

# FRACTALS, CHAOS, POWER LAWS

*Minutes  
from an  
Infinite  
Paradise*

MANFRED SCHROEDER

*Ich sage euch:  
man muss noch Chaos in sich haben,  
um einen tanzenden Stern gebären zu können.*

*Ich sage euch:  
ihr habt noch Chaos in euch.*

*Yea verily, I say unto you:  
A man must have Chaos yet within him  
To birth a dancing star.  
I say unto you:  
You have yet Chaos in you.*

—FRIEDRICH NIETZSCHE,  
*Thus Spake Zarathustra*

QC 174  
.17 S9  
S 345

# Fractals, Chaos, Power Laws

*Minutes from an Infinite Paradise*

Manfred Schroeder

centro universitario de  
comunicación y ciencia



W. H. Freeman and Company  
New York

QC 174  
·1759  
3345

Library of Congress Cataloging-in-Publication Data

Schroeder, M. R. (Manfred Robert), 1926-

Fractals, chaos, power laws: minutes from an infinite paradise/

Manfred Schroeder.

p. cm.

Includes bibliographical references and index.

ISBN 0-7167-2136-8

1. Symmetry (Physics) 2. Self-similarity 3. Recursion 4. Scaling I. Title.

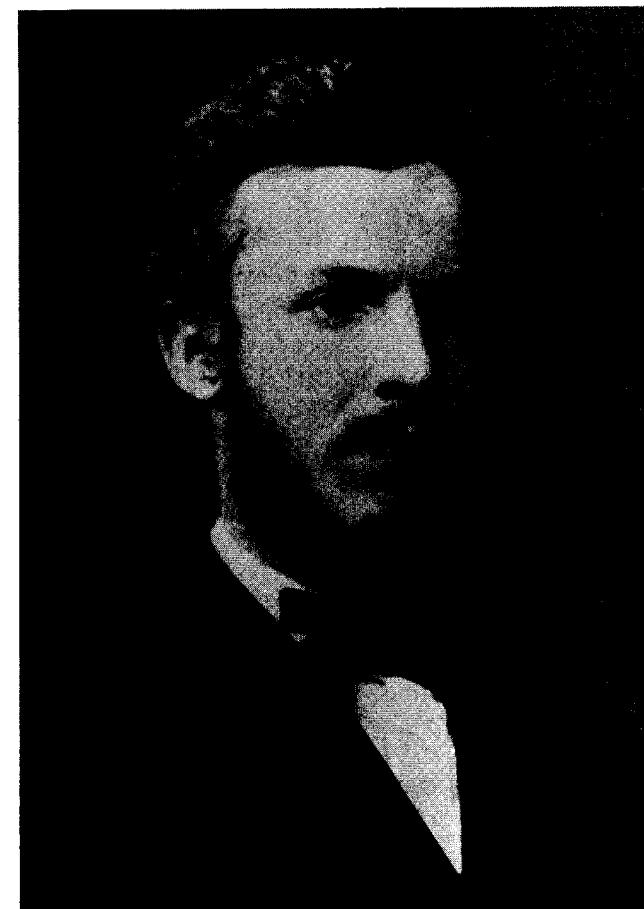
QC174.17.S9S38 1990

530.1—dc20

90-36763

Dedication page: Portrait of Georg Cantor. (Akademie der Wissenschaften zu Göttingen)

To Georg Cantor



Copyright © 1991 by W. H. Freeman and Company

No part of this book may be reproduced by any mechanical, photographic, or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted, or otherwise copied for public or private use, without written permission from the publisher.

Printed in the United States of America

1 2 3 4 5 6 7 8 9 VB 9 9 8 7 6 5 4 3 2 1

From his paradise no one shall ever evict us

DAVID HILBERT, defending Cantor's set theory

# C ONTENTS

## Preface

xiii

## Other titles of interest from W. H. Freeman and Company

**The Fractal Geometry of Nature**

by Benoit B. Mandelbrot, Thomas J. Watson Research Center  
 1982, 460 pages, \$39.95

**Visions of Symmetry: Notebooks, Periodic Drawings, and Related Work of M. C. Escher**

by Doris Schattschneider, *Moravian College*  
 1990, 354 pages, \$39.95

**Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions**

A Scientific American Library Book  
 by Thomas F. Banchoff, *Brown University*  
 1990, 247 pages, \$32.95

**Islands of Truth: A Mathematical Mystery Cruise**

by Ivars Peterson  
 1990, 325 pages, \$19.95

**The New Ambidextrous Universe: Symmetry and Asymmetry from Mirror Reflections to Superstrings**

Revised Edition  
 by Martin Gardner  
 1990, 392 pages, \$19.95

**Tilings and Patterns: An Introduction**

by Branko Grunbaum, *University of Washington*  
 and G. C. Shephard, *University of East Anglia*  
 1989, 446 pages, \$26.95

**1 INTRODUCTION**

1

- Einstein, Pythagoras, and Simple Similarity 3
- A Self-Similar Array of Self-Preserving Queens 4
- A Self-Similar Snowflake 7
- A New Dimension for Fractals 9
- A Self-Similar Tiling and a "Non-Euclidean" Paradox 13
- At the Gates of Cantor's Paradise 15
- The Sierpinski Gasket 17
- Sir Pinski's Game and Deterministic Chaos 20
- Three Bodies Cause Chaos 25
- Strange Attractors, Their Basins, and a Chaos Game 27
- Percolating Random Fractals 30
- Power Laws: from Alvarez to Zipf 33
- Newton's Iteration and How to Abolish Two-Nation Boundaries 38
- Could Minkowski Hear the Shape of a Drum? 40
- Discrete Self-Similarity: Creases and Center Folds 45
- Golden and Silver Means and Hyperbolic Chaos 49
- Winning at Fibonacci Nim 53
- Self-Similar Sequences from Square Lattices 55
- John Horton Conway's "Death Bet" 57

**2 SIMILARITY AND DISSIMILARITY**

61

- More Than One Scale 61
- To Scale or Not to Scale: A Bit of Biology and Astrophysics 63
- Similarity in Physics: Some Astounding Consequences 66
- Similarity in Concert Halls, Microwaves, and Hydrodynamics 68
- Scaling in Psychology 70
- Acousticians, Alchemy, and Concert Halls 72
- Preference and Dissimilarity: Concert Halls Revisited 74

PCC - 3562

<b>3 SELF-SIMILARITY—DISCRETE, CONTINUOUS, STRICT, AND OTHERWISE</b>	81	<b>The Clustering of Poverty and Galaxies</b> 152 <b>Levy Flights through the Universe</b> 155 <b>Paradoxes from Probabilistic Power Laws</b> 155 <b>Invariant Distributions: Gauss, Cauchy, and Beyond</b> 157
The Logarithmic Spiral, Cutting Knives, and Wideband Antennas	89	
Some Simple Cases of Self-Similarity	93	
Weierstrass Functions and a Musical Paradox	96	
More Self-Similarity in Music: The Tempered Scales of Bach	99	
The Excellent Relations between the Primes 3, 5, and 7	102	
<b>4 POWER LAWS: ENDLESS SOURCES OF SELF-SIMILARITY</b>	103	
The Sizes of Cities and Meteorites	103	
A Fifth Force of Attraction	105	
Free of Natural Scales	107	
Bach Composing on All Scales	107	
Birkhoff's Aesthetic Theory	109	
Heisenberg's Hyperbolic Uncertainty Principle	112	
Fractional Exponents	115	
The Peculiar Distribution of the First Digit	116	
The Diameter Exponents of Trees, Rivers, Arteries, and Lungs	117	
<b>5 NOISES: WHITE, PINK, BROWN, AND BLACK</b>	121	
Pink Noise	122	
Self-Similar Trends on the Stock Market	126	
Black Noises and Nile Floods	129	
Warning: World Warming	131	
Fractional Integration: A Modern Tool	131	
Brownian Mountains	133	
Radon Transform and Computer Tomography	134	
Fresh and Tired Mountains	135	
<b>6 BROWNIAN MOTION, GAMBLING LOSSES, AND INTERGALACTIC VOIDS: RANDOM FRACTALS PAR EXCELLENCE</b>	139	
The Brownian Beast Tamed	140	
Brownian Motion as a Fractal	141	
How Many Molecules?	143	
The Spectrum of Brownian Motion	144	
The Gambler's Ruin, Random Walks, and Information Theory	145	
Counterintuition Runs Rampant in Random Runs	146	
More Food for Fair Thought	147	
The St. Petersburg Paradox	148	
Shannon's Outguessing Machine	149	
The Classical Mechanics of Roulette and Shannon's Channel Capacity	150	
<b>7 CANTOR SETS: SELF-SIMILARITY AND ARITHMETIC DUST</b>	161	
A Corner of Cantor's Paradise	161	
Cantor Sets as Invariant Sets	165	
Symbolic Dynamics and Deterministic Chaos	166	
Devil's Staircases and a Pinball Machine	167	
Mode Locking in Swings and Clocks	171	
The Frustrated Manhattan Pedestrian	172	
Arnold Tongues	174	
<b>8 FRACTALS IN HIGHER DIMENSIONS AND A DIGITAL SUNDIAL</b>	177	
Cartesian Products of Cantor Sets	177	
A Leaky Gasket, Soft Sponges, and Swiss Cheeses	178	
A Cantor-Set Sundial	181	
Fat Fractals	183	
<b>9 MULTIFRACTALS: INTIMATELY INTERTWINED FRACTALS</b>	187	
The Distributions of People and Ore	187	
Self-Affine Fractals without Holes	190	
The Multifractal Spectrum: Turbulence and Diffusion-Limited Aggregation	193	
Viscous Fingering	199	
Multifractals on Fractals	200	
Fractal Dimensions from Generalized Entropies	203	
The Relation between the Multifractal Spectrum $f(\alpha)$ and the Mass Exponents $\tau(q)$	205	
Strange Attractors as Multifractals	206	
A Greedy Algorithm for Unfavorable Odds	207	
<b>10 SOME PRACTICAL FRACTALS AND THEIR MEASUREMENT</b>	211	
Dimensions from Box Counting	213	
The Mass Dimension	215	
The Correlation Dimension	220	
Infinitely Many Dimensions	220	
The Determination of Fractal Dimensions from Time Series	223	

<b>Abstract Concrete</b>	<b>224</b>	
Fractal Interfaces Enforce Fractional Frequency Exponents	225	
The Fractal Dimensions of Fracture Surfaces	230	
The Fractal Shapes of Clouds and Rain Areas	231	
Cluster Agglomeration	232	
Diffract from Fractals	233	
<b>11 ITERATION, STRANGE MAPPINGS, AND A BILLION DIGITS FOR PI</b>	<b>237</b>	<b>319</b>
Looking for Zeros and Encountering Chaos	239	
The Strange Sets of Julia	243	
A Multifractal Julia Set	245	
The Beauty of Broken Linear Relationships	249	
The Baker's Transformation and Digital Musical Chairs	251	
Arnold's Cat Map	253	
A Billion Digits for $\pi$	257	
Bushes and Flowers from Iterations	259	
<b>12 A SELF-SIMILAR SEQUENCE, THE LOGISTIC PARABOLA, AND SYMBOLIC DYNAMICS</b>	<b>263</b>	<b>345</b>
Self-Similarity from the Integers	264	
The Logistic Parabola and Period Doubling	268	
Self-Similarity in the Logistic Parabola	272	
The Scaling of the Growth Parameter	274	
Self-Similar Symbolic Dynamics	277	
Periodic Windows Embedded in Chaos	279	
The Parenting of New Orbits	282	
The Calculation of the Growth Parameters for Different Orbits	286	
Tangent Bifurcations, Intermittency, and 1/f Noise	289	
A Case of Complete Chaos	291	
The Mandelbrot Set	295	
The Julia Sets of the Complex Quadratic Map	297	
<b>13 A FORBIDDEN SYMMETRY, FIBONACCI'S RABBITS, AND A NEW STATE OF MATTER</b>	<b>301</b>	<b>357</b>
The Forbidden Fivefold Symmetry	301	
Long-Range Order from Neighborly Interactions	304	
Generation of the Rabbit Sequence from the Fibonacci Number System	307	
The Self-Similar Spectrum of the Rabbit Sequence	308	
Self-Similarity in the Rabbit Sequence	310	
A One-Dimensional Quasiperiodic Lattice	310	
Self-Similarity from Projections	311	
More Forbidden Symmetries	315	
<b>14 PERIODIC AND QUASIPERIODIC STRUCTURES IN SPACE—THE ROUTE TO SPATIAL CHAOS</b>	<b>319</b>	
Periodicity and Quasiperiodicity in Space	320	
The Devil's Staircase for Ising Spins	321	
Quasiperiodic Spatial Distributions	322	
Beatty Sequence Spins	325	
The Scaling Laws for Quasiperiodic Spins	329	
Self-Similar Winding Numbers	330	
Circle Maps and Arnold Tongues	331	
Mediants, Farey Sequences, and the Farey Tree	334	
The Golden-Mean Route to Chaos	340	
<b>15 PERCOLATION: FROM FOREST FIRES TO EPIDEMICS</b>	<b>345</b>	
Critical Conflagration on a Square Lattice	346	
Universality	350	
The Critical Density	353	
The Fractal Perimeters of Percolation	353	
Finite-Size Scaling	354	
<b>16 PHASE TRANSITIONS AND RENORMALIZATION</b>	<b>357</b>	
A First-Order Markov Process	357	
Self-Similar and Non-Self-Similar Markov Processes	358	
The Scaling of Markov Outputs	360	
Renormalization and Hierarchical Lattices	362	
The Percolation Threshold of the Bethe Lattice	363	
A Simple Renormalization	367	
<b>17 CELLULAR AUTOMATA</b>	<b>371</b>	
The Game of Life	373	
Cellular Growth and Decay	375	
Biological Pattern Formation	382	
Self-Similarity from a Cellular Automaton	383	
A Catalytic Converter as a Cellular Automaton	386	
Pascal's Triangle Modulo $N$	387	
Bak's Self-Organized Critical Sandpiles	389	
<b>Appendix</b>	<b>391</b>	
<b>References</b>	<b>395</b>	
<b>Author Index</b>	<b>411</b>	
<b>Subject Index</b>	<b>417</b>	

## Color Plates

**Plate 1** Three heavenly bodies meet. (A) As a double star (yellow and red, approaching from the left) and a single star (blue, closing in from the right) get close to each other, their orbits become wildly chaotic. After some time, they separate again, the double star orbiting off to the upper left and the single star receding to the lower right. (B) In this three-body rendezvous there is a switch of partners. The yellow member of the incoming double star exchanges its red partner for the single blue star, dancing away with its new mate to the lower left. The discarded red partner wanders off alone to the upper right. (Courtesy of Wolf Dieter Brandt.)

**Plate 2** This fanciful self-similar leaf was generated by iterated affine transformations. (Courtesy of Holger Behme.)

**Plate 3** Newton's iteration has three basins of attraction ("countries," shown in red, green, and blue). They meet at a fractal border with the following bizarre property: wherever two countries meet, the third is also present. Paradoxically there are no border *lines*, only three-sided border *posts*. Would international borders so designed promote peace? (Courtesy of Holger Behme.)

**Plate 4** "Rainbow to Infinity," combines a large number of logarithmic spirals in different colors. (Courtesy of Holger Behme.)

**Plate 5** "Clouds over Eastern Europe," a photograph, was taken by the author from his home near the crumbling "wall."

**Plate 6** Real-world fractals. (A) Peeling paint at the Berkeley Heights Swim Club, a multiply connected fractal. (B) Fractal growths in microorganisms: red algae on a rock in Point Reyes National Seashore, California.

**Plate 7** Plants and flowers produced by turtle algorithms. (A) Rose campion, also called dusty miller (*Lychnis coronaria*), raised by Przemyslaw Prusinkiewicz and James Hanan. (B) Vervain (*Verbena*). (C) Lilac twig, grown by Prusinkiewicz and Hanan. (D) "The Garden of L" planted by Prusinkiewicz, Hanan, David Fracchia, and Deborah Fowler. (E) Plant with basipetal flowering sequences. (F) Flower field, fertilized with stochastic L-systems. (Plates 7A–F copyright © 1988 by P. Prusinkiewicz, University of Regina.)

**Plate 8** (A) The Mandelbrot set (black) of complex numbers  $c$  for which the iteration  $z_{n+1} = z_n^2 + c$  with  $z_0 = 0$  stays bounded. Colored areas signify  $c$ -values for which the iteration is unbounded. For  $c$ -values from the main "cardioid" area of the M set, the iteration has a period length equal to 1. The circular disk to the left of the cardioid has a period length of 2, and so on. Each disk or "wart" comprises  $c$ -values for a given finite period length. The M set is a connected set, but it is not known whether it is locally connected. (B) A blow-up in the complex plane reveals a wart and filaments connecting it to smaller replicas of the M set. (C) Another enlargement in the complex plane shows an eleven-armed "whirlpool" in which smaller M sets swim. The arms themselves embrace smaller whorls whose parts, no doubt, contain still smaller whorls. (D) Self-similar descent: zeroing in on one of the "baby" M sets frolicking in the whirlpool reveals its patent parentage, the "mother" M set shown in A. (Courtesy of Holger Behme.)

**Plate 9** Self-organized genetic drift between 16 different "species,"  $n = 1$  to 16 (shown in different rainbow colors). (A) Initially, the different species are randomly intermingled on a square lattice. At every click of the evolutionary clock, each lattice point occupied by species  $n$  will change any of its four nearest neighbors that belong to the species  $n - 1$  to its own species number ( $n = 0$  corresponds to  $n = 16$ ). (B) At a later stage, different genes dominate larger and larger coherent areas, but many fine-grained neighborhoods persist. (C) Still later in the evolutionary process, a new kind of genetic pattern emerges: spirals with periodically repeating species. (D) Although the genetic interactions are strictly local, large spirals are the surviving dominant pattern. As in quasicrystals and numerous other natural phenomena, local rules engender long-range order and global designs. (Courtesy of Holger Behme.)

# PREFACE

*Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection.*

—HERMANN WEYL

The unifying concept underlying fractals, chaos, and power laws is self-similarity. Self-similarity, or invariance against changes in scale or size, is an attribute of many laws of nature and innumerable phenomena in the world around us. Self-similarity is, in fact, one of the decisive symmetries that shape our universe and our efforts to comprehend it.

Symmetry itself is one of the most fundamental and fruitful concepts of human thought [Weyl 81]. By symmetry we mean an *invariance* against change: something stays the same, in spite of some potentially consequential alteration. Mirror symmetry, that is, invariance against “flipping sides,” is perhaps the most widely noticed symmetry. Nature built many of her organisms in nearly symmetric ways, and most fundamental laws of physics, such as Newton’s law of gravitation, have an *exact* mirror symmetry: there is no difference between left and right in the attraction of heavenly (and most earthbound) bodies. However, the nonconservation of parity in radioactive decay—that is, the *violation* of point symmetry in the “weak” interactions—has finally taught even the physicists to take the distinction between right and left seriously.

Another important symmetry is invariance with respect to geometric translation. Our trust in invariance under transpositions in space and time is, in fact, so unlimited that we believe that the laws of nature are the same all over the cosmos—and that they have been, and will remain so, for all time.

An equally momentous symmetry is invariance with respect to rotation. A circle is invariant under rotation around its center by any angle. A square

can be rotated only through angles that are multiples of  $360^\circ/4 = 90^\circ$ ; it is said to have a fourfold symmetry axis. A regular hexagon has a sixfold symmetry. While the rotational symmetry of a flower or a starfish may be imperfect, the exact isotropy found in the fundamental laws of nature is one of the most powerful principles in elucidating the structure of individual atoms, complicated molecules, and entire crystals. Transposition, rotation, and mirror symmetries, acting together, shape crystals from diamonds to snowflakes. And the same three symmetries govern much of what we find pleasing in ornamental designs.

An even more astounding symmetry is the exact identity of like elementary particles. There simply is no difference between an electron here and an electron there—on a distant star, for example. In fact, the perfect identity of photons, the particles of light, has disqualified them from being counted as so many identifiable individuals, resulting in a new kind of particle statistics, discovered by S. N. Bose and rendered palatable by Einstein—a way of counting not heretofore encountered in a world filled with tangible objects.

It was one of the greatest mathematicians of our century, Emmy Noether, who first pointed out the connection between the symmetries of the fundamental laws of physics with respect to displacements in space and time and rotations, on the one hand, and the conservation of linear momentum, energy, and angular momentum on the other. (Noether taught at Göttingen, where David Hilbert, overcoming obstinate prejudice, had finally secured a faculty position for her. In the dismantling of German science in 1933, she was forced to leave Gottingen. She died at Bryn Mawr in 1935.)

Other symmetries have had equally profound consequences in our understanding of the universe we inhabit. Invariance against uniform motion has given us special relativity, a fusion of space and time into space-time and, as its best-known consequence, the equation  $E = mc^2$ . The equivalence of acceleration and gravity postulated by Einstein is the basis of his general theory of relativity, which further revolutionized our appreciation of space, time, and matter.

Yet, among all these symmetries flowering in the Garden of Invariance, there sprouts one that, until recently, has not been sufficiently cherished: the ubiquitous invariance against changes in size, called *self-similarity* or, if more than one scale factor is involved, *self-affinity*. The enormously fruitful concepts of self-similarity and self-affinity pervade nature from the distribution of atoms in matter to that of the galaxies in the universe. And in mathematics, too, self-similarity is deeply entrenched. Some 300 years ago the German philosopher and polymath Gottfried Wilhelm Leibniz used the scaling invariance of the infinitely long straight line for its definition. Cantor sets and Weierstrass functions are other early examples—albeit less smooth—of self-similar structures in mathematics, later joined by Julia sets and other marvels of set theory.

It is perhaps symptomatic that with *set theory* still another abstract branch of mathematics has penetrated the real world. There simply seems to be no limit to Eugene Wigner's “unreasonable effectiveness” of mathematics. Indeed, who would have thought that such utterly mathematical constructions as *Cantor sets*,

invented solely to reassure the skeptics that sets could both have zero measure and still be uncountable, would make a real difference in any practical realm, let alone become a pivotal concept? Yet this is precisely what happened for many natural phenomena from gelation, polymerization, and coagulation in colloidal physics and chemistry to nonlinear systems in innumerable branches of science. Percolation, dendritic growth, fracture surfaces, electrical discharges (lightnings and Lichtenberg figures), and the composition of quasicrystals are best described by set-theoretic constructs.

Or take the weird functions Karl Weierstrass invented a hundred years ago purely to prove that a function could be both everywhere continuous and yet nowhere differentiable. The fact that such an analytic pathology describes something in the real world—nay, is *elemental* to understanding the strange attractors of nonlinear dynamic systems (such as the double swing and the three-body problem)—gives one pause.

The word *symmetric* is of ancient Greek parentage and means well-proportioned, well-ordered—certainly nothing even remotely chaotic. Yet, paradoxically, self-similarity, the topic of this tome, alone among all the symmetries gives birth to its very antithesis: *chaos*, a state of utter confusion and disorder. As we shall endeavor to show, the genesis of chaos is, in fact, closely related to self-similarity and its inherent lack of “smoothness.”

Perhaps not surprisingly, self-similarity entails numerous paradoxes in measurements of time, length, and even musical pitch. Think of Zeno’s tardy turtle, pursued—but never overtaken—by swift Achilles. Why do certain lengths increase without bound when we measure them with ever smaller yardsticks? How would Euclid have explained plane geometric figures whose areas scale not as the squares of their apparent perimeters but as some lesser power, such as 1.77 and other fractional exponents? What should we think of musical sounds that, when scaled *up* in frequency, sound—incredibly—*lower* in pitch? How are such monstrosities possible? And how can we describe them in a consistent, meaningful manner?

Here a particularly felicitous thought by Felix Hausdorff comes to the rescue. His and Abram Besicovitch’s new ways of looking at dimension dethrones it from its integer position and propels it into the realm of real numbers, giving us one of the sharpest tools—the *Hausdorff dimension* and its ramifications—with which to attack the strange sets that self-similarity breeds.

And while recalling some of the glorious names of the past, we should never forget our great contemporary, the inimitable Benoit B. Mandelbrot, who, single-handed, rescued set theory’s most brittle functions and “dustiest” sets from near-oblivion and planted them right in the middle of our daily experience and consciousness. Yes, for all these years, we *have* been living with fractal arteries, not far from fractal river systems draining fractal mountain-scapes under fractal clouds, toward fractal coastlines. But, kin to Molière’s would-be gentleman, we lacked the proper prose—*fractal*, noun and adjective—that Benoit B. begot.

But our story also has a silent and immobile hero: the digital computer. There can be little doubt that computers have acted as the most forceful forceps in extracting fractals from the dark recesses of abstract mathematics and delivering their geometric intricacies into bright daylight. In fact, the impact of fractal images, often of unimagined beauty and appeal, has given computer graphics a surprising new dimension.

## Synopsis

We open our treatise with one of the most charming uses that similarity was ever put too: the young Einstein's proof of Pythagoras's theorem. By adding just a single straight line, in the right place, to a right triangle and applying plenty of similarity, the popular theorem is proved without further prodding.

We then invade the unlimited domain of *self-similarity* as manifest in fractals, multifractals, and the scaling laws of physics, psychophysics, and boundless other fields.

In phase spaces, we encounter deterministic chaos and strange attractors. Percolation and other phase transitions lead us to critical exponents and a hierarchy of different dimensions. Following Poincaré, we immerse ourselves in the self-similarities of iterated mappings, from baker transformations and Bernoulli shifts to logistic parabolas and circle maps. Neither tori, *cantori*, nor Arnold tongues will faze us as we (sur)mount devil's staircases to unwind among the rational winding numbers festooning Farey trees.

And when we talk about nonlinear dynamics we must remember some of the great contributors of recent vintage: Siegel, Moser, Lorenz, Wilson, Feigenbaum, and—last but not least—the great Russian “school” exemplified by such names as Lyapunov, Arnol'd, Sinai, Chirikov, Alexeev, Anosov, Pesin, and the recently deceased master mathematician Kolmogorov.

Cayley trees, also known as Bethe lattices, will provide us with a fitting point of departure for many a practical fractal, such as our bronchial and vascular systems. Cellular automata concern us as models of both biological growth and chemical reactions.

We are also strangely attracted to symbolic dynamics, kneading (and needing) the Morse-Thue sequence and, especially, the Fibonacci rabbit sequence and their discrete self-similarities that, indiscreetly, tell us so much about period doubling, mode locking, frustrated Ising spins, and fivefold symmetric quasicrystals. Many of these subjects were shrouded in mystery and beset by paradoxes before the sharp scalpels, fashioned by scaling and renormalization theories, revealed the underlying tissue and made them tractable. In fact, it is no accident that viable fundamental field theories in physics are renormalizable, as they must be if they are to shun sham scales.

And, of course, we will not hesitate to run down random fractals, from Brownian motion to diffusion-limited aggregation and stock market hiccups

(some hiccups of late!). The poor gambler's ruin and the St. Petersburg paradox will provide further food for fractal reflections.

These, then, are some of the exciting, and sobering, themes sounded in the present volume. The aim of this exposition is to enhance the reader's understanding of self-similarity, perhaps the most pregnant of all of nature's symmetries, and to illustrate the wide-ranging applications of scaling invariance in physics, chemistry, biology, music, and—particularly—the visual arts, as manifested in the recent renaissance of computer graphics through fractal images and their iterative beauty.

## Acknowledgments

This book owes its existence to many sources. Apart from a brief encounter, in my dissertation, with chaos among the normal modes of concert halls, a "nonintegrable" system if there ever was one, my main stimulus came from the early demonstrations by Heinz-Otto Peitgen and Peter Richter of fractal Julia sets. Their beautiful images, and the intriguing mathematics which underlies them, as epitomized in their book *The Beauty of Fractals*, have made a lasting impression on me.

My first meeting with Mandelbrot's work was his analysis of word frequencies in natural and artificial languages, which touched upon my own interests in computer speech synthesis and recognition. Mandelbrot's monumental monograph *The Fractal Geometry of Nature* influenced me immeasurably, as it did so many other people.

I have also greatly benefited from Robert Devaney's books and lectures on chaos, Jens Feder's *Fractals*, Michael Barnsley's *Fractals Everywhere*, and Dietrich Stauffer's charming *Introduction to Percolation Theory*.

I learned a lot about new developments during the 1988 Gordon Conference on Fractals, organized by Richard Voss and Paul Meakin, which brought together many of the world's outstanding experts in the field.

At the University of Göttingen, it was mainly through the work of Werner Lauterborn that I saw chaos in action in nonlinear dynamic systems from cavitation bubbles to Toda chains. My own students at the Drittes Physikalisches Institut provided both stimulus and rectification during a series of lecture courses on self-similarity, fractals, and chaos. Heinrich Henze and Karl Lautscham made skillful additions to the demonstrations that accompanied these lectures. Holger Behme, Wolf Dieter Brandt, and Tino Gramss contributed many of the computer graphics in this book.

The late Walter Kaufmann-Bühler of Springer Verlag, New York, and Ronald Graham, a longtime colleague at Bell Laboratories, provided early encouragement.

Hildegard Franks and Esperanza Plata in Murray Hill, New Jersey, supported by Maija Sutter and Irena Schönke in Göttingen, converted nearly illegible

handwriting to readable typescript. Liane Liebe and Gisela Kirschmann-Schröder took loving care of the illustrations. Anny Schroeder, as during a previous itinerary, furnished much needed logistic support.

Jerry Lyons persuaded me to deliver myself and the manuscript into the hands of W. H. Freeman and Company, and I have not regretted heeding his appeal one bit. I greatly enjoyed working with him; with Georgia Lee Hadler, senior editor; with designer Nancy Singer; and with illustration coordinator Bill Page. The professional competence of my copy editor, Richard Mickey, proved beyond compare.

Last, but not least, I thank Hans-Werner Strube, Allan Hurd, Victor Klee, and Paul Meakin for their expert reviews of the manuscript and their manifold corrections and cogent comments.

*Murray Hill and Göttingen  
May 1990*

*Manfred Schroeder*

# I

## ntroduction

*I want to know how God created this world. I am not interested in this or that phenomenon, in the spectrum of this or that element. I want to know His thoughts; the rest are details.*

—ALBERT EINSTEIN

Nature abounds with periodic phenomena: from the motion of a swing to the oscillations of atoms, from the chirping of a grasshopper to the orbits of the heavenly bodies. And our terrestrial bodies, too, participate in this universal minuet—from the heart beat and circadian rhythms to monthly and even longer cycles.

Of course, nothing in nature is *exactly* periodic. All motion has a beginning and an end, so that, in the mathematical sense, strict periodicity does not exist in the real world. Nevertheless, periodicity has proved to be a supremely useful concept in elucidating underlying laws and mechanisms in many fields.

One reason for the universality of simple harmonic motion is the linearity—or near-linearity—of many physical systems and the invariance with displacement in space and time of the laws governing their behavior.

But there are numerous other phenomena in which linearity breaks down and, instead of periodicity, we get aperiodic or even chaotic motion: the smooth waves on a well-behaved lake turn to violent turbulence in the mountain brook, and the daily sunrise, the paradigm of predictability, is overshadowed by cloud formations, a haven for *chaos*—albeit *deterministic* chaos.

But no matter how chaotic life gets, with all regularity gone to bits, another fundamental bulwark often remains unshaken, rising above the turbulent chaos: *self-similarity*, an invariance with respect to *scaling*; in other words, invariance not

with *additive* translations, but invariance with *multiplicative* changes of scale. In short, a self-similar object appears unchanged after increasing or shrinking its size. Indeed, in turbulent flows, large eddies beget smaller ones, and these spawn smaller ones still—and so on *ad infinitum* (almost). In general, one of the conspicuous consequences of self-similarity is the appearance of exceedingly fine-grained structures, now generally called *fractals* after Benoit B. Mandelbrot, the father of fractals [Man 83].<sup>1</sup>

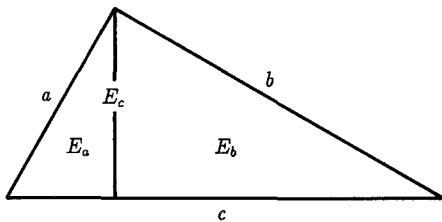
Many laws of nature are independent, or nearly so, of a scaling factor. The fact that scaling usually has a limit (Planck's constant, when things get too small, or the speed of light, when objects fly too fast) does no harm to the usefulness of "thinking self-similar," just as the lack (outside mathematics) of strict periodicity is no great impediment in the real world. In a sense, self-similarity is akin to periodicity on a *logarithmic* scale.

Self-similarity, strict or otherwise, reigns in many fields in many guises, and in this book we shall explore some of the many manifestations of self-similarity in the world around us. Among the topics treated are the following:

- Scaling laws and their exponents in physics, psychophysics, and physiology
- Random walks in the stock market and under the microscope; floods, forest fires, the distribution of galaxies, and other "accidents" with statistical self-similarity
- Scaling invariance, self-similarity, and some of their mathematical models, such as Cantor sets and Julia sets
- Fractals and their characterization by Hausdorff, and other noninteger dimensions; fractal paradoxes and their resolution; Weierstrass functions and Hilbert curves; Koch flakes, Sierpinski gaskets, and other non-Euclidean constructions in two and more dimensions; fat fractals and multifractals
- Iterated mappings and a selection of the ensuing self-similarities
- The logistic parabola and other unimodal maps with universal scaling laws; period-doubling bifurcation to chaos; the Feigenbaum constant, symbolic dynamics and the Morse-Thue sequence; Sharkovskii's universal ordering of orbits
- Complexification of the quadratic map and the Mandelbrot set
- Devil's staircases, Farey trees, Arnold tongues, and modelocking; the "rabbit sequence" and the quasi-periodic route to temporal and spatial chaos; Ising spins and quasicrystals
- Laplace's triangle and cellular automata

---

1. References in brackets are listed alphabetically at the end of the book. The numbers refer to the year of publication.



**Figure 1** Pythagoras's theorem: sketch for proof by the 11-year-old Einstein based on similarity.

In this chapter some of these topics are introduced informally, together with the leading dramatis personae.

## Einstein, Pythagoras, and Simple Similarity

*I will a little think.*

—ALBERT EINSTEIN, in America

When Jacob Einstein taught (Euclidean) geometry to his 11-year-old nephew Albert, the young Einstein—even then striving for utmost parsimony—felt that some of Euclid's proofs were unnecessarily complicated.<sup>2</sup> For example, in a typical proof of Pythagoras's theorem  $a^2 + b^2 = c^2$ , was it really mandatory to have all those extra lines, angles, and squares in addition to the basic right triangle with hypotenuse,  $c$  and sides  $a$  and  $b$ ?

After “a little thinking,” the sharp youngster came up with a proof that required only *one* additional line, the altitude above the hypotenuse (see Figure 1). This height divides the large triangle into two smaller triangles that are *similar* to each other and similar to the large triangle. (Triangles are similar if their angles are the same, which is easily seen to be the case in Figure 1.)

Now, in Euclidean geometry, the area ratio of two similar (closed) figures is equal to the *square* of the ratio of corresponding *linear* dimensions. Thus, the areas  $E_a$ ,  $E_b$ , and  $E_c$  ( $E$  as in German *Ebene*) of the three triangles in Figure 1 are related to their hypotenuses  $a$ ,  $b$ , and  $c$  by the following equations:

$$E_a = m a^2 \quad (1)$$

---

2. I have the story from Schneior Lifson of the Weizmann Institute in Tel Aviv, who has it from Einstein's assistant Ernst Strauss, to whom it was told by old Albert himself.

$$E_b = mb^2 \quad (2)$$

$$E_c = mc^2 \quad (3)$$

where  $m$  is a dimensionless nonzero multiplier that is the *same in all three equations*.

Now a second look at Figure 1 will reveal that the area of the large triangle is, of course, the sum of the areas of the two smaller triangles,

$$E_a + E_b = E_c$$

or, with equations 1 to 3,

$$ma^2 + mb^2 = mc^2$$

Dividing this identity by the common measure  $m$  promptly produces Pythagoras's renowned result

$$a^2 + b^2 = c^2$$

proved here by an 11-year old person<sup>3</sup> by combining two fertile scientific principles that were going to stand the grown-up Einstein in good stead: simplicity and symmetry, of which self-similarity is a special case. Yet the true beauty of Einstein's proof is not that it is so simple, but that it exposes the true essence of Pythagoras's theorem: *similarity and scaling*.

The resemblance of equation 3 to Einstein's later discovery, his famous  $E = mc^2$ , is of course entirely fortuitous. The equivalence of mass  $m$  and energy  $E$ , which is at the basis of nuclear power in all its guises, is a consequence of Lorentz invariance. This invariance, which underlies special relativity, was predicted by Einstein in 1905 after, it seems, several false starts and a "little more thinking" (see Figure 2).

## A Self-Similar Array of Self-Preserving Queens

One of numerous chess problems is the placement of as many queens as possible on a chessboard of a given size so that no queen "attacks" (shares a row, column, or  $\pm 45^\circ$  diagonal with) any other queen. For a  $k \times k$  square board, there can be at most  $k$  nonattacking queens. But are  $k$  peacefully coexisting queens always possible? What if  $k$  is very large? Doesn't the complexity of placing the queens grow exponentially with the size of the board? As we shall see, the placement is actually very simple, even for arbitrarily large boards, if we focus our attention

---

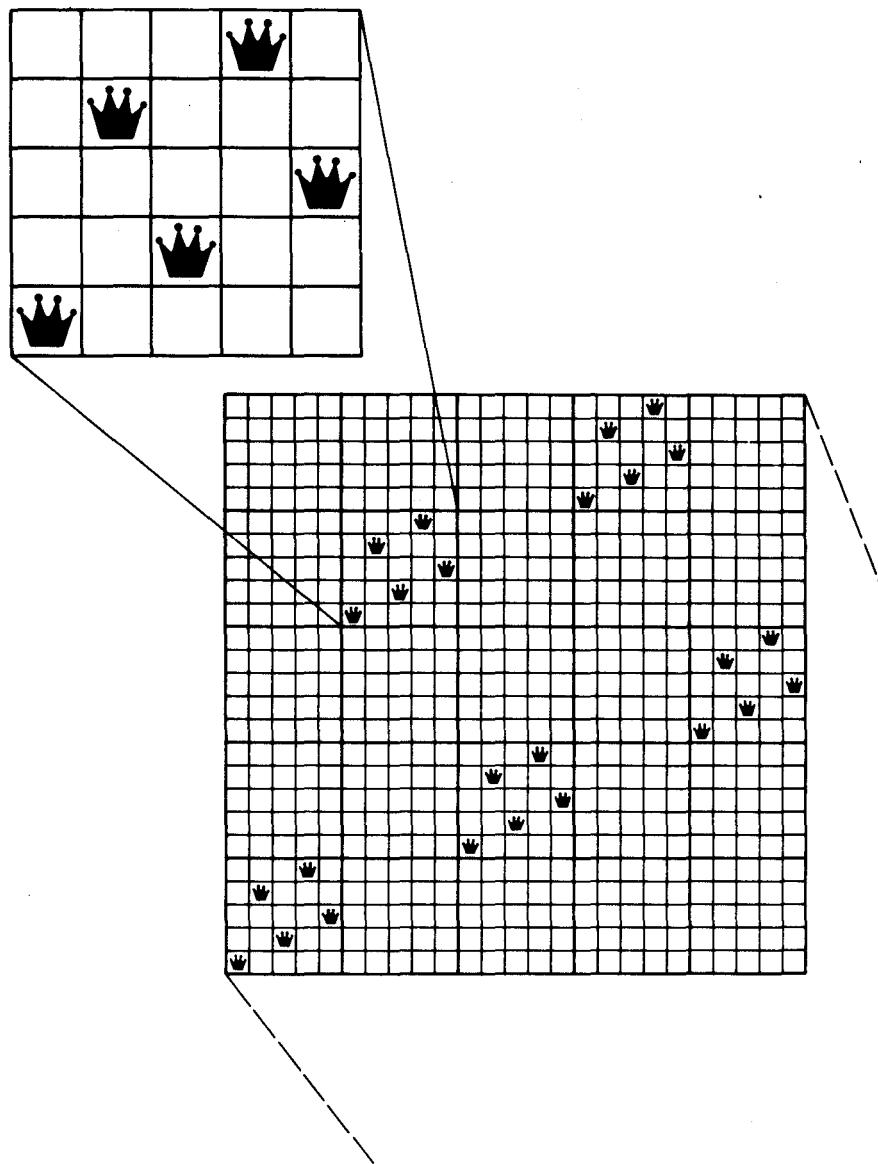
3. Really a "nonperson" at that stage, considering the neglect he suffered in his Munich high school [Pye 85].



**Figure 2** Einstein on the verge of discovering his famous formula  $E = mc^2$ —a cartoonist's view [Har 77]. (© 1991 by Sidney Harris)

on boards for which  $k$  is a pure power of an integer and judiciously exploit the principle of self-similarity in the construction of the solution. (Again, we describe an object or a structure as self-similar if it looks the same when we magnify the object or a properly chosen part of it.)

Figure 3 shows a pattern of queens, the  $5 \times 5$  board sustaining five non-attacking pieces. (This particular placement could have been obtained by a *greedy algorithm*: Starting on the lower left and proceeding column by column, always place the next queen in the lowest position still “eligible.”)



**Figure 3** Five nonattacking chess queens on a  $5 \times 5$  board (top) and solution for the  $25 \times 25$  board derived from the  $5 \times 5$  board by similarity.

From a solution on the  $5 \times 5$  board, we can immediately construct a possible placement for the  $25 \times 25$  board, which can be considered to be composed of  $5 \times 5 = 25$  boards of size  $5 \times 5$ . We simply leave most of those twenty-five  $5 \times 5$  boards empty, except those five that correspond to the positions of the queens in the original board. Figure 3 illustrates the procedure without the need for more words.

To fill a  $125 \times 125$  board with peaceful queens, simply think of it as twenty-five boards of size  $25 \times 25$ , five of which are filled in the by now familiar pattern with the  $25 \times 25$  solutions while the remaining twenty boards are left void. Continuing in this manner, we have, after  $n$  steps, a  $5^n \times 5^n$  board with  $5^n$  pieces.

This process can be extended *ad infinitum* to yield an immaculately *self-similar* distribution of self-preserving queens. Indeed, selecting one of the five occupied subboards of side length one-fifth of the entire board and magnifying it by a factor of 5 will precisely reproduce the entire board. The factor 5 is called the *scaling* or *similarity factor* of the board.

What numbers other than 5 can be used as scaling factors in such self-similar schemes? Can we exploit self-similarity for the construction of boards whose side is not a pure power of an integer (as  $5^n$  is)? The interested reader can find further clues in the illuminating article by Clark and Shisha [CS 88].

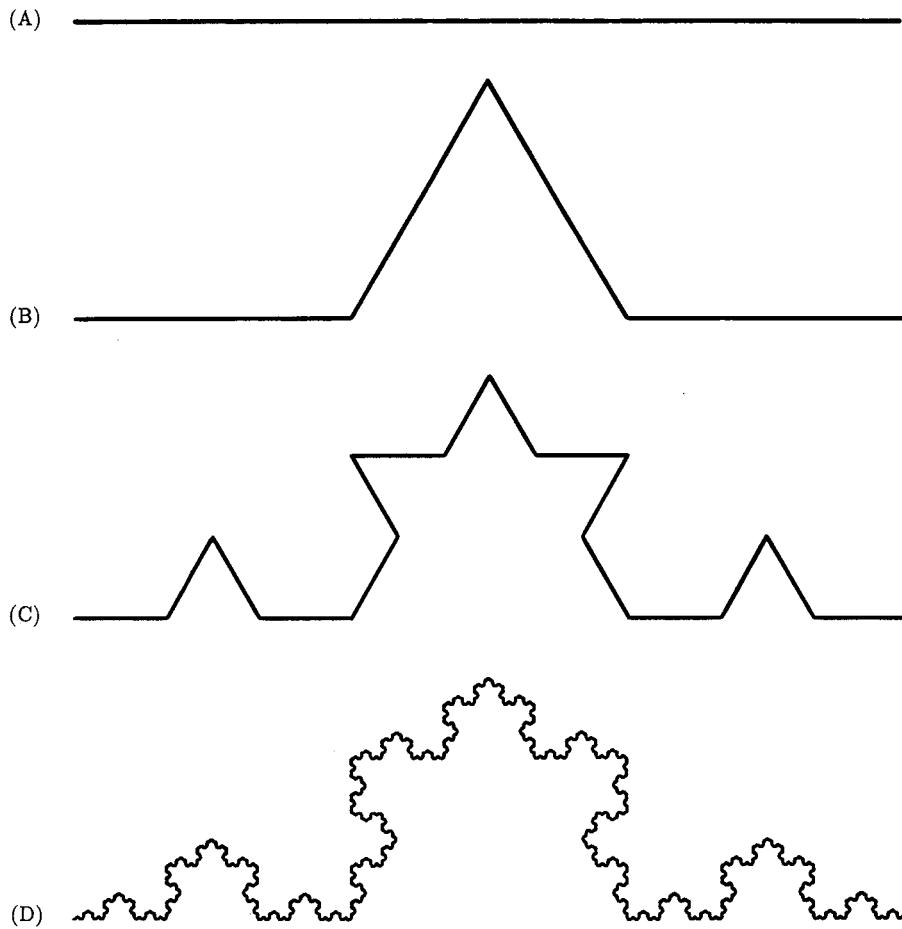
## A Self-Similar Snowflake

Repeating a given operation over and over again—on ever smaller scales—culminates, almost inescapably, in a self-similar structure. Here the repetitive operation can be algebraic, symbolic, or geometric, as in the case of the five dormant queens whom we have just allowed to come alive and multiply without limit, proceeding on the path to perfect self-similarity.

The classical example of such a repetitive construction is the *Koch curve*, proposed in 1904 by the Swedish mathematician Helge von Koch. The basic principle and the final result are equally charming: Take a segment of straight line (Figure 4A, the *initiator*) and raise an equilateral triangle over its middle third as shown in Figure 4B. The result is called the *generator*. Note that the length of the generator is four-thirds the length of the initiator.

Repeating once more the process of erecting equilateral triangles over the middle thirds of straight line segments results in Figure 4C. The length of the fractured line is now  $(\frac{4}{3})^2$ . Iterating the process infinitely many times results in a “curve” of infinite length, which—although everywhere continuous—is *nowhere differentiable*. It is approximated, as far as pen and ink permit, in Figure 4D.

Similarly lamentable “functions,” continuous but without tangents, were first defined a century ago by the German mathematician Karl Weierstrass, just to show his skeptical colleagues (a horrified Hermite among them) that such functions did indeed exist. But other authorities, not least the great Austrian



**Figure 4** Initiator (A) and generator (B) for the Koch curve, the next stage in the construction of the Koch curve (C), and high-order approximation to the Koch curve (D).

physicist Ludwig Boltzmann, saw the light: Boltzmann wrote to Felix Klein (in 1898) that nondifferentiable functions could have been invented by physicists because there are problems in statistical mechanics “that absolutely necessitate the use of nondifferentiable functions.” And his French colleague Jean Perrin went even further when, in 1906, he presaged present sentiment about such mathematical monsters, saying that “curves that have no tangents are the rule, and regular curves, such as the circle, are interesting but quite special.” How politely put! Now, following Mandelbrot, we simply call such nondifferentiable curves *fractals*.

## A New Dimension for Fractals

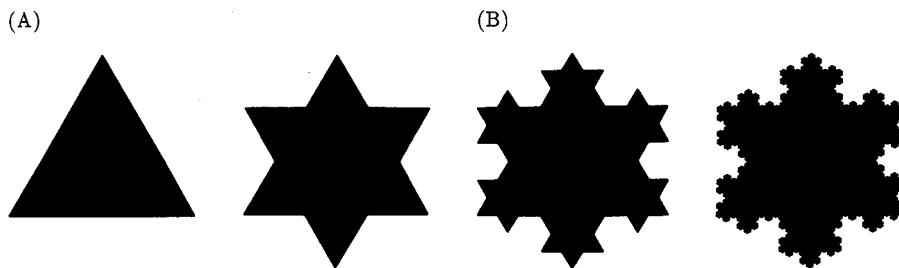
*The universe is not only queerer than we suppose but queerer than we can suppose.*

—J. B. S. HALDANE

Applying the Koch generator (see Figure 4) to an equilateral triangle and painting the interior black results in a solid star of David (Figure 5A). Infinite iteration converges on the Koch snowflake (intermediate stages of the construction are shown in Figure 5B). How long is its perimeter? After  $n$  iterations it has increased  $(\frac{4}{3})^n$ -fold over the perimeter of the initial triangle. Thus, as  $n$  approaches infinity, the perimeter becomes infinitely long. To characterize the perimeter's size, we can therefore no longer use its length. We have to invent a new measure that can distinguish between fractals manufactured from different generators. But while inventing new measures, we want to stay as close as possible to what we have always done when measuring lengths.

For a smooth curve, an approximate length  $L(r)$  is given by the product of the number  $N$  of straight-line segments of length  $r$  needed to step along the curve from one end to the other and the length  $r$ :  $L(r) = N \cdot r$ . As the step size  $r$  goes to zero,  $L(r)$  approaches a finite limit, the length  $L$  of the curve.

Not so for fractals! The product  $N \cdot r$  diverges to infinity because, as  $r$  goes to zero, we enter finer and finer wiggles of the fractal. However, asymptotically, this divergence behaves according to a well-defined homogeneous power law of  $r$ . In other words, there is some *critical exponent*  $D_H > 1$  such that the product  $N \cdot r^{D_H}$  stays finite. For exponents smaller than  $D_H$ , the product diverges to infinity, while for larger exponents the product will tend to zero. This critical exponent,  $D_H$ , is called the *Hausdorff dimension* after the German mathematician



**Figure 5** Initiator and generator for the Koch flake (A) and intermediate stages in the construction of the Koch flake (B).

Felix Hausdorff (1868–1942). Equivalently, we have

$$D_H := \lim_{r \rightarrow 0} \frac{\log N}{\log (1/r)}$$

For the  $n$ th generation in the construction of the Koch curve or snowflake, choosing  $r = r_0/3^n$ , the number of pieces  $N$  is proportional to  $4^n$ . Thus,

$$D_H = \frac{\log 4}{\log 3} = 1.26 \dots$$

The fact that  $D_H$  lies between 1 and 2 is somehow satisfying, because an infinitely long curve is, in some metric sense, more than just a one-dimensional object—without being a two-dimensional area, since the curve does not cover a region in the plane. In fact, we shall soon see that Hausdorff's definition of dimension, which, as we now know, can take on fractional values, makes much sense in many ways. Of course, for a *smooth* curve,  $D_H = 1$ ; and for a smooth surface the number  $N$  of covering disks is proportional to  $1/r^2$  and therefore  $D_H = 2$ . Here  $r$  is the diameter of the  $N$  little disks needed to cover the area. Similarly, for a compact three-dimensional volume,  $D_H$  comes out equal to 3.

Surprisingly, however, for  $D_H$  to equal 2, we do not need an area; a topologically *one*-dimensional entity, a line, suffices. A well-known example is the asymptotically self-similar Hilbert curve (see Figure 6A), which comes arbitrarily close to each point in the unit square. Its construction is illustrated in Figure 6B. The final result is, of course, self-similar. Blow up any appropriately chosen subsquare by a linear factor  $2^n$  and it will resemble the entire figure.

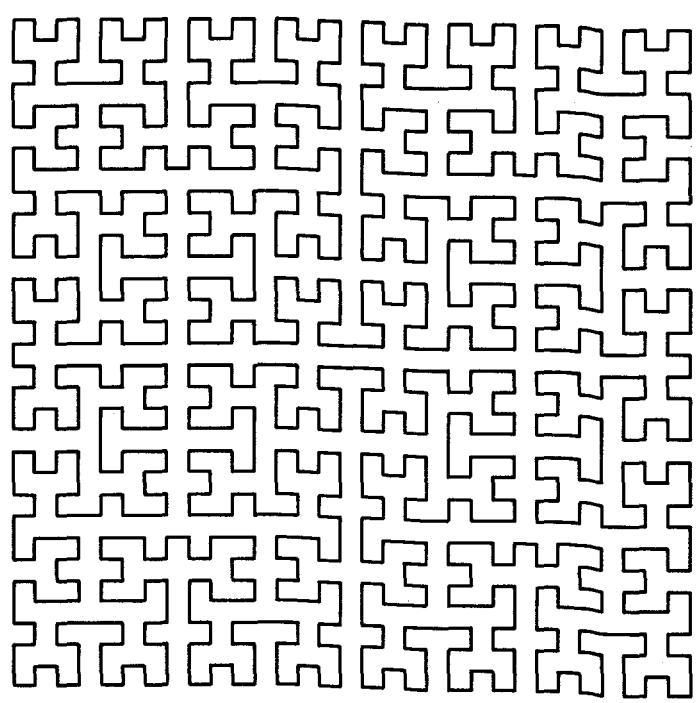
Since the  $n$ th generation of the Hilbert curve consists of  $2^{2n} - 1$  segments of length  $1/2^n$ , its Hausdorff dimension equals 2, as behoves an area-filling curve. Figure 7 shows an artistic variation on the Hilbert curve theme. Can you recognize that the underlying image is a human face?

Adjacent points on the Hilbert curve are adjacent in the unit square, but not vice versa! This property distinguishes the Hilbert curve from broadcast TV scans, which are discontinuous at the line ends,<sup>4</sup> and from Cantor's totally discontinuous mapping of the unit square onto the unit interval, whereby the point in the square  $x = 0.x_1, x_2, x_3, \dots ; y = 0.y_1, y_2, y_3, \dots$  is mapped to the point on the line  $0.x_1, y_1, x_2, y_2, x_3, y_3, \dots$

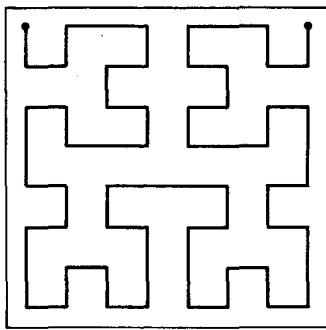
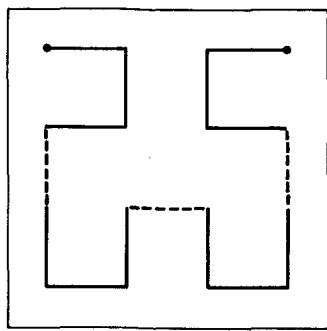
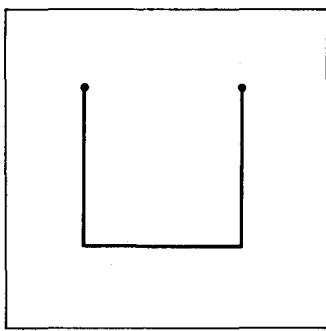
When Cantor first saw that, in this manner, an area could be reversibly mapped to a line, he wrote "I see it, but I don't believe it." But evolution, in constructing our brain, discovered millenia ago that in order to fill a volume

---

4. Interestingly, some sophisticated image-scanning techniques do follow Hilbert's prescription for a space-filling curve. The reason is that points adjacent in time along a "Hilbert scan" are also adjacent in space in the scanned image, making for simpler image processing.

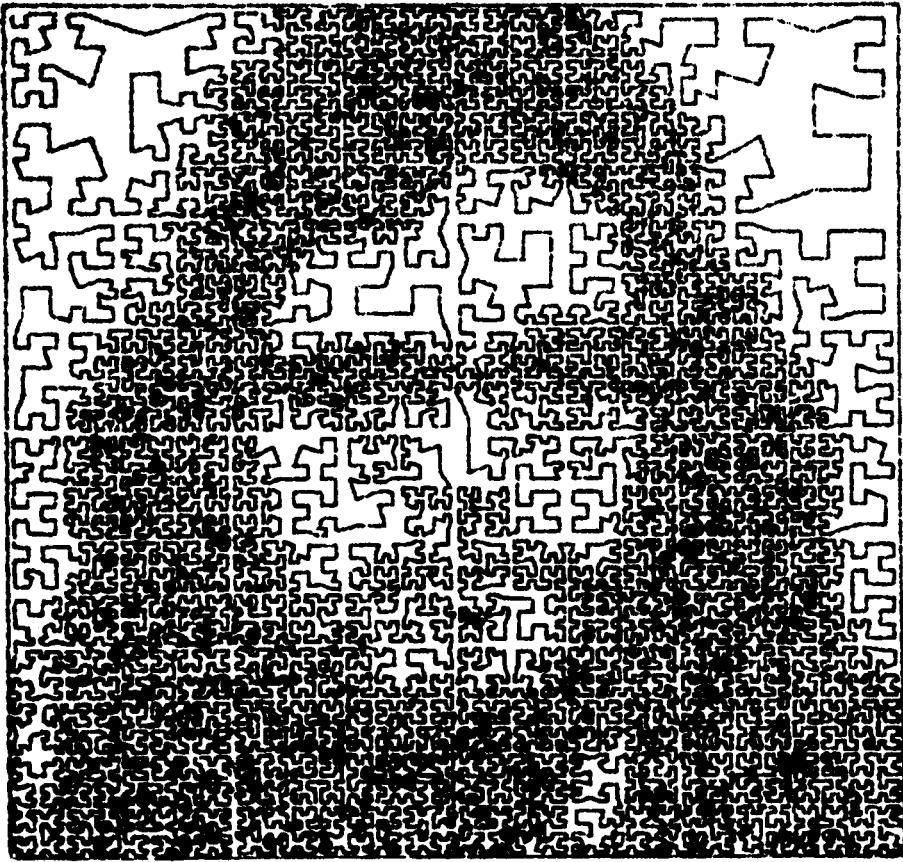


(A)



(B)

**Figure 6** (A) Toward the self-similar Hilbert curve; (B) steps in construction of the Hilbert curve.

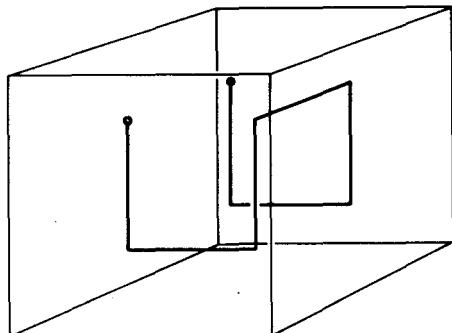


**Figure 7** Contorted Hilbert curve: an artist's version, to be viewed from a distance.  
(Courtesy of H. W. Strube.)

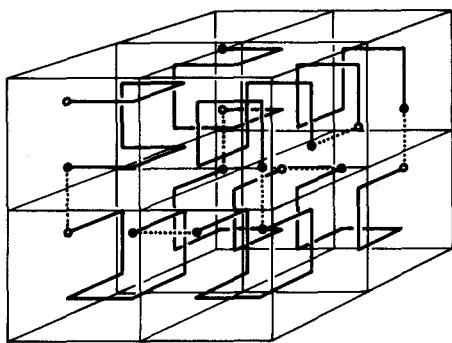
while preserving two-dimensional adjacency, it had to construct the gray matter of our cortex in a folded manner resembling a three-dimensional Hilbert curve.

Hilbert curves in higher-dimensional spaces have also found interesting applications in information theory: the so-called *Gray codes* [Gil 58], so named after their inventor. In a binary Gray code for the integers, only a single bit of the code changes between one integer and the next. Thus, the four integers 0 to 3 are encoded by two binary bits as follows: 0 = 00, 1 = 01, 2 = 11 and 3 = 10 (and not as in the standard binary code, where 2 = 10 and 3 = 11, creating a two-bit jump between the codes for 1 and 2). Figure 8 shows successive stages for the construction of a Hilbert curve in three-space, visualizing generalized Gray codes [Gil 84].

While Cantor's mapping of an area to a line is discontinuous in both directions and the Hilbert curve is continuous in only one direction, there are



(A)



(B)

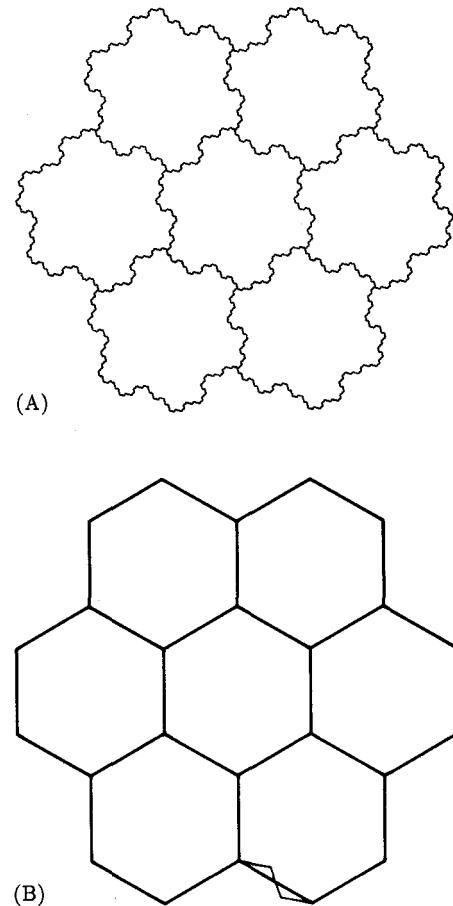
**Figure 8** Constructing a three-dimensional version of the Hilbert curve (A); Hilbert curve for illustrating Gray code (B).

mappings from an area to a line, due to Bernhard Bolzano (1781–1848) and Giuseppe Peano (1858–1932), that are, incredibly, continuous in *both* directions.

### A Self-Similar Tiling and a “Non-Euclidean” Paradox

Look at the seven fractal “tiles” shown in Figure 9A. They are obtained from seven hexagons (see Figure 9B), by breaking up each side into a three-piece zig-zag as shown on one of the sides. If the inner angles of the three pieces are  $120^\circ$ , then the lengths of the three segments will be  $1/\sqrt{7}$  times the length of the unbroken side.

Iterating the breaking up process *ad infinitum* results in a fractal tiling pattern, of which Figure 9A is an approximation. As a result of this construction, the



**Figure 9** (A) Fractal tiles that fit together to cover the plane. The seven tiles shown, taken together, are similar to a single tile, giving rise to a “non-Euclidean” paradox. (B) Tiling hexagons: the initiator of the fractal tiles shown in part A. One generator, consisting of three straight-line segments, is also shown.

perimeter of the entire figure, consisting of seven fractal “hexagons,” is similar to each of the seven hexagons.

Thus, we have found a *self-similar* tiling of the plane based on “hexagons,” where each tile is surrounded by six like tiles. (Note that while regular hexagons do tile the plane, the tiling is not self-similar. A hexagon surrounded by six like hexagons is *not* a larger hexagon.)

Simple inspection of Figure 9A shows that the perimeter of the large fractal hexagon contains the perimeter of the small fractal hexagons precisely three times. Thus, following Euclid’s scaling rule for the areas of similar geometric

figures, the total area should be  $3^2 = 9$  times the area of one of the small fractal hexagons. But it isn't! The area ratio is only 7.

What went wrong? Where did we go astray? Has Euclid finally been caught napping? Well, the ancient Greeks (with the possible exception of one of the Zenos) can continue to rest in peace. Fractal geometric objects like the one illustrated in Figure 9A were never on exhibit in Euclid's school (nor were they used to tile its floors). Euclid probably never considered nondifferentiable functions or bounded curves of infinite length. But later generations of mathematicians did, and since Hausdorff we know that the dimensions of such curves are not necessarily equal to 1 but perhaps exceed 1. For example, the Hausdorff dimension  $D_H$  of the perimeter of our fractal hexagons is  $\log 3/\log \sqrt{7} = 1.12915 \dots$ . Thus, in adapting Euclid's scaling idea, we should raise 3 (the perimeter "ratio") not to the power 2 to obtain the area ratio, but—since the perimeter already has dimension  $1.12915 \dots$ —to the power  $2/1.12915 \dots = 1.77124 \dots$ . This gives an area ratio of 6.999999999 on my pocket calculator—close enough to the true area ratio of 7 to 1 that is immediately apparent in Figure 9A. Thus, we can reformulate Euclid's scaling theorem about similar areas and obtain a more generally valid result, applicable to fractals and nonfractals alike:

*For similar figures, the ratios of corresponding measures are equal when reduced to the same dimension on the basis of their Hausdorff dimensions.*

It is because of properties like this that Hausdorff dimension is such a useful concept. It is *one* proper extension of the concept of dimension to fractal objects, which model, however approximately, a great many phenomena in the real world surrounding us—and *in* us. Just think of the human vascular system, or your lungs with their hierarchical branchings, leading to astonishingly large surface areas that are well described by fractal geometries and Hausdorff dimensions.

Thus, the idea of the Hausdorff dimension has resolved a potentially disastrous paradox by widening our concept of dimension to include fractional and even transcendental values. We shall resume this theme in the body of this book and get to know other fractal paradoxes such as that of a musical chord that sounds *lower* in pitch when reproduced at a higher tape speed! (See pages 96–98 in Chapter 3.)

## At the Gates of Cantor's Paradise

*I place myself in a certain opposition to widespread views on the nature of the mathematical infinite.*

—GEORGE CANTOR

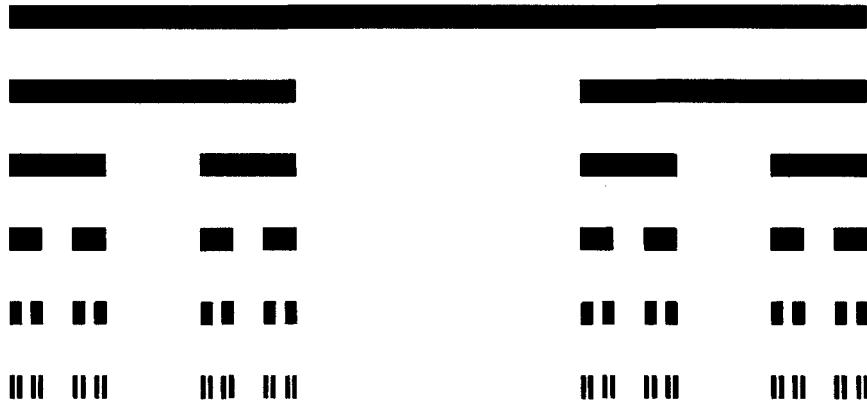
The Hausdorff dimension  $D_H$  is useful not only for characterizing fractal curves of infinite length but also point sets, or "curves" of zero length. Not surprisingly,

for such point sets  $D_H$  is typically less than 1. A famous example is Georg Cantor's original self-similar "middle-third-erasing" set with which he demonstrated, to the astonishment and disbelief of the contemporary mathematical community, that there are sets having measure ("length") zero with *uncountably* many members.

Cantor constructed his highly counterintuitive set as follows. He started with the closed unit interval  $[0, 1]$ , that is, a straight-line segment of length 1 including the two endpoints. He then "wiped away" the open middle third  $(\frac{1}{3}, \frac{2}{3})$  and repeated the process on the remaining two segments of length  $\frac{1}{3}$  (see Figure 10).

Repeating the middle-third wiping-out process over and over again leaves not a single connected line segment; the total length or measure of the remaining set is zero. Yet, as we shall later see, the leftover "dust" still contains infinitely many, in fact *uncountably* many, "points." In fact, one can already appreciate this from the arithmetic description of the Cantor set: its members are precisely all those fractions in the interval  $[0, 1]$  that eschew the digit 1, such as 0.2 or 0.2022.

How do we characterize the content of a set whose length measure is zero? Again Hausdorff offers help. After  $n$  wiping stages, we are left with  $N = 2^n$  straight-line pieces, each of length  $r = (\frac{1}{3})^n$ . Thus, the Hausdorff dimension  $D_H$  equals  $\log 2 / \log 3 = 0.63 \dots$ , a value between 0 and 1, as expected because the Cantor dust is more (a lot more!) than just a point (dimension 0) and much less than a length of line or curve (dimension 1). As in the case of the fractal Koch curve, the value of  $D_H$  is not an integer; in fact, it is a transcendental number.



**Figure 10** Construction of the "middle-third-erasing" Cantor set. It has zero measure, yet is uncountable. Its fractal dimension equals  $\log 2 / \log 3 = 0.63 \dots$

In the course of this book we shall encounter “innumerable” other instances of dusty, Cantor-like sets in a wide variety of settings—such as the celebrated Sierpinski gasket, which we will introduce next.

## The Sierpinski Gasket

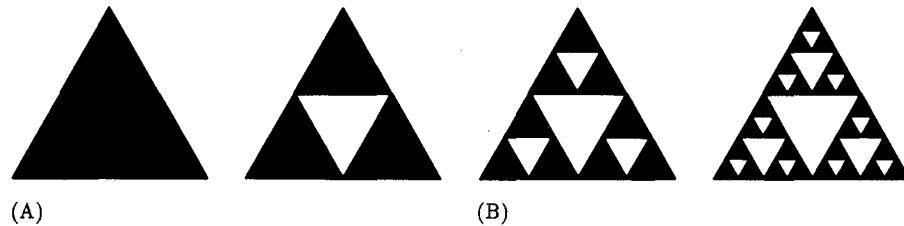
*Fog on Fog.*

—HERMANN WEYL, commenting on  
Cantor’s transfinite numbers

Are there Cantor-like dusts spread out in two dimensions? Yes, there are. Start with the equilateral triangle shown in Figure 11A and remove the open central upside-down equilateral triangle with half the side length of the starting triangle. This leaves three half-size triangles. Repeating the process on the remaining (right-side-up) triangles leaves, after  $n$  iterations,  $N = 3^n$  triangles of side length  $r = r_0(2^{-n})$  (see Figure 11B). The Hausdorff dimension  $D_H$  for the set resulting from an infinite iteration of this procedure, called the *Sierpinski gasket* after the prolific Polish mathematician Waclaw Sierpiński (1882–1969), equals  $\log 3 / \log 2 = 1.58 \dots$ , an irrational number smaller than 2, in spite of the fact that the gasket is embedded in two dimensions.

It is interesting to note that the Sierpinski gasket combines self-similarity with another important, but classical, symmetry: rotation. Indeed, the gasket is congruent to itself when rotated around its center by an angle of  $120^\circ$  (or any integer multiple of  $120^\circ$ ). Such symmetries, combining infinite scaling and finite rotation, can be observed in many fractals—and the prescient works of Maurits Escher (see Figure 12).

Incidentally, the “fractal” dimension of a fractal set is not necessarily a non-integer. For example, the Hausdorff dimension of the self-similar board of non-attacking queens (see pages 4–7) equals 1.



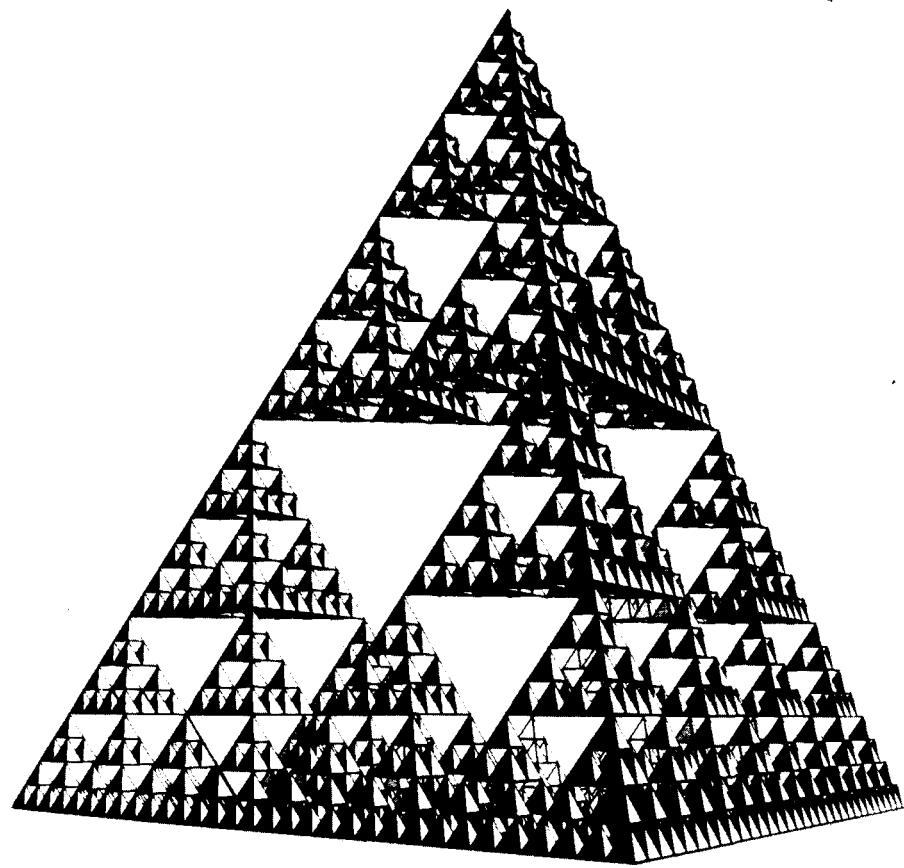
**Figure 11** (A) Generator for the Sierpinski gasket. (B) Toward the Sierpinski gasket: a two-dimensional uncountable set with zero measure and fractal dimension  $\log 3 / \log 2 = 1.58 \dots$ .



**Figure 12** An image by Escher that combines rotational symmetry and infinite scaling [Esc 71].

Figure 13 illustrates a three-dimensional generalization of the Sierpinski gasket. Its construction starts with a regular tetrahedron (a pyramid bounded by four equilateral triangles) from which a half-size upside-down regular tetrahedron has been cut out. This process is repeated on the four remaining tetrahedra and all subsequent tetrahedra to yield the spidery tower shown in Figure 13.

The Hausdorff dimension of this self-similar construction follows immediately from the first step: with  $N = 4$  remaining pieces of size  $r = \frac{1}{2}$ , we have  $D_H = \log 4 / \log 2 = 2$ , a fractal dimension that happens to be an *integer*, but a full unit less than the embedding Euclidean dimension  $d = 3$ .



**Figure 13** A three-dimensional version of the Sierpinski gasket. Its fractal dimension,  $\log 4/\log 2 = 2$ , has an integer value (2), albeit smaller than the dimension of the supporting space (3).

Sierpinski gaskets in two or more dimensions model many natural phenomena and man-made structures. Think of the Eiffel Tower in Paris, designed by Gustave Eiffel. If, instead of its spidery construction, it had been designed as a solid pyramid, it would have consumed a lot of iron, without much added strength. Rather, Eiffel used trusses, that is, structural frames whose members exploit the rigidity of the triangle. (A triangle, in contrast to a rectangle, cannot be deformed without deforming at least one of its sides.) However, the individual members of the largest trusses are themselves trusses, which in turn are made from members that are trusses again. This self-similar construction guarantees high resilience at low weight. The structures of Gothic cathedrals, too, betray great faith in this principle of achieving maximum strength with minimum mass.

And Buckminster Fuller (1895–1983) and his skeletal domes popularized the fact that strength lies not in mass but in *branch points*. In fact, counter to intuition, the Sierpinski gasket and like constructions consist of nothing *but* branch points. (A branch point on a curve has more than two points arbitrarily close to it.) Certain boundary sets (of strange attractors, for example) share this property with the Sierpinski gasket (see pages 38–40, where this exclusive branching is exploited to “settle” an international boundary problem).

The Sierpinski gasket is good for another counterintuitive surprise. For Euclidean bodies in  $d$  dimensions, the volume  $V$  is proportional to  $R^d$ , where  $R$  is some linear measure of size. Surface area  $S$  varies as  $R^{d-1}$ . Thus,  $S \sim V^{(d-1)/d}$ . For example, for  $d=3$ ,  $S \sim V^{2/3}$ . In fact, for the sphere,  $S = 4\pi R^2 = (36\pi)^{1/3} \cdot V^{2/3}$ .

However, for fractal objects this simple Euclidean relation often breaks down. As we have seen, the Hausdorff dimension of the Sierpinski gasket equals  $\log 3 / \log 2 \approx 1.58$ . What is the Hausdorff dimension of its edges? It is easy to see that every time we reduce the yardstick by a factor 2, the number of edge segments goes up by a factor 3. Thus, the Hausdorff dimension of the edges, the “surface” of the Sierpinski gasket, is also  $\log 3 / \log 2$ : “volume” and “surface” have the same dimension. We can also see this by expressing the mass  $M(R)$  of the gasket, that is, the number of points inside a circle of radius  $R$ , as a function of the radius: on average we find  $M(R) \sim R^{1.58}$ . But for the total edge lengths  $S(R)$  inside the circle we find the same dependence:  $S(R) \sim R^{1.58}$ . As a consequence, for the Sierpinski gasket, area  $V$  and edge length  $S$  are proportional to each other:  $V \sim S$ , a paradoxical result indeed.

We shall have the pleasure of meeting the Sierpinski gasket again, both in its original form and, in Chapter 17, in a discrete version, the Laplace triangle modulo 2. In the Meantime, let us relish some of its refreshing implications, such as the board game invented by the mythical Sir Pinski.

## Sir Pinski’s Game and Deterministic Chaos

Consider the following “parlor game” played by two or more persons:

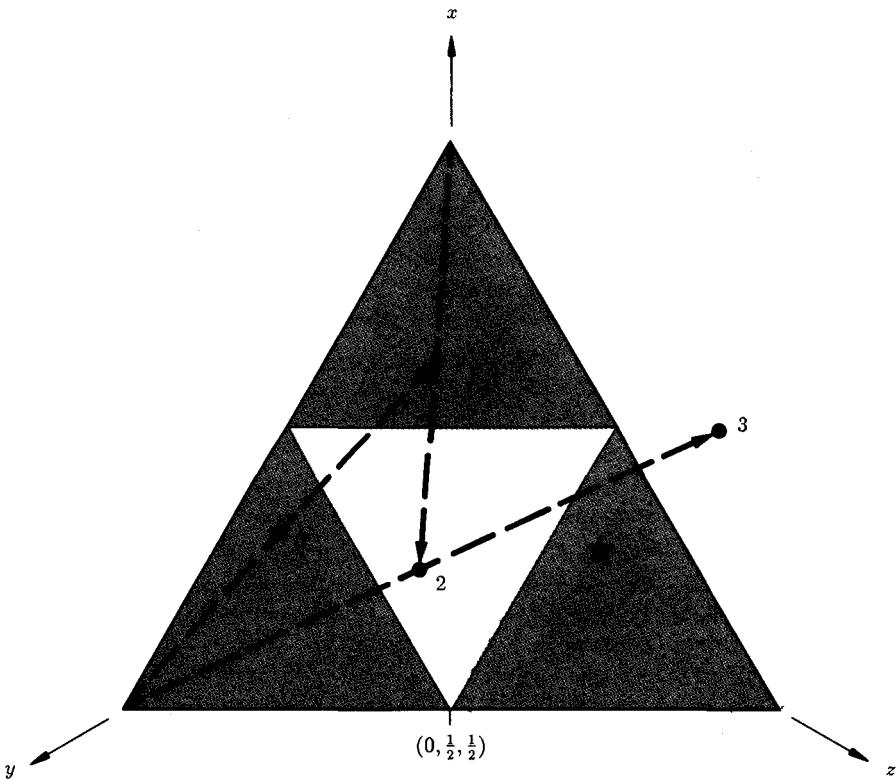
- Each player picks an initial point inside an equilateral triangle.
- Then the player doubles its distance from the nearest corner along a straight line from that corner, thereby arriving at a point  $p_1$ .

The player who can repeat the distance doubling most often without falling outside the triangle wins the game. As we shall see, there are *uncountably* many, but still *very few*, initial points that guarantee winning or a tie, “very few” in the sense that a random choice has a zero probability of infinite survival under the rules of the game.

Figure 14 shows the equilateral triangle with an initial choice, marked by 0, and its three successors or “images” marked 1, 2, and 3. Note that the point 3 already lies outside the triangle; the initial choice is therefore not a good pick. How can we avoid such bad points? We will answer the question first geometrically and then arithmetically.

Note that point 2, which lies inside the small, white (upside-down) triangle, is mapped outside the large triangle. In fact, a little reflection (in more than one sense of the word) will show that all points inside the small white triangle will be mapped to the outside. Thus, the white triangle is *out* as a good starting area—and so, of course, are its preimage and the preimages of the preimage and so on *ad infinitum*. In other words, any point that, sooner or later, is mapped into the white triangle is a loser.

But what *are* the preimages of the white triangle? A little more reflection will show that they consist of three half-size upside-down triangles, one inside

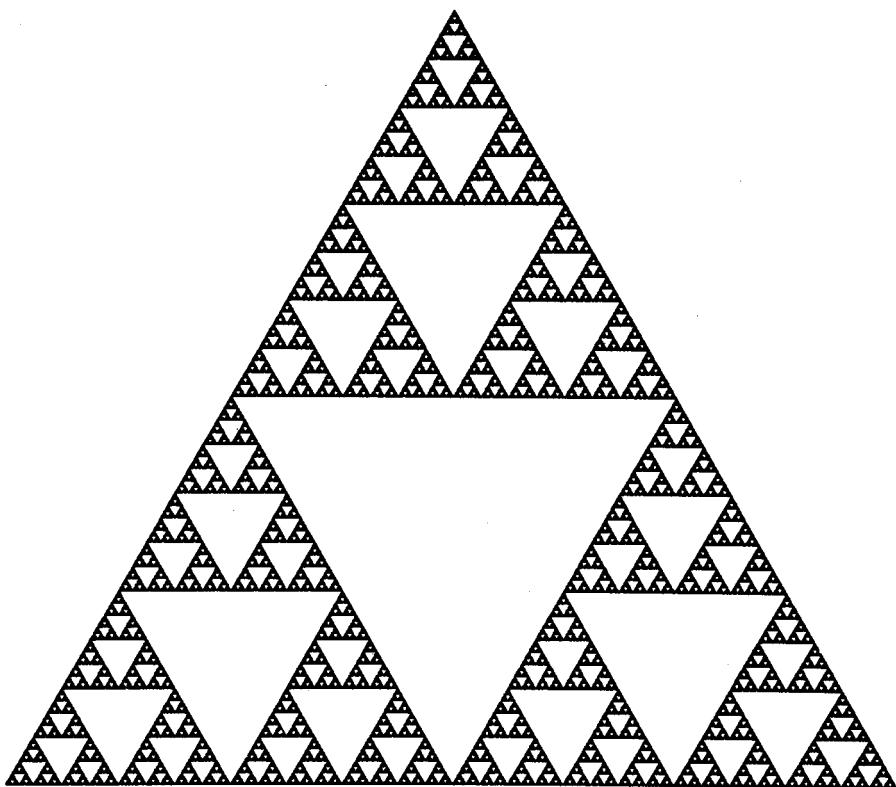


**Figure 14** Sir Pinski's chaos game: How many times can you double your distance to the nearest vertex without leaving the large equilateral triangle?

each of the remaining three dark triangles. And the preimages of these three preimages are *nine* upside-down triangles, again scaled down in side length by a factor 2 and cut out of the centers of the nine remaining quarter-size triangles.

Thus, in delimiting good initial choices, we find ourselves constructing a *self-similar* figure, the well-known Sierpinski gasket (see Figure 15), a Cantor set embedded in two dimensions, with zero area and Hausdorff dimension equal to  $\log 3/\log 2 \approx 1.58$ . Picking an initial point at random, however, will almost certainly land us in white territory, a prelude to the disaster of being eventually mapped outside the big triangle.

In order to avoid potential disputes resulting from poor drafting, the just described Sir Pinski game should be played arithmetically; that is, the initial points and all their images should be stated by their coordinates in a suitable coordinate system. Although two coordinates suffice to locate a point in the plane, a more convenient system, matched to the symmetry of the triangle, uses three coordinates,  $x$ ,  $y$ , and  $z$ , as shown in Figure 14. The corners of the triangle have the



**Figure 15** Sierpinski gasket, the winning set of Sir Pinski's game.

value 1 for the corresponding coordinates, and the opposite sides have the value 0. Thus, the midpoint on the horizontal side, for example, has the coordinates  $x = 0$  and  $y = z = \frac{1}{2}$ .

Of course, a set of three coordinates in the plane is redundant, and the coordinate values cannot be chosen independently, because of the constraint  $x + y + z = 1$ . The points *inside* the triangle are further subject to the constraints  $x > 0$ ,  $y > 0$ , and  $z > 0$ .

Now, how does our mapping, defined by doubling the distance to the nearest corner, look arithmetically? Suppose for our initial choice  $(x_0, y_0, z_0)$  the nearest corner is the lower left ( $y$ ) corner. Then the image of  $(x_0, y_0, z_0)$  is  $(2x_0, 2y_0 - 1, 2z_0)$ . The factors of 2 occurring in this mapping suggest using binary notation for the coordinates. Multiplication by 2 is then a simple left shift of the digits. Thus, the point 0 in Figure 14, which has the approximate coordinates  $(\frac{5}{16}, \frac{39}{64}, \frac{5}{64}) = (0.0101, 0.100111, 0.000101)$ , suffers the following fate:

$$\begin{array}{lll} x_0 = 0.0101 & y_0 = 0.100111 & z_0 = 0.000101 \\ x_1 = 0.101 & y_1 = 0.00111 & z_1 = 0.00101 \\ x_2 = 0.01 & y_2 = 0.0111 & z_2 = 0.0101 \\ x_3 = 0.1 & y_3 = -0.001 & z_3 = 0.101 \end{array}$$

Here  $y_3$  is negative—that is,  $(x_3, y_3, z_3)$  lies outside the triangle—and the player who picked the point  $(x_0, y_0, z_0)$  is eliminated from the game.

Once outside the triangle, the images “escape” to infinity. Arithmetically, what are the *good* initial choices that stay inside the triangle and thus are never eliminated? If we can find a complete answer to this question, we will also have discovered an *arithmetic* description of the Sierpinski gasket to boot!

Then what led to the unwanted negative value of  $y_3$  in the mapping? The answer is that  $y_2$ , the largest preceding coordinate, was smaller than  $\frac{1}{2}$ ; in other words, the first fractional binary digit of  $y_2$  was a 0 and not a 1. Hence, a good initial point must not have 0s in all three coordinates for any of its fractional binary places. This rule is violated by  $(x_0, y_0, z_0)$ , which has only 0s (not a single 1) in the third binary places. Taken together with the constraint  $x + y + z = 1$ , this means that good points  $(x, y, z)$ , that is, members of the Sierpinski gasket, have precisely one 1 and two 0s in every binary place of  $x$ ,  $y$ , and  $z$ .

There is a charming similarity here with the ternary representation of the original (middle-third-erasing) Cantor set, which contains only 0s and 2s and no 1s. In fact, the connection between this arithmetic representation of the Sierpinski gasket and Cantor’s construction is quite close: the first missing 1 of a Cantor number (right behind the “ternary point”) corresponds to the deletion of the interval  $(\frac{1}{3}, \frac{2}{3})$  from the unit interval. The absent 1 in the second ternary place corresponds to the subsequent elimination of the two intervals  $(\frac{1}{9}, \frac{2}{9})$  and  $(\frac{7}{9}, \frac{8}{9})$ , and so on.

What does the absence of three 0s in the binary representation of the Sierpinski gasket mean geometrically? Three 0s in the first binary place to the right of the binary point would mean that neither  $x$ ,  $y$ , nor  $z$  exceeds  $\frac{1}{2}$ . Geometrically speaking, this corresponds to the central half-size upside-down triangle, left white in Figure 14, which is thus excluded from the Sierpinski gasket—as indeed it is in the first step of the geometric construction of the gasket. We could also argue that a 1 in the first binary place of  $x$ ,  $y$ , or  $z$  means that either  $x$ ,  $y$ , or  $z$  is greater than (or equal to)  $\frac{1}{2}$ . Geometrically, these three cases correspond to the three half-size right-side-up triangles (shaded in Figure 14).

What would three 0s in the *second* binary place correspond to geometrically? A little triangular reasoning will reveal that they correspond to quarter-size upside-down triangles cut from the centers of the three half-size shaded triangles left over after the first cutting operation. In general, three 0s in the  $n$ th binary place imply the elimination of  $3^{n-1}$  upside-down triangles of side length  $2^{-n}$  from the  $3^{n-1}$  right-side-up triangles left standing after  $k-1$  cutting operations. Thus, the binary representation of the Sierpinski gasket corresponds, place by place, to its geometric construction. The two descriptions are equivalent.

A proper Sierpinski point is  $(\frac{1}{3}, \frac{2}{3}, 0) = (0.\overline{01}, 0.\overline{10}, 0)$ , for example, which lies on the left side of the triangle, one-third up from the lower left corner. Our distance-doubling mapping will make it alternate, with period length 2, with the point  $(\frac{2}{3}, \frac{1}{3}, 0)$  as is clear both geometrically and from the period length of 2 of the binary fractions for  $\frac{1}{3}$  and  $\frac{2}{3}$ .

Are there periodic points with period length 3? If so, our mapping should be equivalent to a  $120^\circ$  rotation. To find such points, we simply have to consider binary fractions with period length 3. And indeed,  $(0.\overline{010}, 0.\overline{001}, 0.\overline{100}) = (\frac{2}{7}, \frac{1}{7}, \frac{4}{7})$ , which is marked by a star in Figure 14, is such a point. The other two points of its orbit are  $(\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$  and  $(\frac{1}{7}, \frac{4}{7}, \frac{2}{7})$ , reached by twice rotating  $120^\circ$  counterclockwise. The only other period-3 orbit is obtained by interchanging two coordinates—for example, by starting with  $(\frac{1}{7}, \frac{2}{7}, \frac{4}{7})$ , whose two successors are found by  $120^\circ$  rotations clockwise:  $(\frac{2}{7}, \frac{4}{7}, \frac{1}{7})$  and  $(\frac{4}{7}, \frac{1}{7}, \frac{2}{7})$ .

Periodic points exist for all period lengths. Thus, for example, the point  $(1, 0, 0)$ , the upper corner of the triangle, has period length 1; that is, it (and the other two corners) are *fixed points*. We shall later encounter this scenario and similar mappings again, and we shall derive a formula for the number of different orbits of a given period length. (This derivation will involve the *Moebius function* from number theory, a function whose multifarious functions in higher arithmetic one should know about; it “twists things around,” much like the much better known Moebius strip.)

Under the rules of the game, the image of a point in the Sierpinski gasket (called a *Sierpinski point*) is a Sierpinski point. The Sierpinski points therefore form what is called the *invariant set* of the map: once a Sierpinski point, always a Sierpinski point. If you start with a Sierpinski point with irrational coordinates, its orbit may look completely chaotic, but the succession of image points is fully determined by the coordinate values of the initial point. This is why such behavior,

which abounds in nature, is called *deterministic chaos*: the rules governing the "game" are unambiguously deterministic, but the results are ultimately unpredictable because, ironically, the *real* world does not admit the vast majority of *real* numbers, namely, all those with infinite precision.

## Three Bodies Cause Chaos

*One will be struck by the complexity of this figure which I do not even attempt to draw. Nothing more properly gives us an idea of complication of the problem of three bodies and, in general, of all the problems in dynamics where there is no uniform integral*

—HENRI POINCARÉ

A more attractive property of the Sierpinski points is that they are all repulsive points, or *repellors*; that is, a point arbitrarily close to a Sierpinski point will not stay near its images under our mapping, let alone be attracted to it; rather, its distances will *diverge* from the corresponding images of the Sierpinski point. In fact, the divergence will be exponential. The reader with a personal (or impersonal) computer is encouraged to try this and see for him- or herself. The reason for this exponential divergence is easy to see because our mapping corresponds to left shifts of the binary digits that encode the coordinates of the points. Thus, sooner or later, the first "error" bit will arrive at the binary point, which means that the initial difference, no matter how small, will have been magnified to half the height of the triangle. After that, all succeeding bits are random errors; the motion of an initially almost periodic point will become *chaotic*. In fact, this simple example contains the very essence of chaos and accords fully with its definition: small initial errors grow exponentially until they "dominate" any regular motion.

Although we may still not be aware of it, chaotic motion is much more widespread in nature than regular motion [Wis 87]. In fact, the jury is still out on whether planetary motion, the repository of regularity, is not chaotic in the long run. Certainly, Pluto and several other heavenly bodies already "stand" convicted of causing (or suffering) chaos. The smoke rising from a motionless cigarette in still air, first forming a regular ("laminar") flow, becomes a turbulent swirl only a few inches above the ashtray (see Figure 16). And what happens when two stars (a "double star"), encircling each other elliptically (good behavior!), meet a third star? Their regular motion turns wildly chaotic, (see Color Plate 1A). But, just as with human triangular relations, in the end two stars may pair off again to resume a regular orbit, as is the case in Color Plate 1A. However, one member of the initial couple may have switched partners in the course of the chaotic confusion during the three-body encounter (see Color Plate 1B).



**Figure 16** The laminar and the turbulent in cigarette smoke.

Of course, Newton's laws of gravitation, which govern the motions of our three heavenly bodies, are completely deterministic. But the far-future fate of the three partners can depend very sensitively on their initial positions and velocities. Here we have another case of deterministic chaos. In fact, the ultimate stability of the sun's planetary system, including Earth, humankind's common spaceship, has still not been rigorously proved—even without meddling from Nemesis, the hypothetical dark and distant sister of our sun.

We shall encounter more of this chaos, so intimately related to self-similarity, in the body of this book. And for the insatiable reader who has become addicted to disorder, there is James Gleick's recent bestseller *Chaos* [Gle 87] to devour.

## Strange Attractors, Their Basins, and a Chaos Game

*Determinism, like the Queen of England,  
reigns—but does not govern.*

—MICHAEL BERRY

Let us look at the *inverse* mapping of Sir Pinski's game (see pages 20–25) and see whether it holds any surprises (or can teach us a lesson or two). Inverses are generally good to look at for a variety of reasons. For one, repellors turn into attractors (and vice versa). And new concepts arise, such as *basins of attraction* and *strange attractors*.

In the inverse of Sir Pinski's map, we again pick a point inside (or outside) an equilateral triangle, but now we *halve* the distance to the most *distant* corner. We can be pretty sure that halving will never lead to a divergent explosion. But what points will we converge on?

First we remember that in Sir Pinski's game a Sierpinski point will remain a Sierpinski point, and since the map has a unique inverse (except for points with equal distances to two or all three corners), the same will be true for the inverse map.

But what happens to all the other points inside the triangle? Arithmetically, the inverse map looks as follows. If  $x$  is the smallest coordinate (i.e., if the  $x$  corner is the most distant one), then the inverse map of  $(x, y, z)$  is  $((1+x)/2, y/2, z/2)$ . In binary notation the division by 2 means a right shift and adding  $\frac{1}{2}$  means inserting a 1 in the place to the right of the binary point.

Let us start with a non-Sierpinski point, for example,  $(\frac{1}{8}, \frac{1}{4}, \frac{5}{8}) = (0.001, 0.010, 0.101)$ , and follow its course. By our rule it will map into  $(0.1001, 0.001, 0.0101)$ , which will go into  $(0.01001, 0.1001, 0.00101)$ , and so forth. Note that with each mapping we insert exactly one 1 and two 0s in the first place behind the binary point. With each further mapping this triplet will move one binary place to the right. Thus, asymptotically, no matter where we start, we approach

a Sierpinski number, which has exactly one 1 in *all* binary places. (And if we start with a Sierpinski number, we will, of course, stay with the set.)

In fact, we will converge on one of the two period-3 cycles:  $(0.\overline{001}, 0.\overline{010}, 0.\overline{100})$  and its two successors, or  $(0.\overline{001}, 0.\overline{100}, 0.\overline{010})$  and its orbit—which of the two is determined by the coordinate values of the initial point (whether its ordered values constitute an even or odd permutation of  $x, y, z$ ).

This insight, too, can be exploited in a game in which one has to predict, as closely as possible, the twelfth iterate, say, given only a rough initial location. If for the initial point  $x_0 > y_0 > z_0$  holds, then after  $3n$  mappings the image will approach the period-3 point with  $x > y > z$ , namely,  $(0.\overline{100}, 0.\overline{010}, 0.\overline{001})$ , within less than  $2^{-3n}$  (for an initial point inside the triangle). This point is thus the attractor for the  $60^\circ$  sector defined by  $x > y > z$  (whose apex is the center of Figure 14). This sector is its *basin of attraction* for the threefold iterated inverse Sir Pinski map. The five other period-3 points have the remaining five  $60^\circ$  sectors as their basins of attraction. The boundaries of these basins are smooth (in fact, straight) lines, in contrast to many other basins that we shall get to know, which have fractal rims.

The inverse Sir Pinski game is kin to a “game” called *chaos game* invented by Michael Barnsley, as described in his recent book *Fractals Everywhere* [Bar 88]. In Barnsley’s chaos game, players “roll” a three-sided die, marked  $x, y$ , and  $z$ , and halve the distance of a preselected point inside a given triangle to the corresponding corner, also marked  $x, y$ , or  $z$ . (We leave the construction of the die to the reader as an exercise.) Alternatively, a random number generator with three possible outcomes will do.

What is the basin of attraction of the chaos game? The Sierpinski gasket (affinely transformed if the given triangle is not equilateral)! The proof follows directly from our analysis of Sir Pinski’s game. However, the orbit of any initial point, as its iterates approach the attractor, will be completely chaotic. Such an attractor with infinitely many points that form a Cantor-like set is called a *strange attractor*—strange, because familiar attractors consist of either single points (fixed points), finitely many points (periodic orbits), or continuous manifolds that give rise to periodic or aperiodic orbits.

Strange attractors are encountered in many (nonlinear) physical, chemical, and biological systems that are “not integrable” and therefore show ultimately unpredictable, *chaotic* behavior. In fact, the usual “textbook” cases, nicely integrable, are now recognized as singular exceptions; the *real* world outside the textbooks, including romantic attraction, remains largely unforeseeable, moving along strange attractors, sometimes *very* strange attractors indeed.

However, not all is lost; the world is not complete chaos. Strange attractors often do have structure: like the Sierpinski gasket, they are self-similar or approximately so. And they have fractal dimensions that hold important clues for our attempts to understand chaotic systems such as the weather.

Strange attractors have recently found another, most surprising application. Barnsley has shown in his abovementioned book that many ordinary images, be

they black-and-white or in color, can be approximated by a superposition of the strange attractors of a limited number of affine transformations, each transformation occurring with a given probability. An affine transformation in the plane is specified by a rotation, a scaling, and a displacement for each of the two coordinates. Since affine transformations in the plane are thus completely specified by six real numbers, an entire picture can be specified by some multiple of seven numbers, say  $7 \cdot 13 = 91$  numbers.<sup>5</sup>

To understand the approximation of images by strange attractors better, we note first that the Sierpinski gasket consists of three triangular regions, each of which is a contractive affine transformation of the entire gasket. (A contractive transformation decreases the distances between all pairs of points.) Geometrically, these affine transformations correspond to moving a point along a straight line to half the distance to one of the three corners of the Sierpinski gasket. Numerically, the transformations are given by inserting a 1 behind the binary point for  $x$ ,  $y$ , or  $z$  and one 0 each for the other two coordinates while shifting all binary digits of the point  $(x, y, z)$  one place to the right.

To generate the entire gasket, we start with an arbitrary initial point  $(x_0, y_0, z_0)$  somewhere near the gasket and select one of the three transformations—rotation, scaling, or displacement—at random to give us a point  $(x_1, y_1, z_1)$ . On successive mappings, these three transformations are selected independently with given probabilities  $p_1$ ,  $p_2$ , and  $p_3$ .

Because of the rules for inserting 1s and 0s, it is clear that the iterates will soon grow closer and closer to Sierpinski points, that is, members of the gasket. Because of the randomness of selecting the different mappings, the iterates will not become stuck in a “periodic rut” but will “hop” around and “illuminate” the entire gasket.

For equal probability, or  $p_1 = p_2 = p_3 = \frac{1}{3}$ , each of the three parts of the gasket will be visited with equal likelihood. By choosing other values for the  $p_k$ , we can produce different degrees of illumination or shadings for different parts of the attractor.

This process for generating images can be further generalized as follows. Instead of selecting the three corners of an equilateral triangle, we can pick any three points in the plane. In fact, we can specify *any* number of completely general affine transformations, each with its own probability of being chosen. But even with these generalizations, it is still surprising that, using fewer than 100 parameters, realistic looking scenes from nature can be generated by this “strangely attractive” method.

The promises for highly effective image data compression by *iterated function systems*, as the method is called by Barnsley, are mind-boggling—once image

---

5. In number theory, 91 is jocularly known as the smallest composite number that *looks* like a prime, the reason being that there is no simple rule (other than division) to recognize its two factors. But note that 91 times 11 equals 1001, so that for numbers above 1000 divisibility by 7, 11, or 13 can be tested by subtracting the appropriate multiple of 1001. Thus, 9399 is divisible by 13 because 390 is.

*decomposition* in terms of attractors is computationally expedited. Color Plate 2 illustrates an image generated in this fashion.

## Percolating Random Fractals

The Sierpinski gasket is an example of a two-dimensional *deterministic* fractal. Picking a point inside the triangle from which the gasket has been carved, we know immediately whether it is a member of the fractal set or “falls through the cracks.” Many man-made deterministic fractals, like the Sierpinski gasket, are visually attractive and algebraically intriguing. However, most of *nature’s* fractal gaskets are best modeled by *random* fractals, generated by stochastic processes. Among the many cases that have been diagnosed from the point of view of random fractals is the spread of epidemics and forest fires. Other examples of such fractals are random resistor networks, polymer bonds, and, apparently, the ice floes drifting through the Bering Sea.

To make things as discrete and simple as possible, consider a large square lattice whose lattice points are “occupied” independently with probability  $p < 1$  (see Figure 17). The “occupants” could be trees, people, atoms, or whatever; it does not matter. The fraction of the lattice points that are unoccupied or “empty” equals  $1 - p$ . An important question is the following: Do the occupied sites form a *continuous* path from the lower edge of the lattice to the upper edge? A continuous path is defined as a path that goes from an occupied site to a neighboring occupied site. (The neighbors of a site are the sites immediately to the north, east, west, or south of it.) If such a path exists, the lattice is said to *percolate* (as in a coffee percolator, from Latin, “to flow through”). If the occupied sites were occupied by air and the “empty” sites by ground coffee, then the water could indeed percolate through the coffee.

The smallest density  $p$  of occupied sites for which the infinite lattice percolates is called the critical density or *percolation threshold*  $p_c$ . In spite of its simple definition, the exact percolation threshold of sites on the square lattice is still unknown. Massive Monte Carlo simulations put it at approximately 0.59275, with more digits constantly being appended as increasingly more powerful computers are brought to bear.

In addition to *site* percolation, there is *bond* percolation, in which *all* the sites are occupied but the “bonds” from a site to its immediate neighbors occur with probability  $p < 1$ . Missing bonds have a probability of  $1 - p$ . Percolation here means a connected path of *bonds* through the lattice. The bond percolation threshold for the infinite square lattice is known exactly:  $p_c = 0.5$ . But it took two decades of simulation and theory to prove this simple-looking result. Thus, in a large random network of electrical resistors based on a square lattice, electric current could flow between two opposite sides if at least half the bonds were

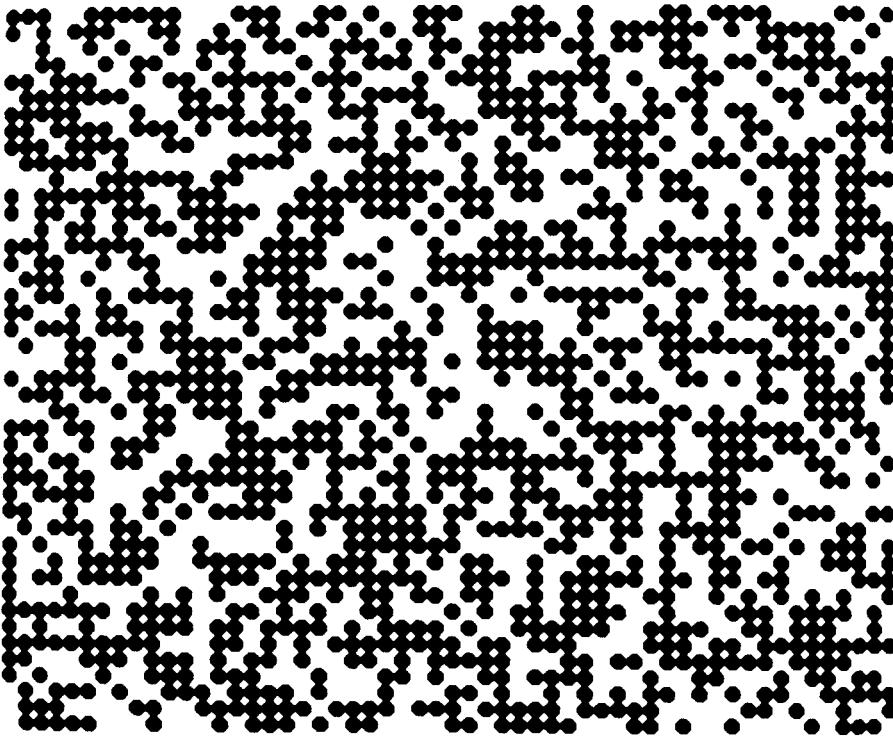


**Figure 17** Square lattice with randomly occupied sites below percolation threshold.

conducting. The resistors through which the current actually flows are called the *backbone* of the cluster; the other resistors are called *dangling bonds*.

At the site percolation threshold ( $p \approx 0.5927$  for the square lattice), the occupied sites of the infinite lattice form clusters of connected sites of all sizes. In fact, their distribution follows a simple power law: the number  $n(s)$  of clusters having  $s$  occupied sites is proportional to  $s^{-\tau}$  with  $\tau = 187/91 = 2.054945$  for the square lattice [Sta 85]. The power law  $n(s) \sim s^{-\tau}$  means that the ratio of the numbers of clusters of two different sizes is independent of cluster size  $s$ ; it depends only on the size ratio.

Figure 18 shows clusters of many sizes from pairs of sites ( $s = 2$ ) to a “spanning cluster” that connects the upper and lower edges. A large lattice, 10 times larger than the one shown in Figure 18, would show exactly the same cluster distribution, except that it could accommodate clusters 10 times larger. Thus, percolating clusters are self-similar or *free of scale*, from the distance of neighboring sites to the size of the entire lattice. However, below the percolation threshold, the upper cutoff length for self-similarity is not the size of the lattice but rather the *correlation length*  $\xi$ , defined as the length over which the probability



**Figure 18** Square lattice, probability of occupied sites equal to percolation threshold. Clusters occur on many size scales. These clusters form a statistically self-similar pattern.

of two sites belonging to the same cluster has decayed to  $1/e \approx 0.368$ . For distances smaller than  $\xi$ , the occupied sites form a fractal; above  $\xi$  Euclidean geometry prevails, with the number of occupied sites  $M(R) \sim R^d$ , where  $d$  is the Euclidean embedding dimension. At the percolation threshold,  $\xi$  diverges to infinity and the probability that two sites, even at an arbitrarily large distance, belong to the same cluster is bounded away from zero.

Percolating clusters, being self-similar fractals, ought to have fractal dimensions. The first measure that comes to mind is the so-called *mass exponent*  $D_m$ , which measures the number of occupied sites (the “mass”)  $M(R)$  within a circle of radius  $R$ :  $M(R) \sim R^{D_m}$ . For Euclidean objects, of course,  $D_m$  equals the Euclidean dimension; for example, the area  $M(R) = \pi R^2$ , or, in other words,  $D_m = 2$ , for a filled disk. But for fractals,  $D_m$  is generally smaller than the Euclidean dimension (also called the *embedding dimension*) that contains (“embeds”) the fractal. For the triadic Cantor set, for example, the fraction of the included set  $M(R)$  grows on average as  $R^{0.63}$ ; that is,  $D_m$  equals the Hausdorff dimension  $D_H = \log 2 / \log 3 \approx 0.63$ . Similarly, for the two-dimensional Sierpinski gasket,  $M(R) \sim R^{1.58}$ ; that is,  $D_m$  again equals the Hausdorff dimension  $D_H = \log 3 / \log 2 \approx 1.58$ .

What is the mass exponent of the percolating cluster on a two-dimensional lattice? Theory gives a value of  $D_m = 91/48 = 1.8958\bar{3}$ , in good agreement with the best values found by simulation [Sta 85].

The integer 91 that appears in the numerical values for both  $\tau$  and  $D_m$  suggests that they are related. In fact,  $\tau - 1 = 2/D_m$ . We shall return at some length to the interesting relationships between characteristic exponents in the chapters on percolation (Chapter 15) and phase transitions (Chapter 16). In Chapter 10, we shall see that in many cases the mass exponent equals the *correlation dimension*  $D_2$ , one—albeit an important one—in an infinite hierarchy of fractal dimensions.

## Power Laws: From Alvarez to Zipf

Homogeneous power laws, like Newton's universal law of gravitational attraction  $F \sim r^{-2}$ , abound in nature—dead and alive alike. Since homogeneous power laws, upon rescaling, remain homogeneous power laws with the same exponent ( $-2$  in Newton's case), such laws are, by definition, self-similar. In other words, Newton's law is *true on all scales*, from the wavelength of light to light-years; it has no built-in scale of its own. Newton's gravitational universe, if we so wished, could be compressed or inflated at will.<sup>6</sup>

The same inverse square law that governs gravitation also describes the falloff of radar power with distance. This simple fact was exploited by German submarines during World War II. By measuring the increase in radar intensity, they could gauge the rate of approach of an enemy plane and dive undersea for safety before the plane could attack.

This tactic worked very well for Grand Admiral Karl Dönitz until the American physicist Luis Alvarez (1911–1988) had a foxy vision, code-named Vixen. Alvarez suggested reducing the radar power so that it would be proportional to the *third* power of the range of the submarine. Thus, while the plane was approaching, the power incident on the unsuspecting U-boat was actually *decreasing*, giving the false impression that the radar plane was flying *away*. A grand idea indeed! (For the attacking plane, however, the received radar power reflected from the boat would still increase as it closed in [Alv 87].)<sup>7</sup>

---

6. Recently, though, some doubt has been cast on the unlimited validity of Newton's law. A still mysterious “fifth force” appears to knock on Newton's underpinnings, adding terms that introduce a natural length scale of a few hundred meters [AZLPAGNCFFMSSBCGHHKSW 89]. At very small scales, Newton's law runs into the *Planck length* ( $10^{-35}$  m), which reminds us that eventually gravitation needs to be properly quantized and endowed with uncertainty.

7. This scheme of Alvarez is somewhat reminiscent of Genghis Khan (“Universal Ruler”) and the wily Mongol tactic perfected by the horsemen of the Golden Horde. While seemingly galloping away from their pursuers, they would actually allow them to close in and then suddenly stand up in their stirrups, turn around in their saddles, and launch their arrows at the dumbfounded enemy.

Another wide-ranging example of a homogeneous law is the one that connects the areas  $A$  of similar plane figures with their diameters, their perimeters, or any other of their linear dimensions  $l$ : areas are proportional to linear dimensions squared, or  $A \sim l^2$ . Of course, this is not true for areas on *curved* surfaces; the radius of curvature introduces a length scale that destroys "truth on all scales." In fact, as everyone knows, distances and areas on the surface of a sphere are limited to a maximum size, given by the radius of the sphere.

In contrast to gravitation, interatomic forces are typically modeled as *inhomogeneous* power laws with at least two different exponents. Such laws (and exponential laws, too) are not scale-free; they necessarily introduce a characteristic length, related to the size of the atoms.

Power laws also govern the power spectra of all kinds of noises, most intriguing among them the ubiquitous (but sometimes difficult to explain)  $1/f$  noise. Thus, the noise in many semiconductor devices is not "white" (i.e., independent of frequency) and not "brown" (with a  $1/f^2$  frequency dependence, like Brownian motion), but has an in-between exponent, which is why it is sometimes called *pink noise*. Pink noise is also a preferred test signal in auditory research, because it has constant power per *octave* (not per hertz) and is thus well matched to the inner ear's frequency scale.

And, as we shall see in the course of our excursion into the world of fractals, power-law exponents do not have to be integers; they can be, and often are, *fractions*.

Not surprisingly, we find homogeneous power laws not only in the inanimate world; they inhabit living nature, and particularly human perception, too. Thus, over much of the auditory amplitude range, subjective loudness  $L$  is proportional to the physical sound intensity  $I$  raised to the three-tenths power:  $L \sim I^{0.3}$ . This means that merely to double the loudness of a rock group of five musicians, say, we have to increase their number *tenfold*, to 50 players of equal power output. (This minor calculation explains the resounding enamoration of popular music makers with electronic amplifiers.)

By the same token, if we want to halve the loudness of a continuous "rumble" emanating from a busy highway, we have to reduce the acoustic noise output by a factor of ten! This may sound difficult, but it is not, at least not from a purely physical point of view: tire noise—the main culprit at steady highway speeds—decreases drastically with decreasing vehicle speed. In fact, the noise intensity is approximately proportional to the *fourth* power of speed.

On the other hand, a tenfold increase in the average intensity of traffic noise caused by a tenfold increase in traffic *density* can raise the rate of complaints by irate residents perhaps a *hundred* fold: one loud truck every 5 minutes may be tolerable, but one every 30 seconds could be a nightmare and would certainly make outdoor conversation nearly impossible. And what is true for trucks is just as true for low-flying aircraft.

Power laws are also ubiquitous in economics. In fact, nearly 100 years ago, the Italian economist Vilfredo Pareto (1848–1923), working in Switzerland, found

that the number of people whose personal incomes exceed a large value follows a simple power law [Par 1896, Man 63a]. Other instances of power laws in economics and the fallacies of trading schemes based on them are discussed by Mandelbrot [Man 63b, 63c].

One of the more surprising instances of a power law in the humanities is Zipf's law connecting *word rank* and *word frequency* for many natural languages. (The word with rank  $r$  is the  $r$ th word when the words of a language are listed with decreasing frequency.) This law, enunciated by George Kingsley Zipf (1902–1950), states that, to a very good approximation, relative word frequency  $f$  in a given text is inversely proportional to word rank  $r$ :

$$f(r) \approx \frac{1}{r \ln(1.78R)}$$

where  $R$  is the number of different words [Zip 49]. Laws like  $f(r) \sim 1/r$  are called *hyperbolic laws*. If we assume  $R = 12,000$ , for example, we find that the relative frequencies of the highest-ranking words (*the*, *of*, *and*, *to*, and so on, in order of rank) are approximately 0.1, 0.05, 0.033, 0.025, and so on.

Figure 19 shows the close match between Zipf's homogeneous power law and actual data. Claude Shannon, the creator of information theory, has used Zipf's law to calculate the entropy of a source of English text that sputters words independently with Zipf's probabilities [Sha 51]. This entropy is given approximately by

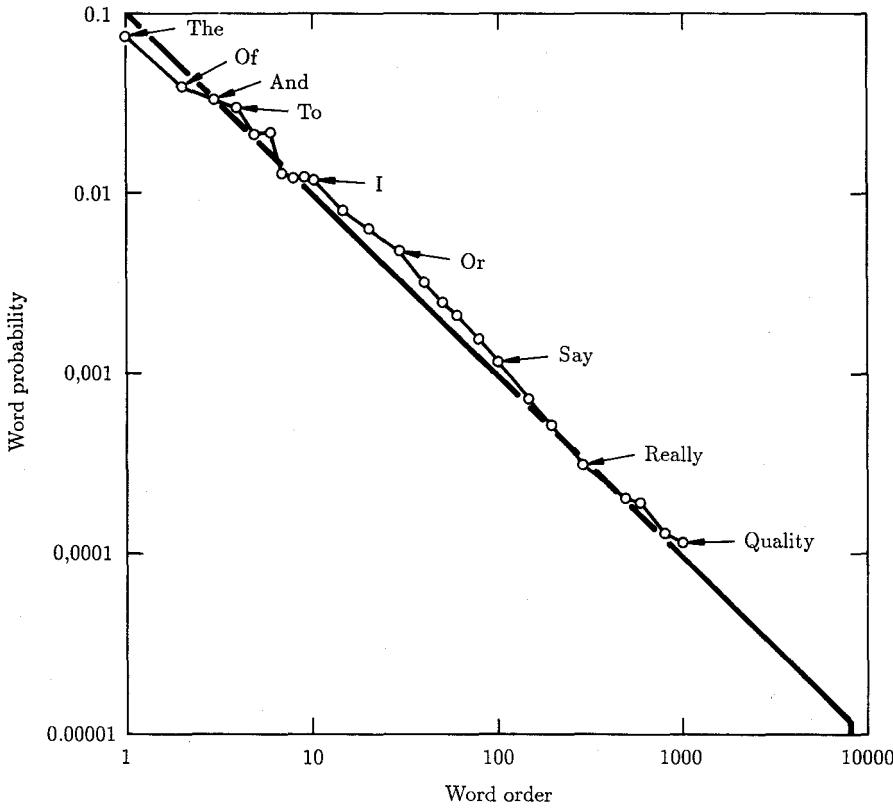
$$H = \frac{1}{2} \log_2 (2R \ln 2R) \quad \text{bits per word}$$

For  $R = 12,000$ , we get  $H \approx 9$  bits per word, while  $R = 300,000$  yields an entropy of about 11.5 bits per word. Of course, this is only an upper bound, because words (though perhaps independent of actions) are not independent of each other—except in random “poetry.” This interdependence of words (“redundancy”) in a meaningful text, of course, reduces the entropy.

Considering that the average length of English words is about 4.5 letters, or 5.5 “characters” including one space between words, we see that the entropy of English text is roughly bounded by 2 bits per character.

Zipf's hyperbolic law, which is applicable not only to the language as such but also to individual writers, has some rather curious consequences. To wit, for a good writer with an active vocabulary of  $R = 100,000$  words, the 10 highest-ranking words occupy 24 percent of a text, while for basic (newspaper?) English with one-tenth the vocabulary ( $R = 10,000$ ), this percentage barely increases (to about 30 percent). Of course, any writer would find it difficult to avoid words like *the*, *of*, *and*, and *to*.

Zipf has endeavored to derive his law from *Human Behavior and the Principle of Least Effort* (the title of his 1949 treatise). But Mandelbrot, in an early effort, has shown that a monkey hitting typewriter keys at random will also produce a



**Figure 19** Word frequency as a function of word rank follows Zipf's law.

"language" obeying Zipf's hyperbolic law [Man 61]. So much for lexicographic *Least Effort!*

A detailed analysis shows that if the monkey's typewriter has  $N$  equiprobable letter keys and a space bar (with probability  $p_0$ ), then his words (defined as letter sequences between spaces) have relative frequencies.

$$f(r) \sim r^{-1 + \log(1 - p_0)/\log N}$$

With  $N = 26$  and  $p_0 = \frac{1}{5}$ , say, the exponent of  $r$  equals  $-1.068$ , only slightly less than  $-1$ . In general, the monkey words can be modeled as a Cantor set with a fractal dimension  $D$  that equals the reciprocal of the exponent of  $1/r$ . In our example,

$$D = \frac{1}{1 - \log(1 - p_0)/\log N} \approx 0.936$$

For a nine-letter alphabet and  $p_0 = \frac{1}{10}$ , the exponent equals  $-1.048$ , corresponding to a Cantor "dust" with  $D \approx 0.954$ . An arithmetic model for the (infinitely many) words of this nine-letter "language" is all the decimal fractions between 0 and 1 in which the digit 0 never occurs (not counting 0s at the end of terminating fractions).

Here are a few "three-letter" words of this language: .141, .241, .643, .442, .692, .121. Of course, .103, .707, and .0̄3 are nonwords because they contain 0s.

Such languages do not have an average rank, but the *median* word rank of our "exemplary" language is an astonishing 1,895,761; that is, it takes the 1,895,761 most frequent words of the language to reach a total probability of one-half. (By contrast, the median word rank of English texts lies between 100 for typical media output and 500 for highly literate writers.) Thus, the monkey, while strictly clinging to Zipf's law, produces a rather wordy (and otherworldly) language.

Another, equally surprising "speech pathology" of the monkey language is the impossibility of constructing a dictionary for it, because its words form an *uncountable* Cantor set. (We would perhaps not be put off by an infinitely thick dictionary, as long as its entries could be sequentially numbered—but we could never countenance an uncountable compendium.)

If the monkey language has a fractal dimension, does it have any self-similarities? It certainly has. Multiply all words of the "decimal language" by 10 and drop the integer part (or, in general, just strike out the leftmost "letter" of each word) and you have another monkey word (most likely forming a nonsensical word sequence). In fact, the words of such languages grow on self-similar trees. Take any branch, no matter how high it is and seemingly small: it is *identical* to the entire tree.

And here we see the difference from natural languages most clearly: commonly spoken and written languages do *not* grow on self-similar trees—or, if we insist on hanging them from such trees (perish the thought), most branches would be dead.

Indeed, in natural languages many letter combinations are nonwords. Nevertheless, numerous English words are *homographs* (identical spellings) of words in other languages. And I do not mean such trivial cases as the uni(n)formed GENERAL, which means the same "thing" in many idioms. No, the interesting instances are "incognates" (unrelated words) such as the English word STRICKEN, which means *to knit* in German, or FALTER (a German *butterfly*) and LINKS (the German *left*). And what about such triplets as ART, which is a German word for KIND, which may mean MINOR in German, which in turn is a technical term in the theory of determinants (in either language). Finally, a fivefold string: ROT-RED-TALK-STEATITE-SOAPSTONES. Who can conceive sextuplets?

There are "literally" hundreds of Anglo-German words like that, and I once composed a (short) German story using only English words. When I showed this story to a German-speaking Hungarian in the United States, his bored comment was "nothing but random poetry"—even after repeated proddings to

look at the text with an open mind, as in visual texture (figure-background) discrimination, one of his research interests. When, half a year later, I showed the same text once more to the same Hungarian friend, this time in Germany, he read it and commented “Interessant! Interessant!!” Talk about the impact of context in human perception! (I leave it as an exercise to the linguistically inclined reader to compose a novel that makes *sense* in both German and English, or any other pair of languages in which at least the letter frequencies are not too different.)

How about the French woman who was amazed at the quantities of “soiled underwear” offered for sale in the United States when she first came upon the common come-on *Lingerie Sale*?

Sometimes a double-duty word engenders a double entendre, or rather a twofold misunderstanding. Shortly after I moved to Göttingen, the building superintendent of the physics institute, who collected my foreign parcel post from customs, went around the campus confiding that “Professor Schroeder is importing *poison* from the U.S.A. It even says so right on the packages: Gift!” Gift indeed, the German word for poison, and cognate to the English gift, because *gift* is something one gives (occasionally, anyhow), as in the surviving *Mitgift*, the bride’s dowry.

When I told this tale to the (research) chemist Francis O. Schmitt of the Massachusetts Institute of Technology, he parried with the perfect misunderstanding in reverse. One of his students had once reported from a postdoctoral stay in Germany how generous indeed the indigenous chemical industry was: every other bottle in his lab was labeled *GIFT!* So, in certain parts of the world, better not to swallow the “presents.”

Of course, not all homographs are quite so harmless. Consider *Not*, the German *emergency*. An Australian friend of mine (a linguist, no less) once found himself trapped inside a building in Austria (was the place on fire?), but every door that he approached repulsed him with a forbidding “verboten” sign saying NOTAUSGANG!—not exit? My increasingly frantic friend, desperately seeking *Ausgangs*, knew enough Latin and German (besides his native English) to properly decode *aus-gang* as *ex-it*. But in the heat of the emergency, he never succeeded in severing the Gordian knot: *Not* is not *not*.

## Newton’s Iteration and How to Abolish

### Two-Nation Boundaries

As every pupil learns, the equation  $z^2 = 1$  has not one but two solutions:  $z = +1$  and  $z = -1$ . But suppose we did not know this; we could then start with some initial guess  $z_0$  and use Newton’s *tangent method* of finding a “closer” approximation  $z_1$ . In our case, Newton’s method gives  $z_1 = (z_0^2 + 1)/2z_0$ . For positive  $z_0$ , the approximation  $z_1$  will lie closer to the solution  $+1$ . For example, for  $z_0 = 0.5$  we get  $z_1 = 1.25$ . In fact, all  $z_0$  whose real part is larger than zero, upon

repeated application of the formula, "migrate" toward  $+1$ . Similarly,  $z_0$  with negative real parts will converge on  $-1$ . Thus, the line in the complex number plane for which the real part of  $z_0$  vanishes (i.e., the imaginary axis) is the boundary between the two *basins of attraction* of the two solutions  $+1$  and  $-1$ , respectively. Easy as pie.

What about  $z^3 = 1$ ? It has, of course, three solutions:  $z = 1$ ,  $z = \omega$ , and  $z = \omega^2$ , where  $\omega$  is the standard abbreviation for  $\exp(i2\pi/3)$ . Starting again with an initial guess, Newton's method now gives  $z_1 = (2z_0^3 + 1)/3z_0^2$  for the next approximation. Iterating this formula, we expect to converge on one of the three solutions ( $1$ ,  $\omega$ , or  $\omega^2$ ), depending on the sector in which the initial value  $z_0$  is located. In other words, we expect the three basins of attraction to partition the complex plane into three  $120^\circ$  pie-shaped pieces. But nothing could be further from the truth, as the English mathematician Arthur Cayley (1821–1895) first noted with utter surprise in 1879. (We shall encounter Cayley again when we consider self-similar trees.)

The real behavior of the harmless looking iteration  $z_{n+1} = R(z_n) = (2z_n^3 + 1)/3z_n^2$  is complex almost beyond belief. For one, there are no pie-shaped pieces for the basins of attraction of  $1$ ,  $\omega$ , and  $\omega^2$ . In fact, there is, in the entire complex plane, not a single connected piece of boundary between two basins. Suppose we have a point  $z_0$  that, upon iteration, converges on  $+1$ , and suppose further that we have another point nearby that converges on  $\omega$ ; then there is always a third point, even nearer to  $z_0$ , that iterates toward the third solution,  $\omega^2$ . It is as if international jealousy (or prudence) abhorred two-nation boundaries and a third country *always* interposed itself between two others.

This kind of incredible behavior, and of such a simple equation at that, has stunned not only mathematical laity but many a hard-boiled professional too, until, from 1918 on, Gaston Julia (1893–1978) and Pierre Fatou (1878–1929) showed that, for iterations of rational functions in general, the boundary points of one basin of attraction are the boundary points of *all* basins. These boundary points form a set that is now called a *Julia set* in Gaston's honor (the complementary set of complex numbers is appropriately called a *Fatou set*). Thus, iterations that have more than two basins of attraction cannot have basin boundaries that are simple connected line segments. Such boundaries must, per force, be fractals consisting of totally disconnected point sets—an infinitely fine sprinkling of uncountable numerical "dust".

Color Plate 3 shows, in red, green, and blue, the three basins of attraction of  $1$ ,  $\omega$ , and  $\omega^2$ , respectively, in the complex Gaussian plane. In the center of the figure ( $z = 0$ ), we see a kind of cloverleaf where the three basins (each represented twice) meet in a single point. The central cloverleaf has three preimages, again cloverleafs, albeit somewhat distorted. These three cloverleafs have nine even smaller cloverleafs as preimages, and so on *ad infinitum*, in a beautiful display of self-similarity. It is in this manner that all boundary points become boundary points of all three attractors, precisely as the point  $z = 0$ . In fact, the Julia set is the set of preimages of  $z = 0$ . But the true dustiness of the set can

never be shown with man-made machinery of finite resolution. In fact, a Julia set that is not the entire complex plane has *no* interior points. So it's all or (almost) nothing for Julia sets.

Instead of attractors, we can define a Julia set also in terms of repellors. Indeed, the Julia set  $J_R$  of a rational function  $R$  comprises all of its (uncountably many) repellors. This makes intuitive sense because  $J_R$  is the *boundary* of  $R$ 's basins of attraction, but does not belong to the attractive basins themselves. However, the fact that the forward orbit of any repellor should "visit" *all* other repellors is a bit surprising.

Interestingly, not only the forward orbit of a repellor, but its backward orbit too (generated by the inverse map), is dense in  $J_R$ . Since repellors become attractors for the inverse map, Julia sets can be computationally constructed in a stable, albeit nonuniform, manner from a single repellor subjected to the inverse map; small errors in the computation will not explode, as they would for the forward map. All this is beautifully explained in *The Science of Fractal Images* by Peitgen and Saupe [PS 88].

We shall come upon Julia sets again on pages 243–248 in Chapter 11, where we will deepen our knowledge of these fascinating and often fractal sets.

## Could Minkowski Hear the Shape of a Drum?

When, in late 1910, the great Dutch physicist Hendrik A. Lorentz delivered the Wolfskehl lectures<sup>8</sup> at Göttingen, he threw in a conjecture that Hilbert (his host) immediately predicted to be unprovable in his lifetime. Lorentz's conjecture, which is important in thermodynamics (for calculating the specific heat of solids), blackbody radiation, and concert hall acoustics, says that the number of resonances  $N_3(f)$ , up to some large frequency  $f$ , depends only on the *volume*  $V$  of the resonator and not on its shape.

Someone in the audience by the name of Hermann Weyl (who later succeeded Hilbert in Göttingen) didn't share the great man's pessimism. In fact, within a short while, Weyl succeeded in proving that, asymptotically, for large  $f$  and for resonators with sufficiently smooth but otherwise arbitrary boundaries.

$$N_3(f) = \frac{4\pi}{3} V \left(\frac{f}{c}\right)^3$$

---

8. Paid for from the proceeds of the (still unclaimed) Wolfskehl Prize, administered by the Göttingen Academy of Sciences and to be awarded for the settlement (one way or another) of Fermat's last theorem. The original amount of the prize was 100,000 gold marks, but inflation, engendered by two world wars, reduced this to 7600 deutsche marks.

where  $c$  is the velocity of sound (or light, in the case of blackbody radiation). The corresponding formula for two-dimensional resonators (think of drums or surface waves on a lake) is

$$N_2(f) = \pi A \left( \frac{f}{c} \right)^2$$

where  $A$  is the surface area of the resonator. The result is asymptotically correct, to order  $f^2$ , again independent of the shape of the boundary (perimeter).

These stunning formulas were later improved by correction terms involving lower powers of  $f$  [HBM 39]. For example, for a given boundary condition, the correction term for  $N_2(f)$  is

$$\Delta N(f) = \frac{1}{2} P \frac{f}{c}$$

where  $P$  is the length of the resonator's perimeter.

What happens if we drop Weyl's smooth-boundary restriction? What if the perimeter is a fractal, with fractal dimension  $D > 1$ ? M. V. Berry surmised [Ber 79] that

$$|\Delta N(f)| = \left( L \frac{f}{c} \right)^D$$

where  $L$  is a length constant and  $D$  is perhaps the Hausdorff dimension of the perimeter. This is a reasonable assumption because the exponent of  $f$  in any of the terms of these formulas, including the correction terms, equals the Euclidean dimension (3, 2, or 1) of the content measure (volume, area, or length) of the resonator. Thus, for a fractal perimeter that has infinite length and fractal dimension  $D$ , the corresponding power of  $f$  might very well be  $f^D$ .

Berry's conjecture that  $D$  was in fact the *Hausdorff* dimension turned out to be wrong in some cases. Rather, as Lapidus and Fleckinger-Pellé have shown, the proper fractal dimension is that of *Minkowski* [LF 88]—another kind of nontrivial dimension, introduced by Hermann Minkowski (1864–1909)<sup>9</sup> for different purposes (and extended to fractals by Bouligand); it does not always coincide with the Hausdorff dimension.

The definition of the Minkowski dimension  $D_M$  for a curve (be it fractal or smooth) is roughly as follows. Let the center of a small circle with radius  $r$  follow the curve to measure the *Minkowski content*, that is, the area  $F(r)$  of the resulting

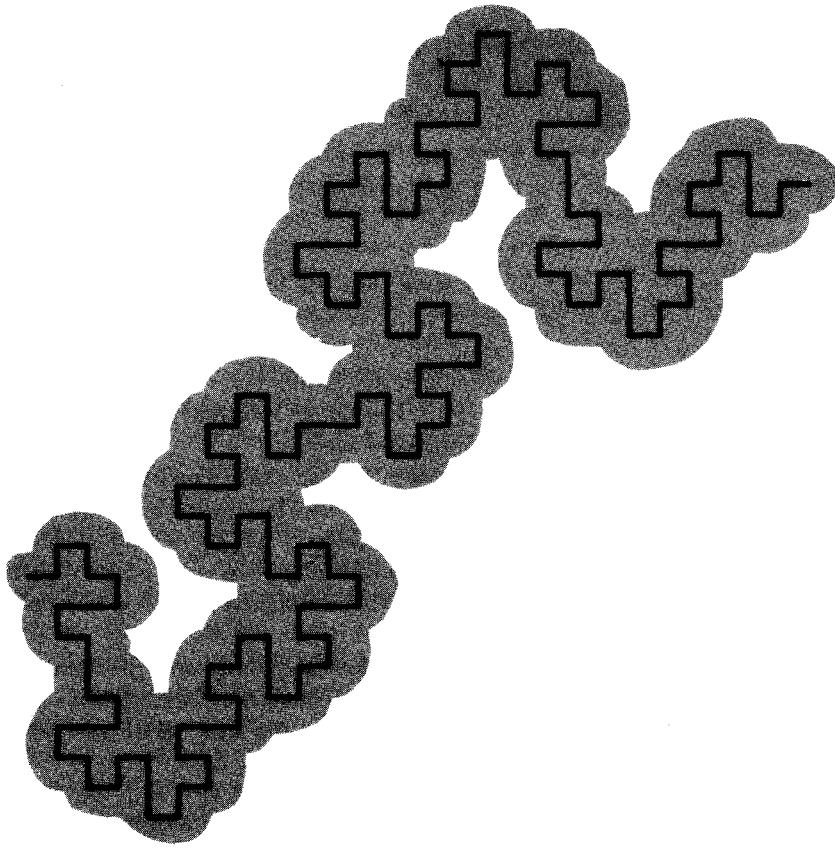
---

9. Minkowski, like Hilbert, was born in Königsberg, where their lifelong friendship began. Minkowski fused geometry and number theory and gave special relativity its proper four-dimensional space suit ("space-time") in preparation for its voyage, under Captain Einstein, into general relativity and modern cosmology.

"Minkowski sausage" traced out by the circle (see Figure 20). Divide the area  $F(r)$  by  $2r$  and let  $r$  go to zero. For a smooth curve, the result will be the length of the curve. But for a fractal "curve," the result may "explode," that is, exceed any finite limit. In fact, the quotient  $F(r)/2r$  will be proportional to  $r^{1-D_M}$ , which—for  $D_M > 1$ —will diverge to infinity for  $r \rightarrow 0$ . The value of  $D_M$  that measures this explosion is defined as the *Minkowski-Bouligand dimension*. Equivalently, we can define  $D_M$  by

$$D_M := \lim_{r \rightarrow 0} \frac{\log F(r)}{\log (1/r)} + 2$$

provided the limit exists. (For some fractals, it is in better taste to distinguish between the two sides of the sausage.) For a smooth curve,  $F(r) \sim r$  and  $D_M = -1 + 2 = 1$ , as expected.



**Figure 20** A "Minkowski sausage" defines the "content" of a curve.

The preceding formula for  $D_M$  is reminiscent of that for the Hausdorff dimension, but note that in place of the number of "covering pieces"  $N(r)$  we have an "area,"  $F(r)$ , the content of the Minkowski sausage. And there is a +2 added to the ratio of the logarithms. (This +2 could be made to disappear, however, by replacing  $F(r)$  by  $F(r)/r^2$  inside the logarithm.)

It is conjectured that for all strictly self-similar fractals the Minkowski dimension is equal to the Hausdorff dimension  $D$ . If they are different,  $D_M$  exceeds  $D$ .

Why is it that the Minkowski, rather than the Hausdorff, dimension controls the number of resonant modes associated with the boundary? Intuitively, the reason is simple. Normal modes need a certain *area* (or volume) associated with the boundary (not the number of covering pieces as for the Hausdorff dimension).

What happens if the resonator domain itself is a fractal, not just its boundary, that is, if the solid isn't solid but has holes on all scales? How does such a fractal "sponge" vibrate? We might conjecture that the foregoing equation for  $N_2(f)$  would then have to be modified to

$$N_2(f) = \left( a \frac{f}{c} \right)^{\bar{d}}$$

where  $\bar{d}$  is an appropriate fractal dimension, called the *spectral dimension*, and  $a$  is some characteristic length. But we have to be careful here, because a fractal with  $D < 2$  embedded in two dimensions is often but a "dust," and how can a *dust* support normal modes? However, as we shall see in the chapter on *percolation* (Chapter 15), there exist, at and above the *percolation threshold*, infinite *connected* clusters of "atoms" that have a finite mass and can support normal modes of vibration. For wavelengths exceeding the typical cluster size, called the *correlation length*, the density of modes is that of a homogeneous body; that is, it is proportional to  $f^{d-1}$ , where  $d$  is the (integer) Euclidean dimension of the space in which the percolating network is embedded. But for wavelengths *below* the correlation length, the normal modes "see" the self-similar fractal structure of the clusters and the mode density exponent drops from  $d - 1$  to  $\bar{d} - 1$ , where the spectral dimension  $\bar{d}$  typically has a fractal value that differs from that of other fractal dimensions (the Hausdorff and Minkowski dimensions, for example).

By analogy to the particles of light—the ubiquitous *photons*—normal modes of vibration, familiar from musical instruments, are commonly called *phonons* when quantized. Phonons are crucial to our understanding of many physical phenomena, including the specific heat of solids and superconductivity, at both low and high temperatures—perhaps even *room* temperature (in Alaska, with windows wide open, no doubt). Phonons live in crystal lattices and feel at home in amorphous substances too. Phonons in fractal media, when they exist, are now often called—what else?—*fractons*. Fractons are believed to play an increasingly important role in our understanding of a vibrant nature.

A related subject is the diffraction of waves from fractal structures ("diffractals"). Since far-field or "Fraunhofer" diffraction is essentially a Fourier transform, the self-similarities (deterministic or statistical) of the scattering fractal must be fully reflected in the diffraction pattern of the incoming radiation, be it electromagnetic, audible, or ultrasound, electrons, neutrons, or neutrinos. (Is neutrino diffraction by the fractal structure of the universe observable?) Clearly, wave diffraction is a sensitive tool not only for classical bodies, but for fractal matter too. Fractal diffraction is also pressed into (military) service to simulate radar clutter with (confusing) detail on many length and size scales.

What happens to the density of normal modes for vibrating fractals whose fractal dimension *exceeds* their Euclidean dimension? Imagine a violin string whose local matter density varies in a Cantor-like way: the middle third of the string, say, has twice the density of the remaining two thirds, which in turn have their central mass densities increased by a factor of 2, and so forth *ad infinitum*. For a "classical" string of length  $L$  with uniform mass density, the number of normal modes is given by

$$N_1(f) = 2L \frac{f}{c}$$

In analogy with Weyl's formulas, we expect the number of modes of the "jazzy" fractal string to vary as

$$N_1(f) = \left( b \frac{f}{c} \right)^{D_M}$$

with  $D_M > 1$ . Here  $b$  is again a characteristic length.

This brings up an interesting and, as it turns out, important question: Can we calculate the (variable) thickness of the string from its resonance frequencies? Such *inverse problems* occur in many guises in many fields. (For example, can we find the location of a tumor inside the brain from the x-ray shadow it casts in different directions? The answer, within limits, is yes—by *computer tomography*.)

For the violin string, unfortunately, the answer is no. However, if we know the resonance frequencies for *two* independent boundary conditions, then we have all the information necessary to calculate the mass distribution of the (lossless) vibrating string.

The solution of the string problem became important at one point in the author's research on basic mechanisms of human speech production (a prerequisite for better-sounding talking computers, without the unfeeling "electronic accent" that can still be heard today when machines "talk"). One would, of course, learn a lot about human speech production if one could deduce the shape of the vocal tract (tongue position, for example) from the recorded speech sounds (which reflect the vocal-tract resonances). Such a capability would also help the deaf and hard-of-hearing as an adjunct to lipreading, because these people could then "see" the (computed) positions of the tongue on a video monitor.

Regrettably, this is difficult for the reasons just stated, namely, that *two* sets of resonances are needed. However, it turns out that determination of the input impedance of the vocal tract, measured at the lips, together with certain other assumptions, permits one to calculate the vocal-tract area function and therefore the motions of the tongue as the subject articulates different speech sounds [Schr 67].

Thus, the question whether we can hear the shape of the vocal tract has to be answered with caution. Yes, we *can* hear it (that is, after all, how we perceive speech)—but the solution is not unique: there are always several different tongue positions that sound alike. This *articulatory ambiguity* is in fact exploited by the ventriloquist, who manages to keep the lips immobile while using other articulators to “take up the slack” from the lips.

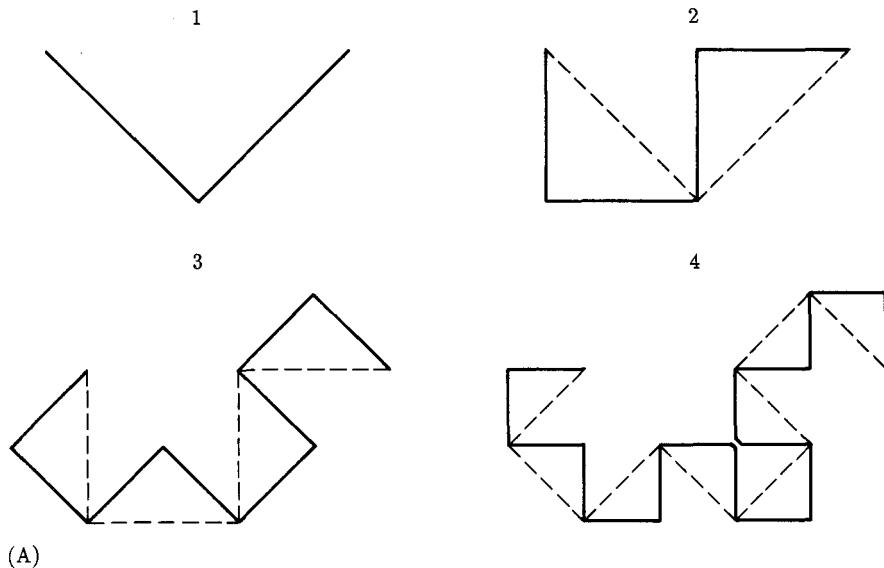
Enough of Lorentz, Hilbert, Weyl, Minkowski—and ventriloquists!

## Discrete Self-Similarity: Creases and Center Folds

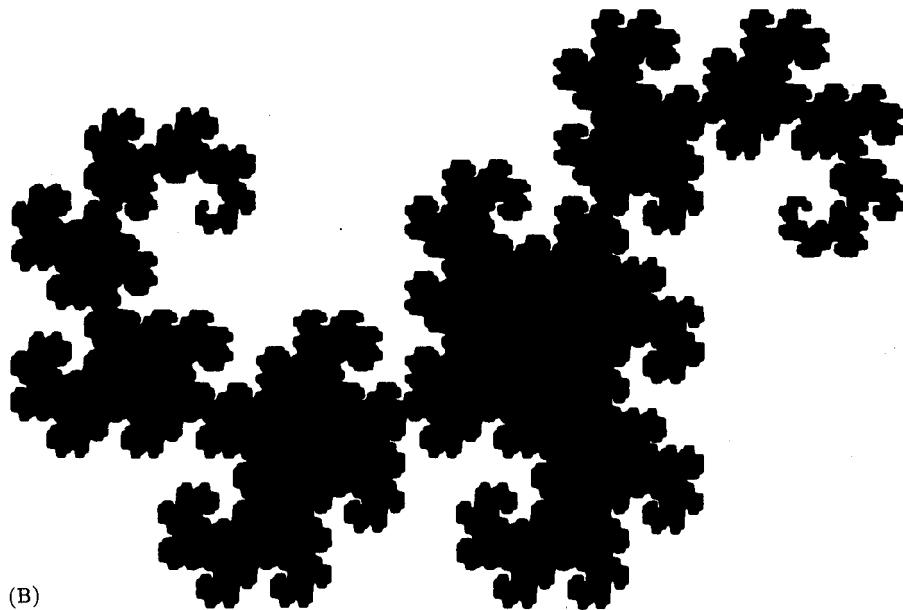
Repetition is a seldom-failing source of self-similarity, beginning with such simple things as paper folding: Take a piece of paper and fold it once. This creates a V-shaped (left-turn) crease (see Figure 21A, generation 1). Fold it over again (parallel to the first fold) and you get three creases, V V Λ: the original center fold V, surrounded by a V on the left and a Λ-shaped fold on the right (Figure 21A, generation 2). Another folding in the same direction yields the crease sequence V V Λ V V Λ Λ (Figure 21A, generation 3). Further folding creates crease sequences of increasing lengths. Each new generation is obtained from the previous one by interpolating alternating V's and Λ's around its letters, beginning with a single V. Thus, the fourth generation reads V V Λ V V Λ Λ V V V Λ Λ V Λ Λ. In an alternative construction, generation  $n + 1$  is obtained from generation  $n$  by copying it, appending a center fold V, and then appending generation  $n$  read backward with V and Λ interchanged [DMP 82]. This operation is equivalent to “pivoting” generation  $n$  around the center fold V (which is what the folding in fact does).

But where is the self-similarity in this crazy succession of creases? Let the untold truth unfold! Pick every other “letter” in V V Λ V V Λ Λ, say, beginning with the second letter (V), an operation appropriately called *unfolding*, and you get the “mother” sequence V V Λ, which, by our alternative construction, must also be the initial part of the daughter sequence V V Λ V V Λ Λ. Thus, the *infinite* folding sequence, obtained in this manner, is precisely self-similar: taking every second (even-numbered) crease recreates the entire sequence. Discrete self-similarity could hardly be simpler.

Can you construct a direct, nonrecursive formula for the  $n$ th letter in the universal crease word? Suppose  $n$  is written as a binary number and the first digit to the left of the first 1 is . . .

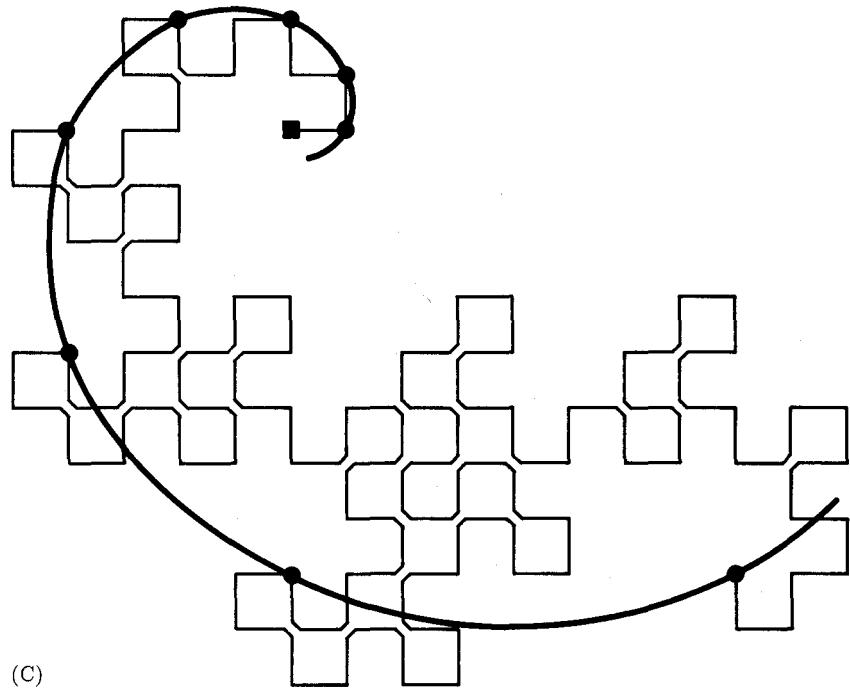


(A)



(B)

**Figure 21** (A) Basic dragon curve, generated by right-angle creases. (B) Self-similarity revealed in later generations of the dragon curve. (C) Center creases (marked by dots) fall on a self-similar logarithmic spiral.



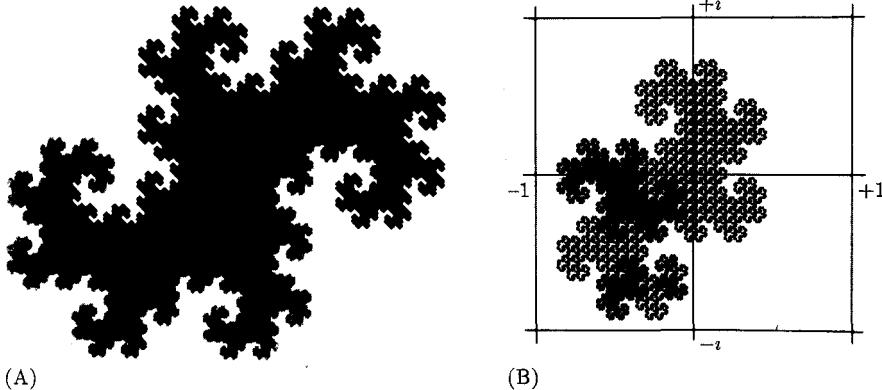
**Figure 21 (continued)**

The self-similarity inherent in the crease sequence can be brought to visual life by making all creases right angles (see Figure 21A). The fractal so generated is known as the *dragon curve*, because later generations (see Figure 21B) resemble a dragon. Two such dragons produce the *twin dragon* [DK 88]. The dragon curve is self-similar (see Figure 21C). The successive center folds, marked by little dots, fall on a logarithmic spiral, one of the basic (and smooth!) self-similar objects, with many interesting applications (see pages 89–92 in Chapter 3).

A *twin dragon* (see Figure 22A) comes alive in a noteworthy number system using a *complex* base. With the advent of digital computers, the binary system, using only the two digits 0 and 1, became the most widely used notation for numbers.<sup>10</sup> Nowadays computers deal a lot with complex numbers, that is, numbers having two “components”: a real part and an imaginary part. Complex numbers thus require *two* sets of binary numbers. It would be nice, of course, if

---

10. True, Claude Shannon once built a computer called THROBAC based on the *Roman* numerals (I, II, III, IV, and so on), but this exercise in masochism somehow did not catch on (in contrast to Shannon’s information theory, which continues to shine brightly).



**Figure 22** (A) Twin dragon. (B) The proper fractions in the binary number system for complex numbers using  $1 - i$  as a base: a mirrored twin dragon. Twin dragons tile the plane [DK 88].

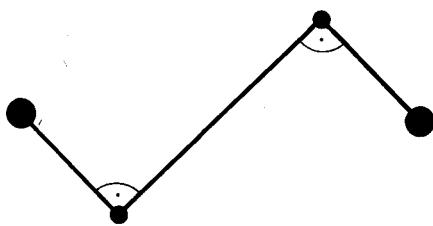
complex numbers, too, could be written as *single* binary numbers, but that seems impossible (eschewing such foul play as interleaving digits).

Yet there exists a complex number system using only the two digits 0 and 1, but its base is not 2. Obviously, the base must be a complex number, and its magnitude must not exceed  $\sqrt{2}$ . Otherwise, two binary digits would have to cover a magnitude range larger than 2. If we further call for evenhandedness between the real and imaginary parts, then a best-base bet would be  $(1 - i) = \sqrt{2} \exp(-i\pi/4)$  (or one of the other three primitive eighth roots of 16).

Of course, we are paying a penalty in “programming” when using this ingenious system. For example, the number 2, which is simply 10 in the real-valued binary systems, is a little more “complex” (in both senses of the word) when using the base  $1 - i$ .

Actually, calling the system based on  $1 - i$  complex is something of an understatement. Figure 22B shows all those numbers in the complex (Gaussian) number plane that are *proper fractions*, that is, numbers in which only *negative* powers of the base appear, such as  $0.1 = (1 - i)^{-1}$  or  $0.\overline{1} = (1 - i)^{-1} + (1 - i)^{-2} + \dots = i$ . (Note that the periodic fraction  $0.\overline{1}$  does *not* equal 1, as it does in the real binary system.)

As is readily apparent from this illustration, the proper fractions occupy a simply connected area with a fractal perimeter, the *twin dragon*. But, in contrast to the Koch flake and the fractal “hexagons” that we encountered in previous sections, the twin dragon’s skin is born of a generator consisting of pieces of different lengths: one piece of length  $r_1 = 1/\sqrt{2}$  and two pieces at right angles of length  $r_2 = r_1^2$  (see Figure 23). The Hausdorff dimension  $D_H$  of such fractals is a straightforward generalization of the formula for  $N$  equal-length pieces, which



**Figure 23** The generator for the twin dragon's skin: Hausdorff dimension 1.52 . . .

can be written  $Nr^{D_H} = 1$ . Replacing the unique length  $r$  of the generator by  $N$  different lengths  $r_1, r_2, \dots, r_N$  results in

$$r_1^{D_H} + r_2^{D_H} + \dots + r_N^{D_H} = 1$$

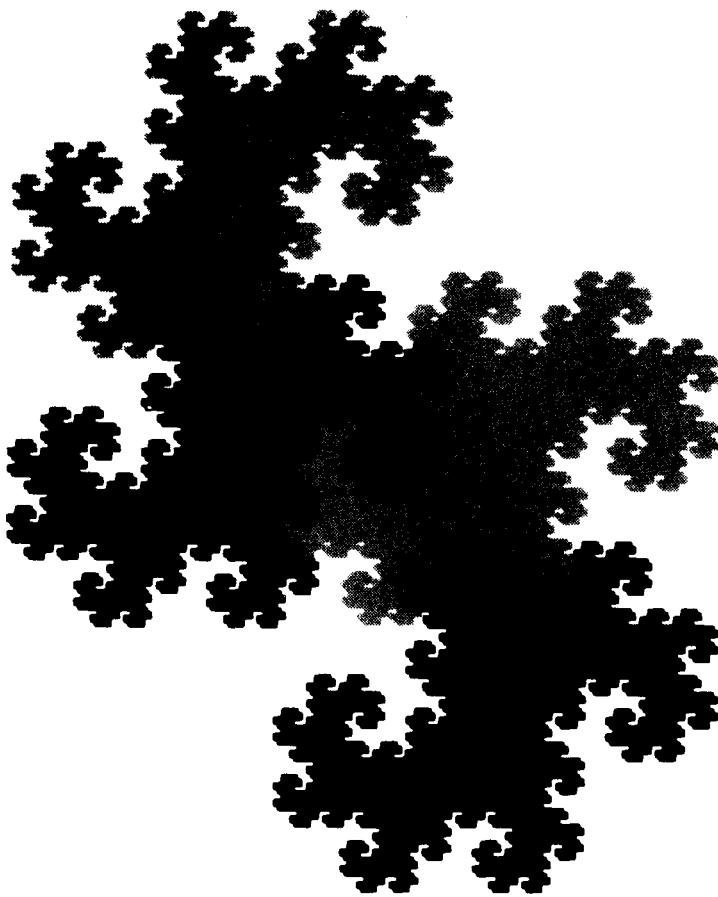
Of course, for  $r_1 = r_2 = r_3 = r_4 = \frac{1}{3}$  we get the old value for the Koch flake:  $4(\frac{1}{3})^{D_H} = 1$  or  $D_H = \log 4/\log 3$ .

To obtain  $D_H$  for the twin-dragon generator, we have to solve a cubic equation in  $r_1^{D_H}$ , namely,  $r_1^{D_H} + 2r_1^{3D_H} = 1$ . Although there are closed formulas involving radicals for cubic equations, I prefer a more conservative approach: my pocket calculator tells me that  $r_1^{D_H} = 0.5897545123 \dots$  and, with  $r_1 = 1/\sqrt{2}$ ,  $D_H = 1.523627 \dots$

The twin dragon can be cut up into four pieces similar to itself (see Figure 24). Thus, according to our generalization of Euclid's scaling theorem for fractal figures (see pages 13–15), the skin of the mother dragon must contain the skin of one of the four child dragons not 2 times but  $2^{D_H} = 2.875 \dots$  times. (Now, of course, one wishes for the radical solution to see what this irrational ratio could possibly mean.)

## Golden and Silver Means and Hyperbolic Chaos

Iteration, as was noted before, is one of the richest sources of self-similarity. Given the proper jump start, the repeated application of some self-same operation, be it geometric, arithmetic, or simply symbolic, leads almost invariably to self-similarity. Take for example the simple rule  $F_{n+2} = F_{n+1} + F_n$ . Starting with  $F_0 = 0$  and  $F_1 = 1$ , this recursion generates the well-known Fibonacci numbers: 0, 1, 1, 2, 3, 5, 8, 13, 21, . . . What is self-similar about them? Multiplying each number by 1.6 and rounding to the nearest integer, we get 0, 2, 2, 3, 5, 8, 13,



**Figure 24** The twin dragon contains four smaller replicas similar to itself, but its fractal skin violates Euclid's scaling law for areas and perimeters of similar figures.

21, 34, . . . —the same sequence, except for a few initial terms (and perhaps later ones).

Taking ratios of successive numbers, we find  $F_n/F_{n+1} = 0, 1, 0.5, 0.\bar{6}, 0.6, 0.625, 0.615 \dots, 0.619 \dots$ —numbers that appear to approach some constant. In fact, a little arithmetic shows that these ratios approach the irrational number  $\tau = (\sqrt{5} - 1)/2 = 0.618 \dots$ , the famous golden mean that tells us how to subdivide a piece of straight line so that the ratio of the shorter segment to the larger equals the ratio of the larger to the whole. Thus, the  $n$ th Fibonacci number should equal, approximately, some constant times  $\tau^{-n}$ . In fact, the approximation is uncannily close: simply divide  $\tau^{-n}$  by  $\sqrt{5}$ , which yields 0.4 . . . , 0.7 . . . ,

1.1 . . . , 1.8 . . . , 3.0 . . . , 4.9 . . . , 8.0 . . . ; and round to the nearest integer and—presto—the Fibonacci numbers, even for  $n = 0$ .

What about the golden mean  $\gamma$  itself? Does it hide any self-similarities? Perhaps they are revealed if we write  $\gamma$  down in the proper number notation: not in “dumb” decimal (0.61803 . . .), not in “bitsy” binary (0.100111 . . .), and not in any other system that elevates some base number beyond its proper station. Rather, let us try a more natural representation, namely, *continued fractions*. Written as a continued fraction,  $\gamma$  becomes  $[1, 1, 1, \dots]$ . In general,  $[a_0, a_1, a_2, \dots]$  is shorthand for

$$\cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \dots}}}$$

which should be banished even as a typist’s punishment.

Figure 25 shows a rendering of the periodic continued fraction for the golden mean  $\gamma$  that is geometrically self-similar. But what is self-similar about  $\gamma$  *arithmetically*?

The continued fraction for a given positive irrational number  $\alpha < 1$  is calculated as follows: set  $x_0 = 1/\alpha$  and apply the iteration  $x_{n+1} = 1/\langle x_n \rangle_1$ , where the pointed brackets with the subscript 1 mean “take the remainder modulo 1” (for example,  $\langle \pi \rangle_1 = 0.14 \dots$ ). Then the continued fraction for  $\alpha$  is  $[[x_0], [x_1], [x_2], [x_3], \dots]$ , where  $\lfloor \cdot \rfloor$  stands for rounding down to the nearest integer and the left-most term is the integer part of  $\alpha$ . Since all terms (except the first) in the continued fraction for the golden mean equal 1, the number  $\gamma$  is a *fixed point* of the iteration  $x_{n+1} = 1/\langle x_n \rangle_1$ , also called the *hyperbolic map*.

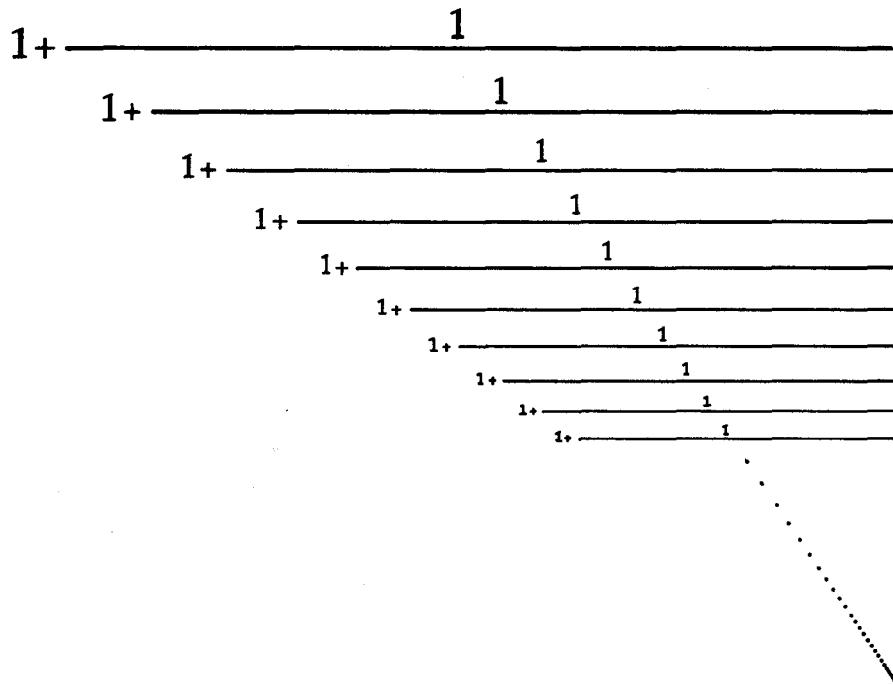
The hyperbolic map is particularly simple to execute if the “condemned” number  $x_0$  is given as a continued fraction: simply move all terms of  $[a_0, a_1, a_2, \dots]$  one place to the left and drop the first term:  $[a_1, a_2, \dots]$ . Thus, the golden mean  $\gamma = [1, 1, 1, \dots]$  is indeed a fixed point of the hyperbolic map.

This map is also called the *Gauss map* because Gauss derived many of its properties, including the invariant distributions of  $x$  and  $a_k$  [Schr 90].

Are there other such precious numbers expressible as periodic continued fractions with periodic length 1? Note that  $1/\gamma = \gamma + 1$ . If we replace the +1 by any other positive integer, we get the *silver means*  $\tau_n$ , defined by  $1/\tau_n = \tau_n + n$ , which have the continued fractions  $\tau_n = [n, n, n, \dots] = [\bar{n}]$ .

The silver means  $[\bar{n}]$  play an enormous role in a sheer, limitless wonderland of applications, encompassing curious quasicrystals, (easy) Ising spins, the mode-locking route to chaos, the “multiplication” of rabbits—and some even curioser games, such as the Fibonacci fleecing, effected by the golden mean.

These numbers, like so many self-similar objects, contain the seeds of chaos. Try iterating the hyperbolic map, starting with a silver mean, on a computer of any finite precision: after a while the result will be utter chaos. As an example,



**Figure 25** Geometrically self-similar continued fraction for the reciprocal golden mean.

take as a starting value  $(\sqrt{13} - 3)/2 = [3, 3, 3, \dots] = 0.3027756 \dots$ , the silver mean  $\tau_3$ . Just keeping the first decimal digit, the hyperbolic map iterated on a small pocket calculator gives 0.3 eight times, followed by 0.2, 0.8, 0.2, 0.6, 0.4, 0.0, 0.2, and so on—a completely unpredictable sequence with a totally chaotic tail.

Similarly, the successive continued fraction obtained from  $\tau_3$  by iterating the hyperbolic map eventually become chaotic: showing only the first term,  $[3, \dots]$  occurs nine times, followed by  $[1, \dots]$   $[4, \dots]$ ,  $[1, \dots]$ ,  $[2, \dots]$ ,  $[10, \dots]$ ,  $[4, \dots]$ ,  $[11, \dots]$ ,  $[90, \dots]$ ,  $[1, \dots]$ , and so forth. Where does this chaos come from? The continued fraction of any irrational number does not terminate; it has infinitely many terms. But finite machines have finite precision, and no matter how high the precision, the less significant terms, starting with some term, must be indeterminate. Thus, in our example,  $\tau_3$  is represented *not* by  $[\overline{3}]$ , where the line over the 3 stands for infinitely many 3s, but by  $[3, 3, 3, 3, 3, 3, 3, 3, 3, 1, 4, 1, 2, 10, 4, 11, 90, 1, \dots]$ . After iterating the hyperbolic map nine times, the random digits have “moved to the front” and take over after that in a characteristic case of chaos.

## Winning at Fibonacci Nim

Let us put the golden mean  $\gamma$  through some of its paces right now. Consider the two integer sequences.

$$a_k = \lfloor k/\gamma \rfloor = 1, 3, 4, 6, 8, \dots$$

$$b_k = \lfloor k/\gamma^2 \rfloor = 2, 5, 7, 10, 13, \dots$$

called a pair of Beatty sequences. Between them, they “exhaust” the positive integers. Note that  $b_k = a_k + k$  and that  $a_k$  is always the smallest positive integer not already used up by  $a_n$  and  $b_n$  for  $n < k$ .

This property of the Beatty sequences just given leads to an interesting game called *Fibonacci nim* (also named Wythoff’s nim): two players alternate taking coins from two piles, always at least one coin, and if a player takes coins from both piles, then he must take the same number of coins from both. The player who takes the last coin(s) wins.

Suppose initially there are 7 and 12 coins in the two piles and I have the first move. To win, I have to leave my opponent a Beatty pair  $(b_k, a_k)$ , which I can always attain from a non-Beatty pair. The Beatty partner of the smaller<sup>11</sup> number ( $7 = b_3$ ) is  $a_3 = 4$ ; thus I take  $12 - 4 = 8$  coins from the larger pile, leaving my opponent the Beatty pair  $(7, 4)$ , which means that he has now lost, because he cannot obtain another Beatty pair.

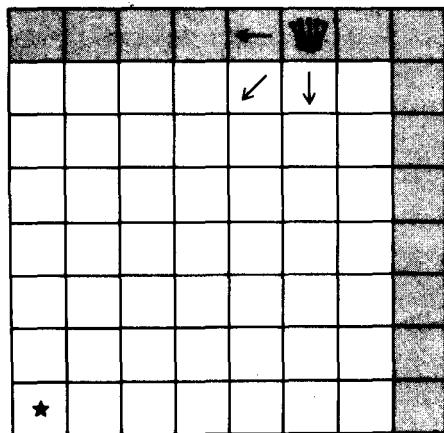
Suppose he leaves me the pair  $(5, 4)$ , which I cannot convert into  $(7, 4)$ . But notice that the difference between 5 and 4 is 1. Thus, by taking 3 coins from each pile, I can realize the Beatty pair with the difference 1, namely,  $(2, 1)$ . I leave it to the reader to convince himself that no matter which of the four possible options my opponent now chooses, I can always take the last coin(s) and thereby win. Once you receive a Beatty pair, you cannot, by yourself, recover from it, and you are beaten.

It is interesting to note that there is a simple board game, called “corner the lady” [Gar 89], that is equivalent (“homomorphic”) to Fibonacci nim. Take a chessboard and place a “queen” anywhere in the top row or the rightmost column, shown in gray in Figure 26A. Two players alternate moving the queen either “west,” “southwest,” or “south.” To which cells should I move the queen so that, no matter what my opponent does, I have the last move to the starred square and win?

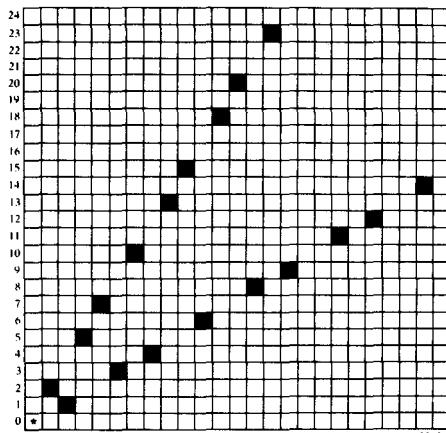
The queen’s moves to the west or south correspond, of course, to taking tokens from either of the two piles in the nim game. And the moves to the southwest correspond to taking on *equal* numbers of tokens from both piles. The

---

11. If both numbers are equal, I take all the coins and win immediately.



(A)



(B)

**Figure 26** (A) The board game “corner the lady” of Rufus P. Isaacs: The initial positions for the queen are shown in gray. The goal is to reach the lower left corner. (B) The safe squares for the lady are shown in dark shading. They are all those squares for which two opposing sides are pierced by one of two straight lines whose slopes equal the golden mean and its reciprocal.

“safe” squares are therefore whose coordinates correspond to our Beatty pairs derived from the golden mean:  $(1, 2), (3, 5), (4, 7), (6, 10), (8, 13)$ , and so on. These safe squares are black in the large board shown in Figure 26B.

Is there a simple, purely *geometric*, method of finding all the safe squares? There certainly is. Draw a straight line from the lower left corner with slope equal to the golden mean (see Figure 26B). The squares whose east *and* west sides are pierced by the straight line are all the lower safe squares. (The upper safe squares are their images mirrored at the  $45^\circ$  diagonal.)

We also remark in passing that Fibonacci’s “multiplying” rabbits have known this strategy since long ago—which is why the nim version of the game is called *Fibonacci nim*.

In his *Liber Abaci*, published in 1202, the Italian mathematician Leonardo da Pisa, better known as Fibonacci (son of Bonacci), considered the question of how many rabbit pairs one would have after  $n$  breeding seasons, starting with a simple immature pair and assuming the following idealized rules for the growth of their numbers [Fib 1202]:

- Rabbits mature in one season after birth.
- Mature rabbit pairs produce one new pair every breeding season.
- Rabbits never die.

It is easy to see that, with these rules, the number of rabbit pairs  $F_n$  in the  $n$ th generation must equal the sum of the number of rabbit pairs in the two preceding generations:  $F_n = F_{n-1} + F_{n-2}$ . Starting with  $F_1 = F_2 = 1$  yields the justly famous sequence of Fibonacci numbers: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, . . . , which appears in innumerable situations.

But the rabbits produce still another number sequence, a binary bit sequence which I have called the *rabbit* sequence [Schr 90]. Consider the two maps  $0 \rightarrow 1$  ("young rabbits grow old") and  $1 \rightarrow 10$  ("old rabbits stay old and beget young ones"). Beginning with a single 0, continued iteration gives 1, 10, 101, 10110, . . . , resulting in the infinite rabbit sequence 1011010110110 . . . . Is this sequence self-similar? Naturally it is. Just underline all the "10" pairs—10 1 10 1 10 1 10 1 10 . . . —and read them as 1s, and read the nonunderlined 1s as 0s, and the infinite rabbit sequence reproduces itself: 10110101 . . . !

Where are all the 1s in the rabbit sequence? They occupy the places numbered 1, 3, 4, 6, 8, 9, 11, 12, . . . , which is the first of our golden-mean Beatty sequences,  $a_k$ . And the 0s are located at places 2, 5, 7, 10, 13, which is our second Beatty sequence,  $b_k$ . So, apparently, the rabbits know the ropes of Fibonacci nim—but they have to go on multiplying. . . .

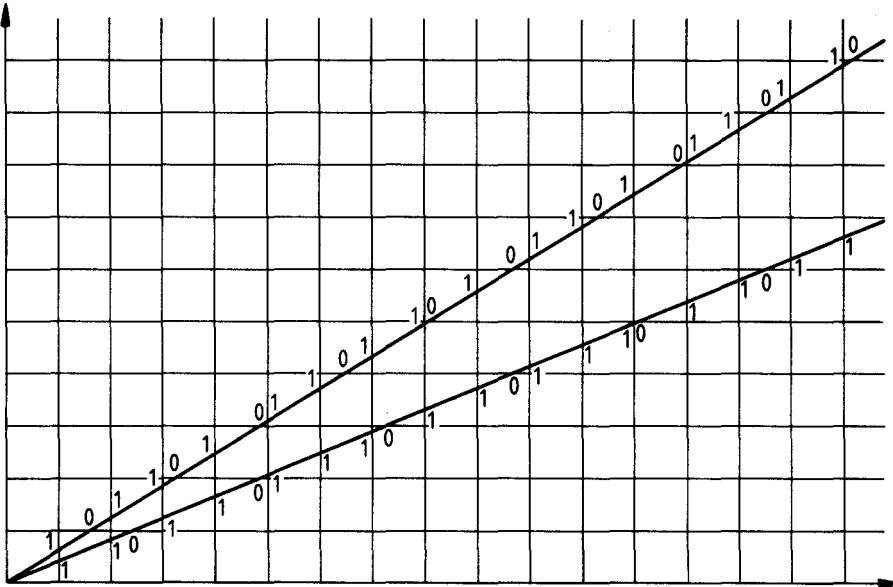
There is still another curious connection between the rabbit sequence and the Fibonacci numbers, discovered by John Horton Conway [Gar 89, p. 21]. The "rabbit constant," defined by the binary fraction .1011010110110 . . . obtained by putting a "decimal" point in front of the rabbit sequence, equals the continued fraction  $[2^0, 2^1, 2^1, 2^2, 2^3, 2^5, 2^8, 2^{13}, 2^{21}, 2^{34}, \dots]$ , where the exponents are none other than the Fibonacci numbers.

We leave it to the reader to concoct similarly mean games based on a silver—rather than the golden—mean. What are the corresponding rabbit multiplication rules for  $\tau_2 = [2] = \sqrt{2 - 1}$ , say?

## Self-Similar Sequences from Square Lattices

Take a square lattice (see Figure 27) and draw a straight cutting line through the origin with a slope equal to the golden mean  $\gamma = [\bar{1}] = 0.618 \dots$ . Next, write a 1 every time the inclined line crosses a vertical lattice line and write a 0 when it crosses a horizontal lattice line. The resulting sequence of 1s and 0s is aperiodic, because the golden mean is irrational, yet it has infinite long-range order because the sequence is deduced from an infinite rigid square lattice.

What other properties are there to the infinite sequence, which begins 1011010110110 . . . ? It is self-similar in the following sense. Consider each 1 the beginning of a new "sentence," and abbreviate the sentence 10 by 1 and the other possible sentence, 1, by 0. The result is the original (infinitely long) "novel": 10110101 . . . . In fact, our novel is not that novel after all: it is the same old-young rabbit sequence we have just encountered in the Fibonacci nim



**Figure 27** Square lattice and straight line with golden-mean slope  $\gamma$  generates the rabbit sequence 10110... The lower straight line has the silver-mean slope  $\sqrt{2} - 1$  and generates another self-similar binary sequence.

game, only here the rabbits emerge from a square lattice. (I am sorely tempted—but will resist for the moment—to call such a lattice *lettuce*.)

If, instead of abbreviating the golden-mean novel, we want to *write* it in the first place, we can start (like most authors) with nothing (0) and iterate the “rabbit” mapping  $0 \rightarrow 1, 1 \rightarrow 10$ .

Conversely, the cutting sequence for  $\tau_2$  is generated by the iterated mapping  $0 \rightarrow 1$  and  $1 \rightarrow 110$ .

For a cutting line with a slope corresponding to the next silver mean,  $\tau_3 = [\bar{3}] = (\sqrt{13} - 3)/2$ , the cutting sequence is 11101110111011110110... Where are the periods ("full stops") between sentences? And how do we write the novel by iteration starting from scratch (0)?

Self-similar sequences, similar to the ones just exhibited, have recently gained notoriety as generators of one-dimensional *quasiperiodic* “lattices” whose generalizations to three dimensions are good mathematical models for a newly discovered state of matter called *quasicrystals* (see Chapter 13).

Sequences, similar to the one discussed here, have also become prominent in computer graphics—specifically, the digitization of straight lines. For *any* slope  $m$  (not just the golden or silver means), the staircase function that best approximates a straight line running through an integer lattice is characterized by a sequence of 0s and 1s, called a *chain code*. Each horizontal step in the staircase is represented by a 1 in the chain code, and each vertical step by 0. Chain codes have the following property: one symbol occurs always in isolation, and the other symbol occurs in runs with at most two different run lengths. If the two run lengths differ, they differ by 1. In fact, for  $m \geq 1$ , the run lengths are  $\lfloor m \rfloor$  and  $\lceil m \rceil$ . (The “gallows”  $\lfloor \cdot \rfloor$  stand for rounding up to the nearest integer.) For  $m < 1$ , the run lengths are  $\lfloor 1/m \rfloor$  and  $\lceil 1/m \rceil$ . The run lengths are different only for noninteger  $m$  or  $1/m$ .

One method of recoding chain codes with different run lengths by two symbols is to assign one symbol to the shorter run length and the other symbol to the longer run length. The result of this recoding is another sequence of symbols of which one occurs in isolation and the other occurs with at most two run lengths that differ by 1. This invariance was discovered by Azriel Rosenfeld in 1973 [RK 82]. But the underlying number-theoretic question, namely, whether a given sequence of integers can be represented by rounding a linear function, was already addressed by one of the Bernoullis [GLL 78].

These results have been applied to efficient image coding and in picture recognition, specifically, for distinguishing straight lines from curved contours [WWM 87]. Present research in this area focuses on computationally efficient algorithms for detecting straight-line segments [KS 87].

A related recoding simply counts the number of places between two successive symbols that occur with different run length. Thus, the chain code 1011010110110 . . . is recoded as 10110101 . . . I conjecture that this recoding of chain codes corresponds to a left shift in the continued fraction representing the slope  $m$ .

## John Horton Conway’s “Death Bet”

John H. Conway, the prolific British mathematician—now serenely ensconced in Princeton, New Jersey—became widely known, even outside mathematics, through his ingenious game called *life* (see Chapter 17, Cellular Automata). During a captivating talk entitled “Some Crazy Sequences” at AT&T Bell Laboratories in Murray Hill, New Jersey, on July 15, 1988, Conway delivered himself of one



**Figure 28** John Horton Conway wearing self-similar horned sphere. (Drawing by Simon Fraser; courtesy of J. H. Conway.)

more direct proof that mathematics is not only great fun but outright funny. After a few preliminary reminiscences about the Fibonacci numbers and the like, he introduced a sequence,  $a(n)$ , that began, harmlessly enough, 1, 1, 2, 2, 3, 4, 4, 5, . . . . The simple iterative law for this sequence is

$$a(n) = a(a(n - 1)) + a(n - a(n - 1))$$

with  $a(1) = a(2) = 1$  for starters. As in the Fibonacci sequence, each new term is the sum of two previous terms.

Numerical evidence suggests that  $a(n)/n$  approaches  $\frac{1}{2}$  as  $n$  becomes large, and Conway challenged the audience to find an  $n_0$  such that for all  $n > n_0$  the absolute error  $|a(n)/n - \frac{1}{2}|$  is smaller than 0.05. Since he and his wife (also a mathematician) found the sequence rather intractable, he offered \$100 for the first solution to reach him. In a barely audible aside (but clearly detectable on the videotape that was made of his talk), he offered a \$10,000 bet to the first finder of the *smallest* such  $n_0$ .

Precisely 34 days later, on August 18, 1988, Colin Mallows, a most capable colleague at Bell, presented the solution, including a formal proof, to Conway's \$10,000 question:  $n_0 = 317\,337\,5556$ . I have written the solution as a U.S. telephone number, in Indiana, incidentally, because it was suggested that Conway was just kidding, that he knew the solution all along and that—upon dialing (317) 337-5556—his voice would come on the line to reassure the keen caller that he had the right number, all right, but was, unfortunately, a bit too late. (Actually, when dialing  $n_0$ , one gets a recorded message to "try again." Try again? After all the trouble to get the number in the first place!

As one might have guessed, the sequence  $a(n)$ , being generated by a simple recursion, is replete with appealing self-similarities that contain the clue to the problem's solution. These self-similarities were speedily brought to light by Mallows, a statistician and data analyst, employing nothing more sophisticated than straightforward numerical computation and graphic displays.

We leave it to the PC-equipped rapt reader to discover for himself the tip-off self-similarities and other symmetries of Conway's sequence  $a(n)$  or the simpler  $b(n) = 2a(n) - n$  (to take out the trend). What happens to  $b(n)$  for  $n = 2^m$ , and why? And what is the Hausdorff dimension of the fractal function to which a properly normalized  $b(n)$  between  $n$  and  $2n$  converges as  $n$  goes to infinity? Is there a simple *direct* formula for  $b(n)$ ? As a warm-up workout, the reader may want to unravel the run-length law of  $a(n)$ .

Mallows, the grand winner, and Conway later agreed that Conway had meant to offer *one* thousand dollars instead of ten thousand. So Conway sent another check, for the smaller amount, but Mallows kept the original prize check for framing. (Figure 28 shows John Horton Conway wearing a self-similar head-gear, the aptly named *horned sphere*.)