figure 1A). After bisecting the interval once, we find two probabilities, $1 - p$ and p, for the two halves of the unit interval. With $1 - p < p$, we obtain the single-step distribution shown in Figure 1B for $p = \frac{2}{3}$. Bisecting once more results in four intervals and the probability distribution shown in Figure 1C. A third bisecting results in distributions shown in Figure 1D. In the limit of infinitely many bisectings, we obtain a self-affine probability distribution: the left half of the distribution stretched by a factor of 2 in the horizontal direction and by a factor of $1/(1 - p)$ in the vertical direction reproduces the entire distribution [PS 86].

If we consider $p = 0.7$ as characteristic for the distribution of people on the earth as a whole, then 18 bisections of the earth's land area will leave just two hermits living on an area of $576\ \text{km}^2 = 24\ \text{km} \times 24\ \text{km}$ in the sparsest region, say, in central Siberia, while 8 million people will share the same area in a dense megalopolis. According to this simple bisecting model, most people (3.5 billion) live in 60,000 communities of 20,000 to 300,000 people each.

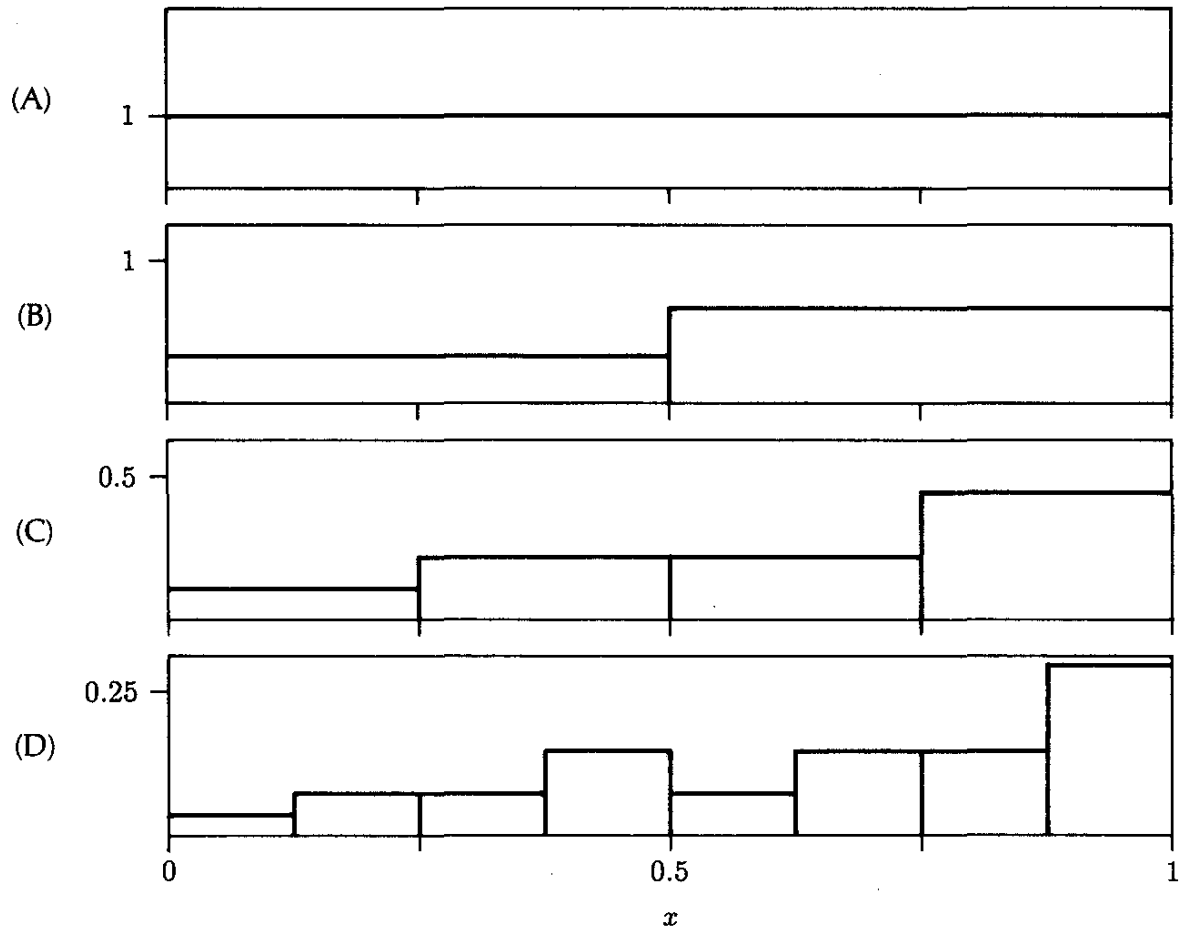


Figure 1 Constructing a self-similar distribution by iterated bisecting [PS 86].

And what is true of people and ore obtains for photons too. Take a beam of light from an old-fashioned incandescent lamp and cut it into two equal portions. The number of photons in the two halves will not be the same, nor will the numbers be equal in the four quarter beams, and so on, for each bisection. Or take electromagnetic cavity radiation, produced in a hollow space whose walls are heated to a given temperature. (A tiny hole in the cavity's wall will emit the justly famous blackbody radiation, as described by Planck's law.) The number of photons in one of the cavity's phase cells or modes of oscillation is distributed according to a geometric distribution: add one photon, and the probability decreases by a constant factor $m/(m + 1)$, where m is the expected number of photons (given by the "Bose-Einstein" distribution). The variance σ^2 of this distribution equals $m^2 + m$, where the first term (m^2) stems from the heat-induced random fluctuations of the classical electromagnetic field. The second term (m) reflects the "granularity" of the energy due to Einstein's photons, the particles of light whose existence he deduced from the added m in $\sigma^2 = m^2 + m$. (This granularity was originally introduced by Planck to match theory to experiment.)

For large m, $\sigma^2 \approx m^2$; and bisecting the cavity's volume will result in an expected proportion p of photons in one half and $1 - p$ in the other half, with $p \approx 0.6$ for a cutting direction that maximizes p—independent of the *number* of

photons. For large m the geometric distribution is in fact scale-invariant or self-similar, because $\sigma \sim m$. Thus, repeated bisections will continue to apportion the available photons in a self-similar branching ratio $p/(1-p) \approx 1.5$ until their number becomes so small that σ/m is no longer constant and the existence of individual photons destroys strict self-similarity.

In laser light, by contrast, photons obey a Poisson distribution with $\sigma^2 = m$. Thus, the scale of the distribution, $\sigma \sim \sqrt{m}$, is not proportional to its mean m, and self-similarity does not obtain.

Self-Affine Fractals without Holes

The repeated bisecting and multiplying of proportions in each half by $(1-p)$ for each left half and by p for the right half intervals, which we exercised in the preceding section, is a special case of a multiplicative random process. Infinite iteration of such a process results in a *self-affine* distribution of densities (see Figure 1, which shows the initial uniform density over the unit interval and the results of the first three iterations).

After n iterations, the probability in the interval $m \cdot 2^{-n} < x < (m+1)(2^{-n})$ is given by $p^k(1-p)^{n-k}$, where k is the number of 1s in the first n binary places of x. For example, for $n = 6$ and $x = \frac{1}{5} = 0.00110011\ldots$, $k = 2$. Hence the proportion equals $p^2(1-p)^4$, as it does for the entire interval $0.001100 = \frac{12}{64} < x < 0.001101 = \frac{13}{64}$ of length 2^{-6} that contains $x = \frac{1}{5}$. In the sixth iteration, there are $\binom{n}{k} = \binom{6}{2} = 15$ intervals with this density, namely, those 15 intervals whose x values have precisely two 1s in their first six binary places. The most frequent density for $n = 6$ corresponds to $k = 3$. It occurs $\binom{6}{3} = 20$ times, the leftmost interval beginning at $x = 0.000111 = \frac{7}{64}$. The reason the binary notation for x describes this kind of distribution is that each 0 in the binary expansion of x corresponds to a left half interval and each 1 to a right half interval.

Note the incipient self-affinity in this recursive construction: the right half of each distribution (see Figure 1) equals the left half times $p/(1-p)$, and the entire distribution tends to be invariant as the left half is stretched by a factor of 2 in the horizontal direction and a factor of $1/(1-p)$ in the vertical direction

In the limit as $n \to \infty$, the distribution $P(x)$ over the entire unit interval equals the one over the left half interval stretched horizontally by a factor of 2 and vertically by the factor $1/(1-p)$:

$$P(x) = \frac{1}{1-p} P\left(\frac{x}{2}\right) \tag{1}$$

This is a functional equation that we shall encounter again in an intriguing

gambling strategy (see pages 207–210). The factors 2 and $1/p$ are the two scaling factors of this self-affine fractal.

How else can we characterize this fractal function? The usual Hausdorff dimension D, based on the limit as $r \to 0$ of $\log N/\log(1/r)$, is of little help here. After n iterations the number of pieces N equals 2^n and the length r of each piece equals 2^{-n}. Thus $D = 1$, reflecting the fact that the fractal shown in Figure 1 has no holes.

On the other hand, if we focus on the percentages and their distribution, we find that after $n = 2m$ iterations a large number of the segments, namely, $\binom{n}{m}$, have a probability of $(1-p)^m p^m$. Their locations within the unit interval are precisely all those half-open intervals of length 2^{-n} whose abscissa values x have an equal number of 0s and 1s in the first n binary places of x. For example, for $n = 4$, these $\binom{4}{2} = 6$ special intervals are given by $x = 0.0011\ldots \triangleq (\frac{3}{16}, \frac{4}{16})$, $x = 0.0101\ldots$, $x = 0.0110\ldots$, $x = 0.1001\ldots$, and $x = 0.1010\ldots$, and $x = 0.1100\ldots$. (Here the triple dots indicate all possible combinations of 0s and 1s, thereby defining an *interval* and not just a single value.)

In general, there are $\binom{n}{k}$ segments of length 2^{-n} with density $(1-p)^k p^{n-k}$, representing a total probability

$$p_{k/n} = \binom{n}{k}(1-p)^k p^{n-k}(2^{-n})$$

Note that

$$\sum_{k=0}^{n} P_{k/n} = 1$$

Using Stirling's formula for factorials and ignoring an immaterial factor $(n/2\pi k(n-k))^{1/2}$, we can write

$$\binom{n}{k} \approx \left(\frac{k}{n}\right)^{-k}\left(1-\frac{k}{n}\right)^{k-n} = 2^{nH(k/n)} \tag{2}$$

where H is the entropy function, well known from thermodynamics and information theory:

$$H(\xi) := -\xi \log_2 \xi - (1-\xi)\log_2(1-\xi)$$

shown in Figure 2.

The fractal dimension of the set of all those subintervals having the common probability $(1-p)^k(p)^{n-k}$ is given by

$$\tilde{f}(\xi) = \lim_{n\to\infty} \frac{\log \binom{n}{k}}{n \log(1/r)} \tag{3}$$

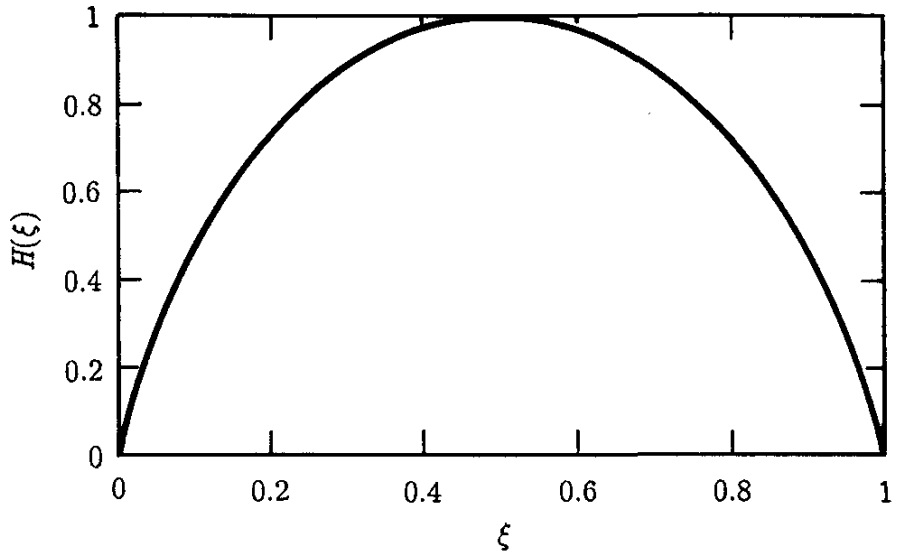


Figure 2 The entropy function.

where $\xi = k/n$. With $r = \frac{1}{2}$ and the approximation in equation 2, we have

$$\tilde{f}(\xi) = H(\xi) \tag{4}$$

Thus, depending on the value of ξ, representing a given probability, we get different fractal dimensions for the support of that probability. In fact, $\tilde{f}(\xi)$ ranges from the smallest value $\tilde{f}(\xi) = 0$ for the lowest and the highest densities ($\xi = 0$ and $\xi = 1$, respectively) to a maximum value $\tilde{f}(x) = 1$ for $\xi = 0.5$. This is one reason why such fractals are called *multifractals*. Note that the subsets of the unit interval that correspond to a given fractal dimension $\tilde{f}(\xi)$ are scattered all over the unit interval and are intimately intertwined with subsets of other dimensions. This is another characteristic feature of multifractals. Thus, for $\xi = 0.25$, for example, the fractal dimension $\tilde{f}(\xi)$ equals approximately 0.811 and the corresponding subsets are all those *points* $0 \leq x \leq 1$ for which the binary expansion of x has 25 percent 0s and 75 percent 1s. For example, $x = 0.0\overline{111} = \frac{7}{15}$ is one of uncountably many such points. Again, the fractal dimension of this Cantor-like dust equals 0.811 . . . .

After having considered the fractal dimension $\tilde{f}(\xi)$ of the *support* of a multifractal, we ask how the *probabilities* $(1 - p)^k p^{n-k}(2^{-n})$ of equation 1 scale as we let $n \to \infty$. For this purpose we introduce the *Lipshitz-Hölder exponent* $\alpha(\xi)$, which is defined to ensure that the product $p^k(1 - p)^{n-k} r^{-n\alpha(\xi)}$ does not diverge to zero or infinity as $n \to \infty$. Thus, the Lipshitz-Hölder exponent, which characterizes the singularities of the probabilities, is given by

$$\alpha(\xi) = \frac{\xi \log p + (1 - \xi) \log (1 - p)}{\log r} \tag{5}$$

where $\xi = k/n$, as before, and $r = \frac{1}{2}$ for our particular bisecting process.

As equation 5 shows, $\alpha(\xi)$ is a *linear* function of ξ that is monotonically increasing for $p < 0.5$. For $\xi = 0$, $\alpha = \alpha_{min} = -\log_2 (1 - p)$; and for $\xi = 1$, $\alpha = \alpha_{max} = -\log_2 p$. Thus, for $p = 0.3$, for example, the Lipshitz-Hölder exponent α ranges from $\alpha_{min} = 0.51$ to $\alpha_{max} = 1.74$. The value α_{min} represents the least probable part of the multifractal and α_{max} the most probable.

Although the fractal subsets of a multifractal are perfectly deterministic, as opposed to random fractals, they exhibit much less geometric regularity than the original Cantor set. For example, at the twelfth stage of construction, the triadic Cantor set consists of 2^{12} pieces of length $3^{-12} \approx 2^{-19}$, which form a regular geometric pattern. Its Hausdorff dimension, we recall, is $D \approx 0.631$.

A multifractal without holes having about the same Hausdorff dimension ($D \approx 0.629$) is characterized by binary fractions with a proportion of 1s equal to 3/19. Its nineteenth stage of construction consists of $\binom{19}{3} = 969$ pieces, which like the triadic Cantor set, have length 2^{-19}. But these intervals form a rather irregular pattern, as dictated by the binary fractions of length 19 containing three 1s. By contrast, *ternary* fractions with missing 1s, which describe the triadic Cantor set, result in a well-ordered pattern.

Another irregularity of our multifractal is betrayed by the *number* of pieces. If one estimated its Hausdorff dimension by the number of pieces at the nineteenth stage of construction, the result would be $\tilde{D} = \log 969/\log 2^{19} \approx 0.522$, which is considerably less than the asymptotic value $D \approx 0.629$. Even at the 190th stage of construction, generating $7.74 \cdot 10^{34}$ pieces, the estimated value of $D = 0.610$ still falls short of the final value by 3 percent. Such subsets of multifractals thus demonstrate how slowly estimates of fractal dimensions can converge if a fractal is not self-similar—even it the structure from which the fractal is derived is perfectly self-affine (Figure 1).

To illustrate the irregularity of such multifractals, Figure 3 shows successive stages of construction of the subset with $k = \lfloor n/2 \rfloor$ of the multifractal illustrated in Figure 1.

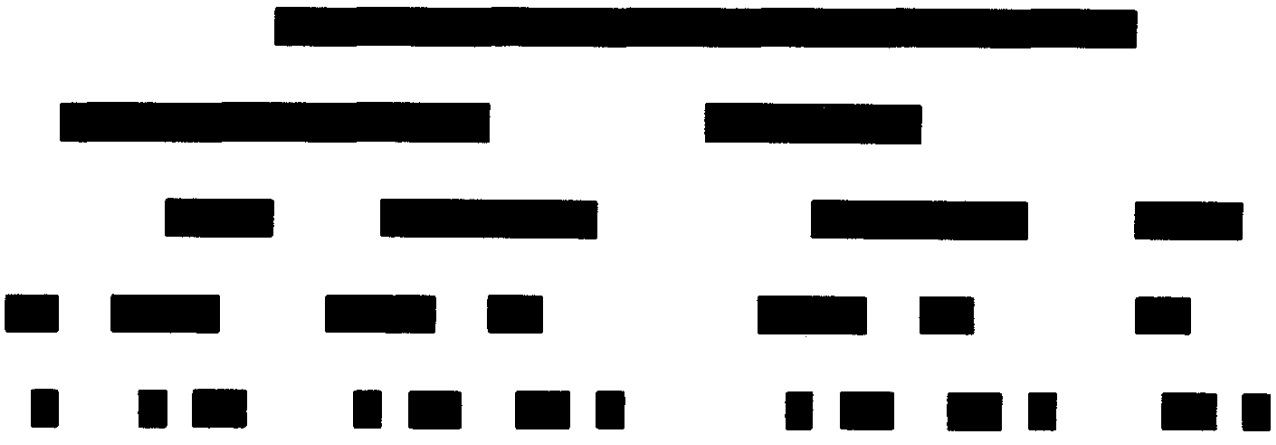


Figure 3 Successive stages of construction of the subset with the highest probability of occurence of the multifractal illustrated in Figure 1.

The Multifractal Spectrum: Turbulence and Diffusion-Limited Aggregation

In most applications, one has no direct access to the variable ξ [Fed 88]. In fact, ξ, being related to a very specific bisecting process, is often irrelevant. The important parameters for describing a multiplicative random process, like the one considered in the preceding sections, are the fractal dimension $\tilde{f}$ of the support, the Lipshitz-Hölder exponent α of the density distributions, and their relation $f(\alpha) := \tilde{f}(\xi(\alpha))$, called the "strength of the singularity" α, that is, the Hausdorff dimension of its support, or simply the *multifractal spectrum*. In our example, in which α is a linear function of ξ, the spectrum $f(\alpha)$ is simply a stretched and transposed version of $\tilde{f}(\xi)$ (see Figure 4).

The maximum of $f(\alpha)$ occurs for $\xi = 0.5$. According to equation 5, the corresponding value of α is $\alpha_0 = -\frac{1}{2}\log_2 p(1-p) = (\alpha_{\min} + \alpha_{\max})/2$. The value of the multifractal spectrum equals $f(\alpha_0) = 1$, that is, the Hausdorff dimension of the unit interval (or any interval). For multiplicative random processes on a fractal (as opposed to an interval) with Hausdorff dimension D, the highest value of $f(\alpha)$ equals D. In other words, the maximum of the multifractal spectrum $f(\alpha)$ equals the Hausdorff dimension of the support of the process.

Another special point of $f(\alpha)$ is the value $f(\alpha_1)$ at which its slope, $df/d\alpha$, equals 1. With

$$\frac{df}{d\alpha} = \frac{d\tilde{f}}{d\xi} \cdot \frac{d\xi}{d\alpha} = \frac{\log \xi - \log (1-\xi)}{\log p - \log (1-p)} \tag{6}$$

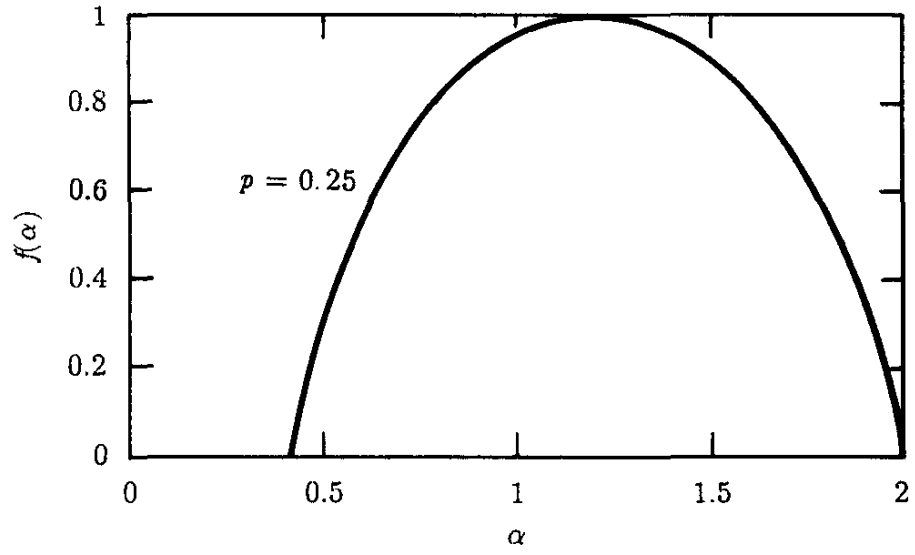


Figure 4 Multifractal spectrum of multiplicative random process.

we see that, for $df/d\alpha = 1$, the corresponding value of ξ is given by $\xi_1 = p$. With equation 5 we then obtain $\alpha(\xi_1) = \alpha_1 = H(p)$, which, according to equation 4, equals $\tilde{f}(p) = f(\alpha_1)$. Hence, $f(\alpha_1) = \alpha_1$, and $f(\alpha_1)$ lies on the tangent of the $f(\alpha)$ curve with slope 1 through the origin. The value $f(\alpha_1)$ equals the information dimension D_1 (see pages 203–207).

Figure 5 shows the multifractal spectrum for the energy dissipation in fully developed turbulence along a one-dimensional straight-line path through the turbulent flow [Man 87, Man 88]. The turbulent regions form the *support* of the multifractal. The experimental points are from different physical realizations of turbulence (such as atmospheric turbulence, boundary-layer turbulence, and turbulence in the wake behind a circular cylinder or wire grid). Note that these measurements are well matched by a single $f(\alpha)$ curve, the best match being obtained for $p = 0.3$. Thus, it seems that turbulence is indeed well modeled by multifractals as originally suggested by Mandelbrot [Man 74].

Another beautiful example of a multifractal phenomenon is diffusion-limited aggregation (DLA) as analyzed by Meakin and his coworkers [MSCW 85, MCSW

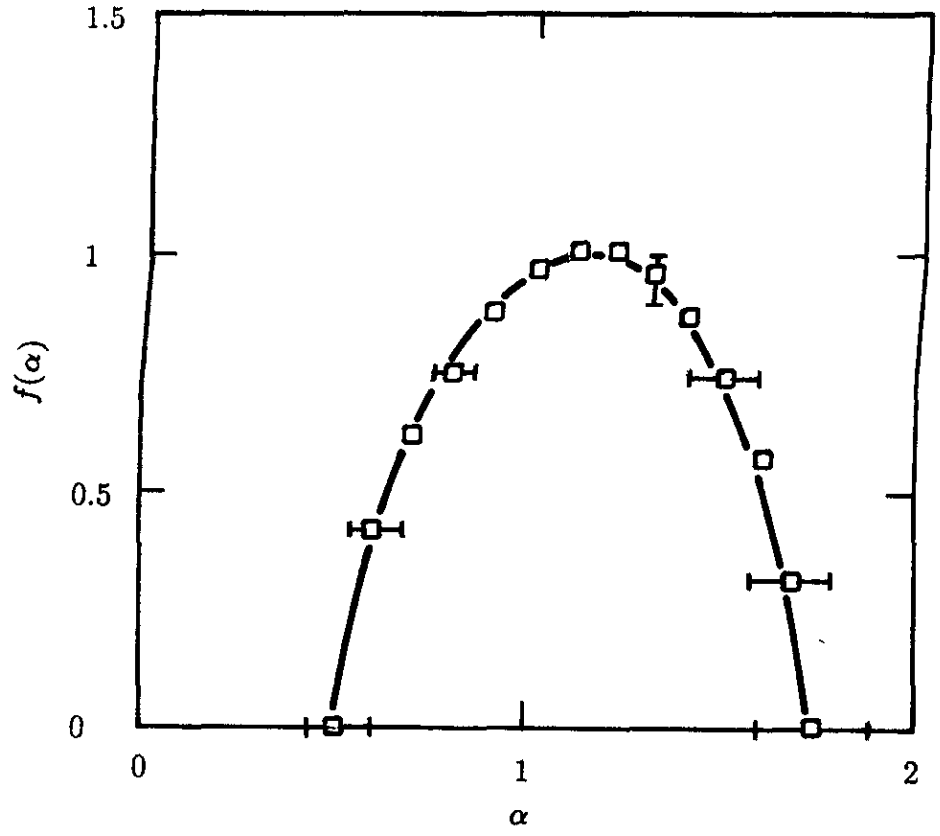


Figure 5 Multifractal spectrum in turbulence [MS 87].

86]. In DLA, single molecules perform a random walk until they become a "stuck" on the aggregate, producing attractive random fractals (see Figure 6), reminiscent of certain biological growth patterns and "Lichtenberg" figures of electrical breakdown on insulating surfaces (see Figure 7 [NPW 84]). These patterns are characterized by a dendritic design with "fjords" on many size scales. The reason for this structure in DLA is that a wandering molecule will settle preferentially near one of the tips of the fractal, rather than inside a deep fjord; the probability

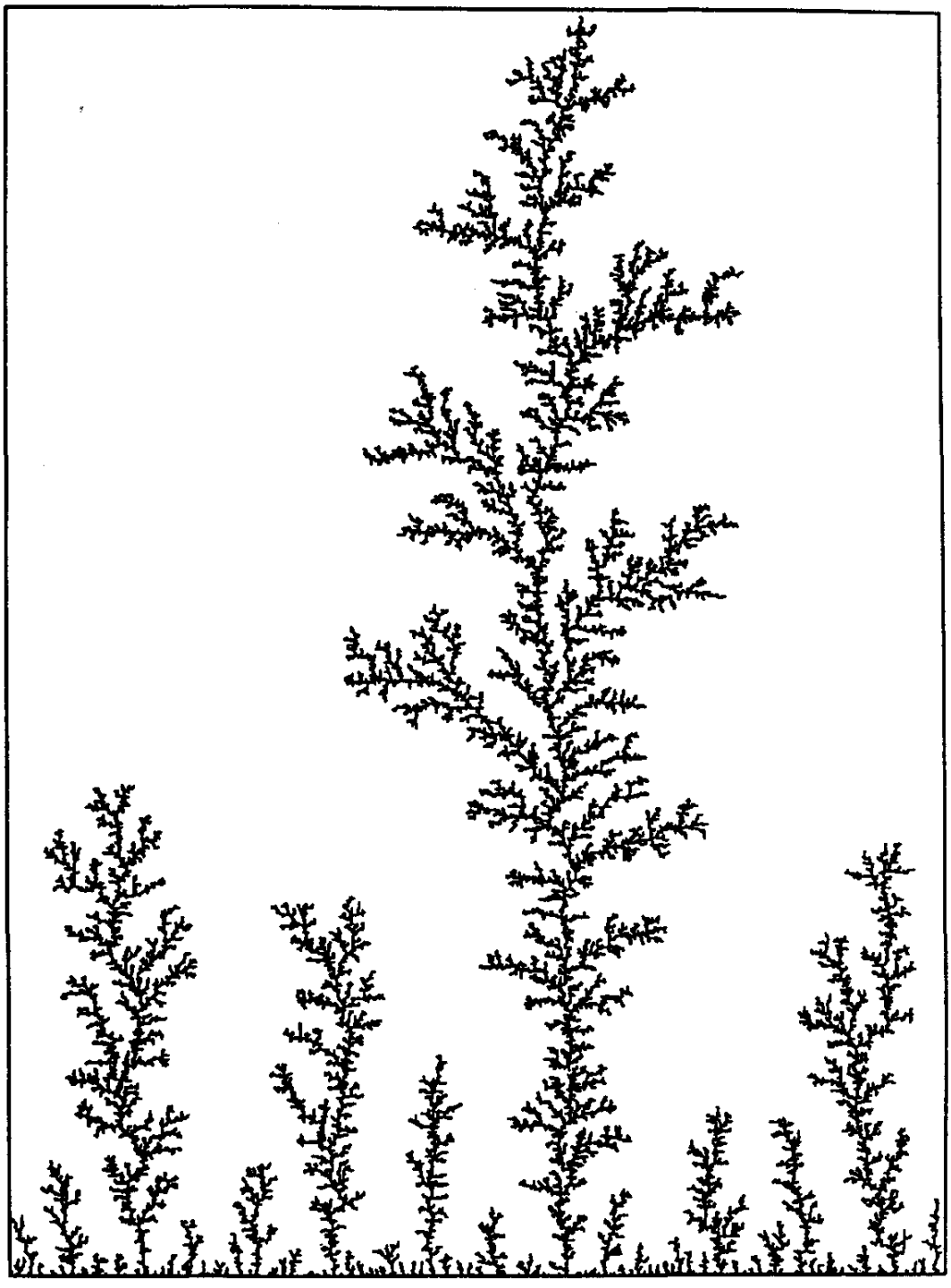


Figure 6 Crystal growth by diffusion-limited aggregation (DLA) [HFJ 87].

of penetrating a deep fjord without having become stuck earlier is simply too low. Thus, different sites have different growth probabilities, which are high near the tips and decrease with increasing depth inside a fjord. This is the precise paradigm for which multifractals are tailor-made.

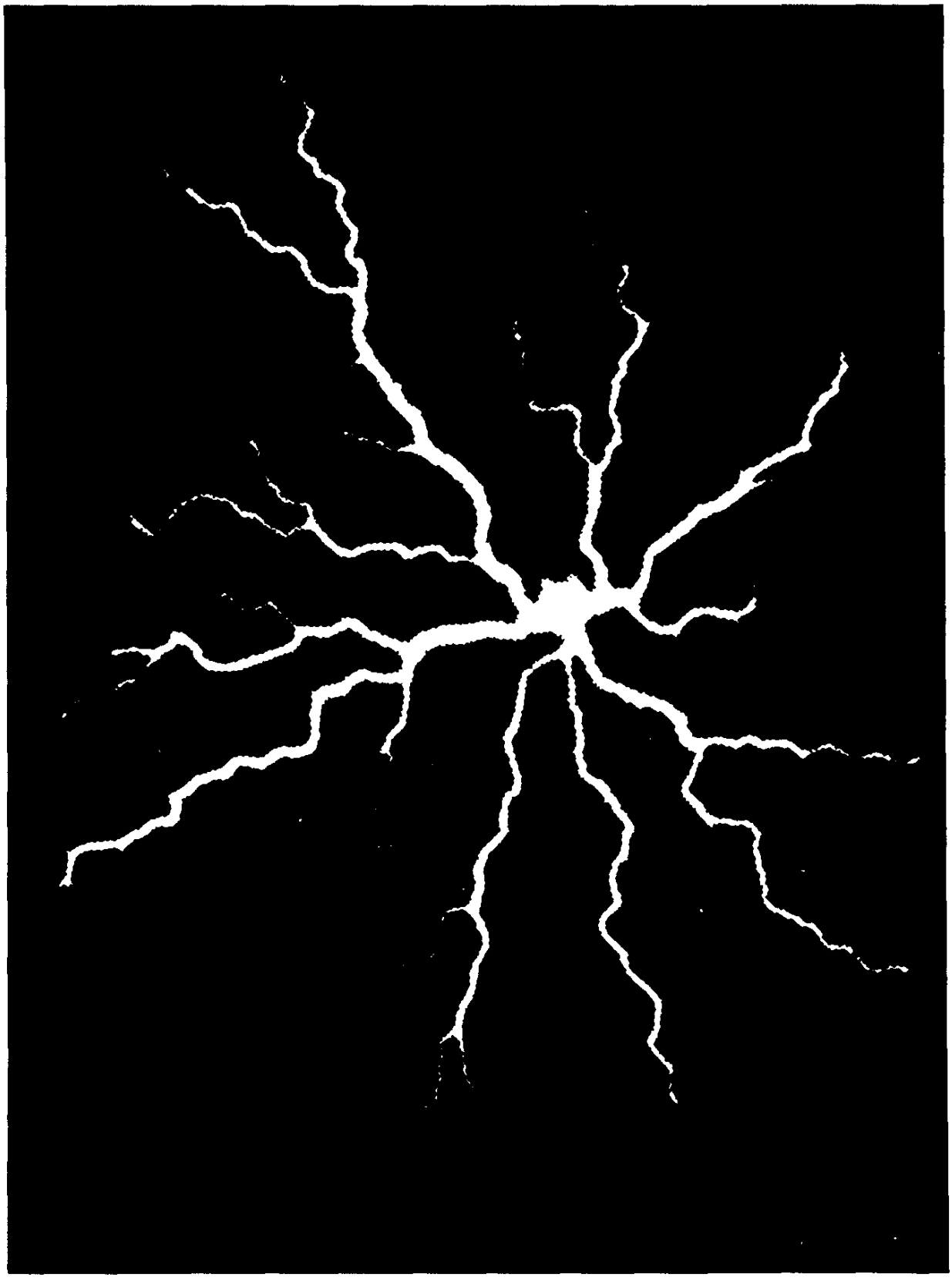


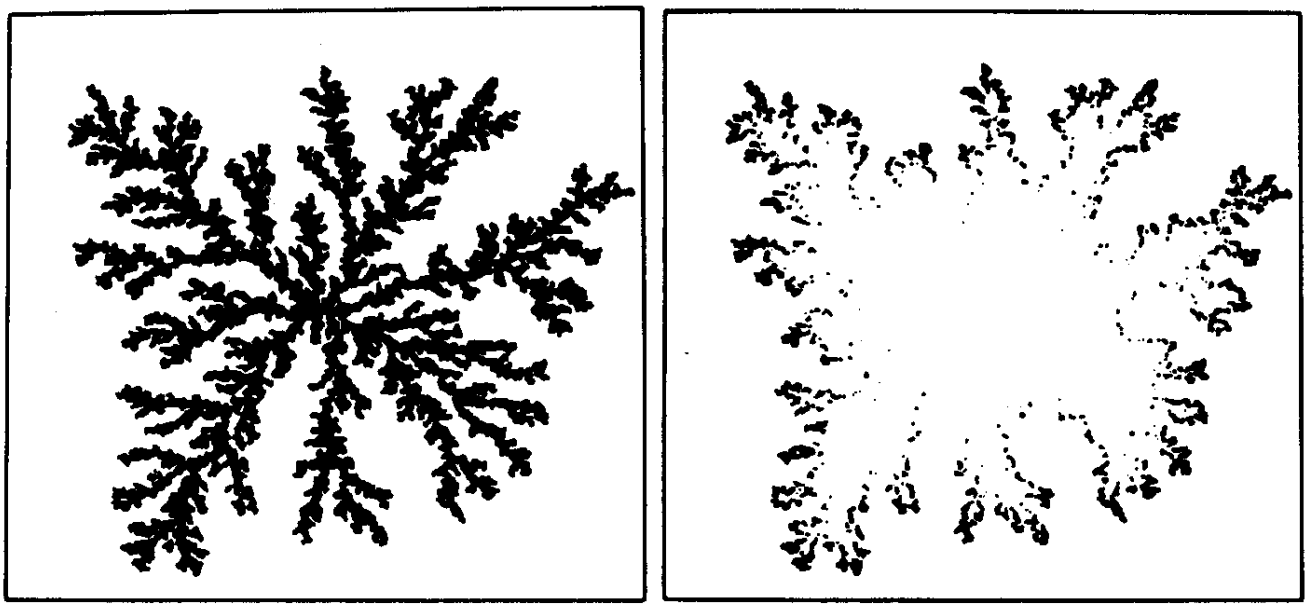
Figure 7 Lichtenberg figure of electrical discharge. Fractal dimension $D \approx 1.7$ [NPW 84].

Figure 8A shows a two-dimensional DLA cluster and Figure 8B shows the sites with a relatively high growth probability. As expected, these high-probability growth sites favor the tips of the cluster. The experimental multifractal spectrum $f(\alpha)$, which extends from $\alpha_{min} \approx 0.5$ to $\alpha_{max} > 5$, models the theoretical one quite well; see the review by Stanley and Meakin and the references cited there [SM 88].

Figure 9 shows the result of an early computer simulation of DLA by Witten and Sander. The fractal dimension for two-dimensional DLA is found to lie near 1.7. This means that the mass of the aggregate increases with its linear dimension L as $L^{1.7}$ and the average density goes as $L^{1.7}/L^2 = L^{-0.3}$—that is, it decreases, in accordance with the visual appearance of such growth patterns. In three-dimensional DLA the fractal dimension is typically near 2.5 [WS 83].

The visual similarity between Lichtenberg figures and DLA patterns is not accidental. Both processes are governed by the Laplace equation of potential theory, the gradient of the potential corresponding to the diffusion field in DLA. The surface of the DLA cluster is an equipotential surface. In this approach to DLA, particles will attach themselves preferentially at those sites of the cluster for which the potential gradient is high, which is near the tips.

In lightning and Lichtenberg figures,and similar electrical breakdowns, the potential is, of course, the *electrical* potential. The growth of a lightning stroke or a discharge occurs preferentially in the direction of the highest gradient of the potential. The deep "fjords" of the pattern, by contrast, are well shielded electrically and therefore experience little or no growth. This correspondence between potential theory and fractal growth has been fully confirmed in careful measurements and numerical solutions of the potential equation [NPW 84].



(A) (B)

Figure 8 (A) DLA cluster. (B) Sites with high probability of growth [MCSW 86].

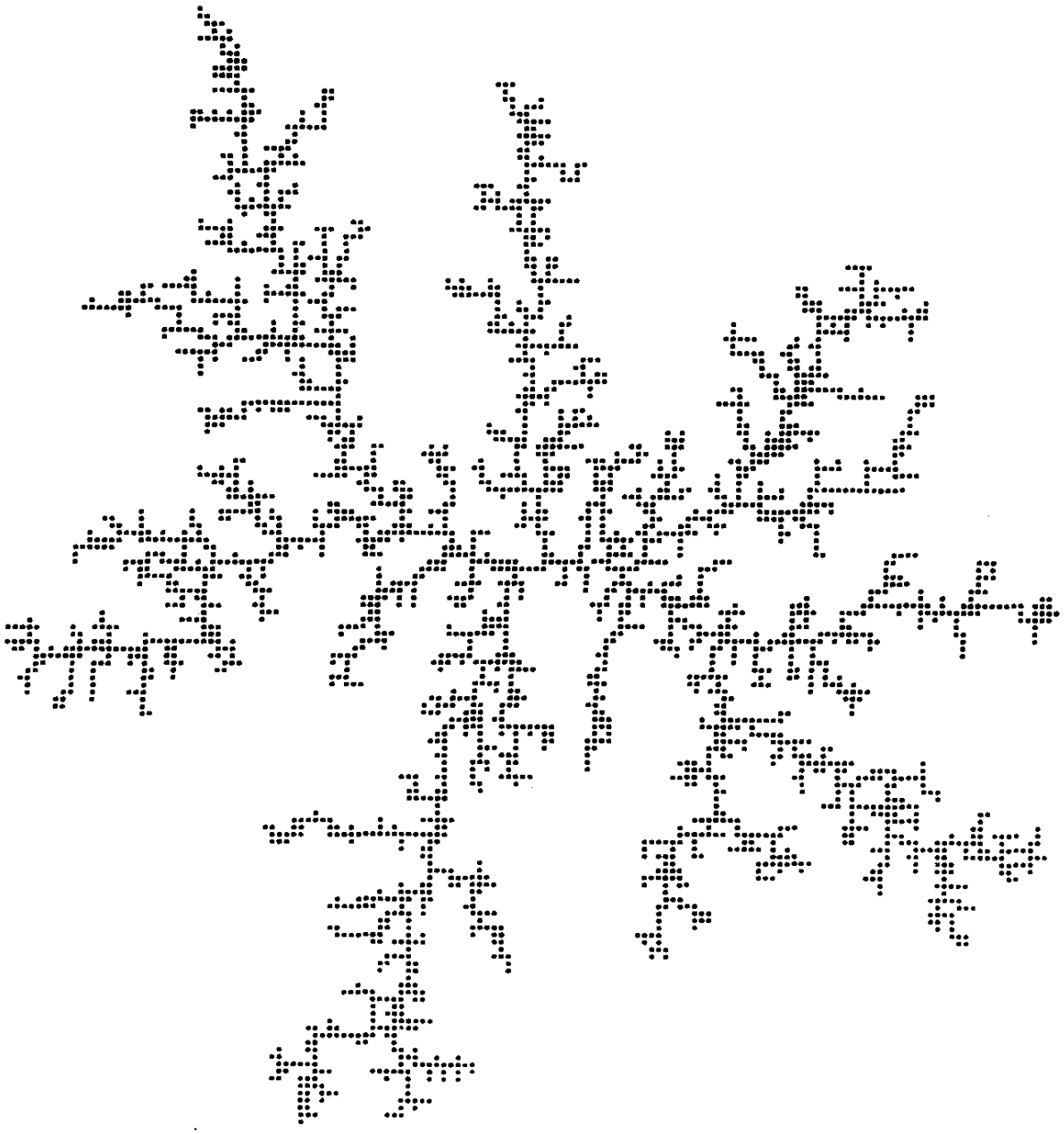
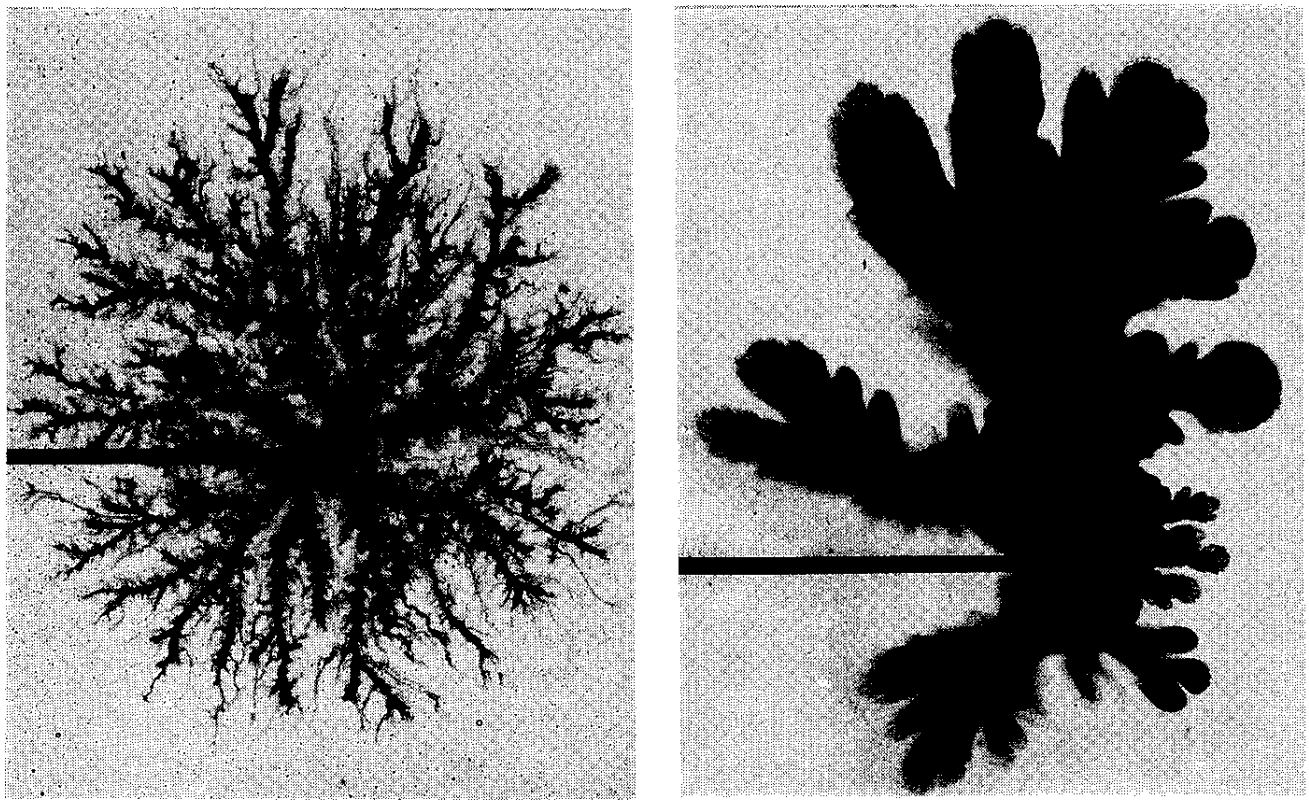


Figure 9 Fractal cluster grown by computer simulation of diffusion-limited aggregation [WS 83].

Viscous Fingering

A related problem is *viscous fingering*, observed most conveniently at the interface of two liquids between two glass plates. For two miscible liquids, like gelatin and water, a DLA-like structure appears when one liquid invades the other see (Figure 10A). This growth process is the result of a hydrodynamic instability between the liquids. As in DLA, any small "bump" on the interface will tend to grow predominantly a the tips because the pressure gradient, which drives the growth, is largest at the tips.



(A) (B)

Figure 10 Viscous fingering at the interface of two liquids. The shape of the "fingers" depends on the mutual miscibility of the liquids. (A) A dark-colored water was injected through a central cannula (visible as a black bar) into fluid gelatin, which is highly miscible with water. (B) A less miscible concentrated sugar solution was injected into the gelatin.

For two *immiscible* liquids, like glycerine and oil, the fingers are much wider because the surface tension between the two liquids prevents the formation of thin dendrites—that is, tips with high curvature (see Figure 10B). Because of this kind of fingering, much of the oil stays in the ground when water is injected with high pressure into oil-containing shale. By increasing the surface tension of the water through additives, the tips can be made even rounder, thereby lowering the fractal dimension and increasing the amount of oil that can be extracted before the first water arrives at the point where the oil is extracted.

Multifractals on Fractals

Considering the awe that multifractals have inspired in some quarters, the generalization that leads from fractals to multifractals is surprisingly simple. We will

next consider multifractals on a *fractal* support. Take the unit interval from which a number of open intervals have been removed, leaving individual line segments of lengths r_i, separated by empty "holes." We associate with each line segment r_i a weight or *probability* p_i. Iterating this generalized process of removal and assignment of probability, we arrive at a generalization of the Cantor set with probabilities associated with each speck of "dust." This is the prototypical multifractal on a fractal support that we want to study in this chapter.

Associating a probability p_i with each segment r_i will allow us to model fractal growth processes in which the different segments correspond to the different sites at which growth takes place. The probabilities p_i represent the different growth rates at these sites as in diffusion-limited aggregation (see pages 193–199). In the application to strange attractors, the segments r_i converge on the different values a dynamic variable can assume, called the support of the attractor, while the probabilities p_i model the frequencies with which the segments are "visited."

While a self-similar Cantor set, generated from segments of equal length, is characterized by a single scaling exponent, the Hausdorff dimension D, multifractals are described by *two* scaling exponents, one for the supporting fractal and one for the probabilities.

To properly introduce these two scaling exponents, we first recall the definition of the Hausdorff dimension. The Hausdorff dimension D of a set is given by the limit as $r \to 0$ of the expression $\log N/\log (1/r)$, where N is the smallest number of pieces of diameter r to completely cover the set:

$$D := \lim_{r \to 0} \frac{\log N}{\log (1/r)} \tag{7}$$

This definition of D can also be rendered in the following implicit form:

$$\lim_{r \to 0} Nr^D = c \qquad 0 < c < \infty \tag{8}$$

where c is a constant. Equation 8 brings out an important property of the Hausdorff dimension: it is *the* exponent that keeps the product Nr^D finite and nonzero as $r \to 0$. If D is altered even by an infinitesimal amount, this product will diverge either to 0 or to ∞.

In the recursive construction of a self-similar set, the number of pieces N after n iterations is N_G^n, where N_G is the number of pieces of the generator. Similarly, r equals r_G^n, *where* r_G is the length of the segments of the generator. (These segments are assumed to have equal lengths at this point.) Thus, instead of equation 8, we may write

$$\lim_{n \to \infty} (N_G r_G^D)^n = c \tag{9}$$

which of course requires

$$N_G r_G^D = 1$$

or

$$D = \frac{\log N_G}{\log (1/r_G)} \tag{10}$$

Hence, for a strictly self-similar set it is not necessary to take the limit as $r \to 0$ as in equation 7. It suffices to use the parameters of the generator N_G and r_G.

For generator segments of different length r_i, equation 9 becomes

$$\lim_{n \to \infty} \left(\sum_{i=1}^{N} r_i^D \right)^n = c \tag{11}$$

which implies

$$\sum_{i=1}^{N} r_i^D = 1 \tag{12}$$

For the generalized generator with line segments r_i and probability p_i, we introduce *two* exponents, the exponent τ for the support intervals r_i, and the exponent q for the probabilities p_i. Thus, instead of the limit in equation 11, we consider

$$\lim_{n \to \infty} \left(\sum_{i-1}^{N} p_i^q r_i^\tau \right)^n \tag{13}$$

and we ask for the values of q and τ for which expression 13 stays finite—in other words, which q and τ satisfy

$$\sum_{i-1}^{N} p_i^q r_i^\tau = 1 \tag{14}$$

It is obvious from equation 14 that there are no unique values of q and τ. Rather, there is a continuous range of exponents $\tau = \tau(q)$ corresponding, as we shall see, to a continuum of fractal dimensions.

In the case of the original Cantor set ($N = 2$, $r_i = \frac{1}{3}$), equation 14 reads, with $p_i = \frac{1}{2}$,

$$\left(\frac{1}{2}\right)^q \left(\frac{1}{3}\right)^\tau + \left(\frac{1}{2}\right)^q \left(\frac{1}{3}\right)^\tau = 1$$

which has the solution $\tau = (1 - q) \log 2/\log 3$. Injecting the Hausdorff dimension $D = \log 2/\log 3$ for the original Cantor set, we can also write $\tau = (1 - q)D$.

This relation between τ and D is also borne out by comparing expression 13 with equation 11: for $q = 0$, τ corresponds to D, in agreement with the relation $\tau = (1 - q)D$. This comparison suggests that $\tau/(1 - q)$ may play the role of a *generalized dimension* D_q that agrees with the Hausdorff dimension for $q = 0$ but may be different for other values of q.

Another method of creating new dimensions is discussed in the following section.

Fractal Dimensions from Generalized Entropies

In his attempt to generalize the concept of entropy of a probability distribution, the Hungarian mathematician A. Rényi introduced the following expression based on the moments of order q of the probabilities p_i:

$$S_q := \frac{1}{q-1} \log \sum_{i=1}^{N} p_i^q \tag{15}$$

where q is not necessarily an integer [Ren 55]. For $q \to 1$, the definition in equation 15 yields the well-known entropy

$$S_1 = -\sum_{i=1}^{N} p_i \log p_i \tag{16}$$

of a discrete probability distribution. The definition in equation 15 can therefore be considered, as was Rényi's intent, a *generalized entropy*.

Taking a cue from Rényi, we define the generalized dimensions

$$D_q := \lim_{r \to 0} \frac{1}{q-1} \frac{\log \sum_{i=1}^{N} p_i^q}{\log r} \tag{17}$$

where p_i is the probability that the random variable falls into the ith "bin" of size r. The parameter q ranges from $-\infty$ to $+\infty$. Note that, for a self-similar fractal with equal probabilities $p_i = 1/N$, the definition in equation 17 gives $D_q = D_0$ for all values of q. For such a fractal, going to the limit as $r \to 0$ is not necessary. Thus

$$D_q = \frac{1}{q-1} \frac{\log N(1/N)^q}{\log r} = \frac{\log N}{\log (1/r)}$$

which is independent of q.

It is clear that for $q = 0$, the definition in equation 17 agrees with that for the Hausdorff dimension D. For this reason we call the D_q *generalized dimensions*,

hoping that they will prove another potent tool in describing multifractals. This is indeed the case. In fact, the D_q are uniquely related to the two exponents q and τ for the general multifractal.

This relation can easily be deduced from the limit in expression 13 by introducing N *constant* bin sizes $r_i = r$, which, for $n \to \infty$, does not affect the values of τ and q for which the limit in expression 13 converges. Thus,

$$\tau = \tau(q) = -\lim_{r \to 0} \frac{\log \sum_{i=1}^{N} p_i^q}{\log r}$$

and, with the definition in equation 17.

$$\tau(q) = (1 - q)D_q \tag{18}$$

For a self-similar fractal, the dimensions D_q can be obtained directly form the p_i and r_i of the generator using equation 14 and the identity in equation 18:

$$\sum_{i=1}^{N} p_i^q r_i^{(1-q)D_q} = 1 \tag{19}$$

For $q = 1$, $\tau(q) = 0$ and D_q is given by

$$D_1 = -\left.\frac{d\tau}{dq}\right|_{q=1}$$

which, with equation 14, becomes

$$D_1 = \frac{\sum_{i=1}^{N} p_i \log p_i}{\sum_{i=1}^{N} p_i \log r_i} \tag{20}$$

or, for N equal probabilities $p_i = 1/N$,

$$D_1 = \frac{N \log N}{\sum_{i=1}^{N} \log (1/r_i)} \tag{21}$$

For $q \to 1$, the definition in equation 17 yields

$$D_i = \lim_{r \to 0} \frac{S_1}{\log r}$$

where S_1 is the entropy of the probabilities p_i given by equation 16.

This entropy and D_1, also called the *information dimension*, play an important role in the analysis of nonlinear dynamic systems, especially in describing the loss of information as a chaotic system evolves in time. In this context, the entropy S_1 is called the *Kolmogorov entropy*.

For $q = 2$, equation 17 yields the so-called *correlation dimension*

$$D_2 = \lim_{r \to 0} \frac{\log \sum_{i=1}^{N} p_1^2}{\log r}$$

which, in addition to D_0 and D_1, is another important fractal dimension. Its main practical advantage is the relative ease with which it can be determined for"practical" fractals (see Chapter 10). The theoretical importance of D_2 lies in its close relation with the fundamental concept of correlation. In fact, we will show in Chapter 10 that D_2 is determined by the "correlation function" of the fractal set, that is, the probability of finding, within a distance r of a given member of the set, another member. Thus, measuring D_2 comes down to a simple counting process.

In principle D_q can be determined for all q in accordance with its definition (equation 17). In practical applications, however, one sometimes encounters difficulties for $q > 0$ because positive q diminish the terms with small p_i (corresponding to the "rarely visited" parts of the fractal). As a result the limit as $r \to 0$ converges very slowly. This drawback can be overcome by calculating the numerator in equation 17 for both r and $r/2$ and requiring that their *ratio* equal 1 as $r \to 0$; see Halsey et al. [HJKPS 86].

The Relation between the Multifractal Spectrum $f(\alpha)$ and the Mass Exponents $\tau(q)$

In the preceding sections, we have introduced two different functions for describing a multifractal:

- the multifractal spectrum $f(\alpha)$ that describes the fractal dimension f of a subset with a given Lipshitz-Hölder mass exponent α, and
- the generalized fractal dimensions D_q or, equivalently, the exponents $\tau(q) = 1 - q)D_q$

Since both functions, $f(\alpha)$ and $\tau(q)$, describe the same aspects of a multifractal, they must be related to each other. In fact, the relationships are

$$\tau(q) = f(\alpha) - q\alpha \tag{22}$$

where α is given as a function of q by the solution of the equation

$$\frac{d}{d\alpha}(q\alpha - f(\alpha)) = 0 \tag{23}$$

Conversely, if the fractal dimension D_q or the exponents $\tau(q)$ are known, the multifractal spectrum is given by

$$f(\alpha(q)) = \tau(q) + q\alpha(q) \tag{24}$$

where $\alpha(q)$ is given by

$$\alpha(q) = -\frac{d}{dq}\tau(q) \tag{25}$$

which, with equation 24, implies

$$\frac{df}{d\alpha} = q \tag{26}$$

Equations 24 and 25 give a parametric description of the $f(\alpha)$ curve in terms of q. These two equations represent a *Legendre transform* from the variables q and τ to the variables α and f. Such transformations play an important role in thermodynamics in converting, for example, energy as function of volume and entropy into free energy as a function of volume and temperature.

In fact, the analogies between multifractals and statistical mechanics go much further than a change of variables mediated by the Legendre transform. As the physicist will appreciate, equation 14 is patterned after one of the most powerful analytical tools in statistical mechanics, the *partition function* invented by the noted theoretical physicist and chemist J. Willard Gibbs (1839–1903). Because of this mathematical analogy, several of our parameters are formally equivalent to such thermodynamic concepts as energy (α), free energy (τ/q), entropy (f), and temperature ($1/q$).

Strange Attractors as Multifractals

Strange attractors are among the most important realizations of multifractals. An attractor of an iteration $x_{n+1} = f(x_n)$ is a single point or an indecomposable bounded set of points to which starting values x_0 from the attractor's "basin of attraction" converge as $n \to \infty$. A *strange* attractor is an attractor for which the iterates x_n depend sensitively on the initial x_0; that is, arbitrarily close initial values will become macroscopically separated for a sufficiently large n. Strange attractors are fractal dusts with a fractal dimension smaller than the Euclidean dimension of the embedding space.

A prototypical strange attractor is furnished by the *logistic parabola* (see Chapter 12) defined by the "quadratic map" $f(x) = rx(1 - x)$. For $r < r_\infty = 3.5699456\ldots$, the iterates of $f(x)$, $f^{(n)}(x)$, with $0 \leq x \leq 1$, are periodic with period length 2^m. But for $r = r_\infty$, the iterates are aperiodic and converge on strange attractor. This strange attractor is a Cantor set that is asymptotically well modeled by a generator with two intervals with lengths $r_1 = 0.408$ and $r_2 = r_1^2$, and with equal weight $p_1 = p_2 = 0.5$.

The Hausdorff dimension $D = D_0$ of this model attractor is given by equation 19 for $q = 0$:

$$r_1^D + r_2^D = 1$$

which, for $r_2 = r_1^2$, has the solution

$$r_1^D = \frac{\sqrt{5} - 1}{2} = 0.618\ldots$$

Thus, $D = D_0 \approx \log 0.618/\log 0.408 \approx 0.537$.

The information dimension is given by equation 21. With $N = 2$ and the values for r_i just used one obtains $D_1 = 0.515$. For all other q and $p_1 = p_2 = 0.5$, equation 19 gives

$$r_1^\tau + r_1^{2\tau} = 2^q$$

which yields

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

With $r_1 = 0.408$, this gives $D_{-\infty} = -1/\log_2 r_1 = 0.773, \ldots, D_{-1} = 0.561$, $D_0 = 0.537$, $D_2 = 0.497$, $D_3 = 0.482, \ldots, D_\infty = -1/\log_2 r_1^2 = 0.387$.

The entire multifractal spectrum $f(\alpha)$, calculated from D_q, is shown in Figure 11. Because $df(\alpha)/d\alpha$ equals q, the maximum of $f(\alpha)$ equals $D_0 = 0.537$, while $f(0)$ corresponds to $df(\alpha)/d\alpha = \pm\infty$, yielding the values $\alpha_{\min} = D_\infty = 0.387$ and $\alpha_{\max} = D_{-\infty} = 0.774$, which are within 2.5 percent of the best numerical values [HJKPS 86]. This discrepancy, albeit small, reflects the fact that, contrary to our assumption, the period-doubling strange attractor is not exactly self-similar.

A Greedy Algorithm for Unfavorable Odds

In Chapter 6 (pages 150–152), we outlined a strategy for optimally playing a game of chance when the odds are favorable. And we advised against gambling when the odds are unfavorable. Nevertheless, for those who persist in the face

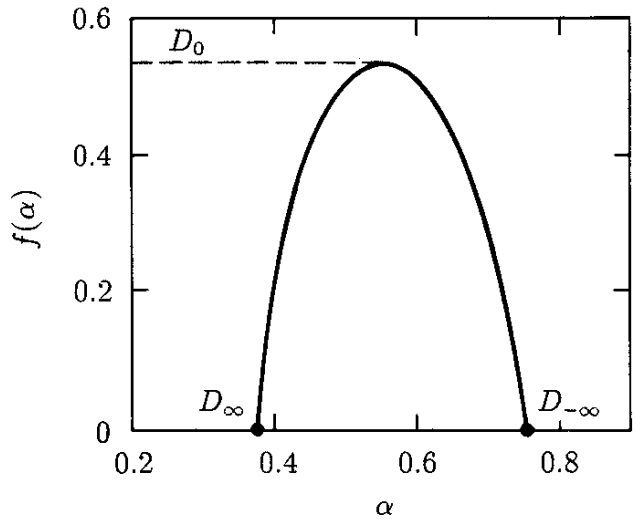


Figure 11 Multifractal spectrum of iterates of quadratic map at period-doubling accumulation point [HJKPS 86].

of adversity and are willing to risk everything, there is a small chance of "making a killing" if they follow a simple rule based on maximum avarice: a *greedy algorithm.*

Suppose a player has an initial capital of K dollars and the bank has B dollars, and the probability of winning a single play is p, with $p < 0.5$. Of course, a reasonable person, intent on a positive expected gain, *should not* play against odds that are less than even ($p = 0.5$). But suppose our player is willing to risk his entire capital for a small chance of winning all the money in the bank. From equation 8 in Chapter 6 we see that the probability of achieving this goal, while betting always exactly one dollar, is given by

$$p_K = 1 - q_K = \frac{r^K - 1}{r^B - 1} \tag{27}$$

where $r := (1 - p)/p > 1$. An unbiased roulette wheel with 36 positive numbers and only one 0 gives $p = \frac{18}{37}$, or $r = \frac{19}{18}$, for an "even" chance. More generally, K in equation 27 is the player's initial capital divided by the size of his bet and B is the bank's capital divided by the bet.

Say that the player has an initial capital of \$1000 ($K = 1000$) and the bank has \$10,000 ($B = 10{,}000$); then, according to equation 27, the probability p_K of breaking the bank is approximately 10^{-211}. In other words, our player looses all his money with probability indistinguishable from certainty.

A less timid player might wager \$10 on every play, thereby changing K to 100 and B to 1000 in equation 27. This "improves" his chances of winning \$10,000 to $p_K \approx 10^{-21}$. And an even bolder player, always wagering \$100, corresponding to $K = 10$ and $B = 100$, is rewarded with $p_K \approx 0.0025$. This suggests that the higher the wager, the better the chances of winning the pot.

Why should this be so? Obviously, a player who wagers very little, like \$1 or \$10, has to play very often reaching \$10,000 (*if ever*), and every time one plays, one gives the unfavorable odds another chance to come into play. This

suggests that always betting all one's money would be an optimum strategy in the face of unfavorable odds. More specifically, as long as $x = K/B$ is smaller than $\frac{1}{2}$, one would bet all and, upon winning, would double one's money. Thus, if $P(x)$ is the probability under this greedy algorithm of ultimately winning all, then we have the following functional recursion:

$$P(x) = pP(2x) + (1 - p) \cdot 0 \qquad 0 \leq x \leq \frac{1}{2} \tag{28}$$

Of course, if $x > \frac{1}{2}$, one only has to wager the difference $B - xB$ to reach B. In this case, even if one loses, one still has $K - (B - K) = 2K - B$ dollars to continue playing [Bill 83]. Hence,

$$P(x) = p \cdot 1 + (1 - p)P(2x - 1) \qquad \frac{1}{2} \leq x \leq 1 \tag{29}$$

Now the only remaining question is how to calculate $P(x)$ from the equations 28 and 29. Of course, for $0 < p < \frac{1}{2}$, $P(0) = 0$ by equation 28 and $P(1) = 1$ by equation 29. For other values of x, we note that $P(x)$ must be a *self-affine* function. Indeed, according to equation 28, compressing x by a factor of 2 and multiplying P by p reproduces the left half of $P(x)$.

In fact, $P(x)$ is precisely the function that we encountered earlier in this chapter (with p replaced by $1 - p$, see Figure 1 and equation 1). However, like many other self-similar or self-affine functions, $P(x)$ is not well behaved: its derivative vanishes almost everywhere except at values of x which form a Cantor set. In other words, $P(x)$ is a devil's staircase—see Figure 12, where $P(x)$ is plotted for $p = 0.25$.

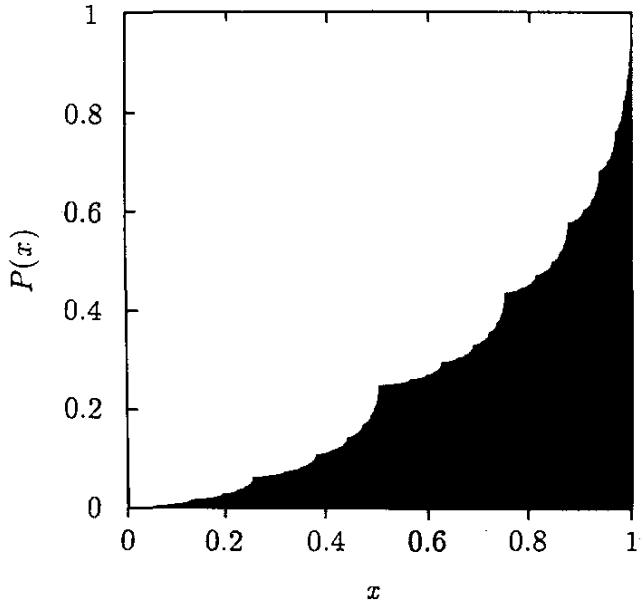


Figure 12 The cumulative probability of winning the pot when playing according to a "greedy" algorithm: a self-affine devil's staircase.

To calculate $P(x)$, we express x as a binary fraction. If this fraction does not terminate, we truncate it after n places. In our example, $x = \frac{1}{10}$ has the periodic binary expansion $x = 0.0\overline{0011}$, which we can approximate by $\tilde{x} = 0.00011 = \frac{3}{32}$. Now $P(\tilde{x})$ is given by the following expression:

$$P(\tilde{x}) = p_0p_0p_0p_0p_0 + p_0p_0p_0p_0p_1 + p_0p_0p_0p_1p_0 \tag{30}$$

where $p_0 = p$ and $p_1 = (1 - p)$. The three sequences of indices correspond to all binary fractions with $n = 5$ places which are smaller than $\tilde{x} = 0.00011$; that is, 0.00000, 0.00001, and 0.00010. Thus, for $p_0 = \frac{18}{37}$ and $p_1 = \frac{19}{37}$, $P(\tilde{x}) = p_0^4(p_0 + 2p_1) \approx 0.084$. Since $\tilde{x} < x$ and $P(x)$ is a nondecreasing function, $P(x)$ is actually somewhat larger than 0.084. This is quite an improvement over the winning probability $p_K = 0.0025$ for constant \$100 bets. And even constant \$1000 bets give only $p_K = 0.077$.

The three terms in equation 30 reflect the three different routes to success under our greedy algorithm if we stop after five plays. For example, the third term, $p_0p_0p_0p_1p_0$, represents three wins, followed by a loss and a final win. The corresponding capital sequence is, in thousands of dollars: 1, 2, 4, 8, 6, 10. The capital sequence corresponding to the second term in equation 30, $p_0p_0p_0p_0p_1$, is 1, 2, 4, 8, 10, 10. The first term, $p_0p_0p_0p_0p_0$, has the same capital sequence because once the goal (\$10,000) is reached, the player stops playing.

Of course, if the player isn't ruined after five plays, he can continue to play and improve his chances of breaking the bank. This corresponds to truncating x in $P(x)$ after more than five places. For example, $\frac{51}{512}$ is an excellent approximation to $\frac{1}{10}$, giving $P(\frac{51}{512}) > 0.09$. For irrational x, which have aperiodic binary fractions, the calculation of $P(x)$ is of course an infinite process. But for rational x, with periodic binary fractions, there is a closed formula, which we leave the reader to discover. So what is $P(\frac{1}{10})$?

And what is the expected duration of the game until either the player or the bank is broke under the greedy algorithm? Like $P(x)$, it should be a self-affine function, and it should be less than the expected duration of the game with any constant wager [Fel 68]:

$$D_K = \begin{cases} \dfrac{1}{p(r-1)}\left[K - B\dfrac{r^K - 1}{r^B - 1}\right] & r \neq 1 \\ K(B - K) & r = 1 \end{cases} \tag{31}$$

For $p = \frac{18}{37}$, $r = \frac{19}{18}$, and \$1 bets, $K = 1000$ and B = 10,000, giving an expected duration of $D_K = 37{,}000$ plays. For \$100 wagers, $K = 10$ and B = 100, giving a duration of 358 plays. And for \$1000 wagers, $K = 1$ and $B = 10$, resulting in a duration of only 25 plays. Thus, the greedy algorithm should have an expected duration of less than 25 plays.