THE FRACTAL GEOMETRY OF NATURE

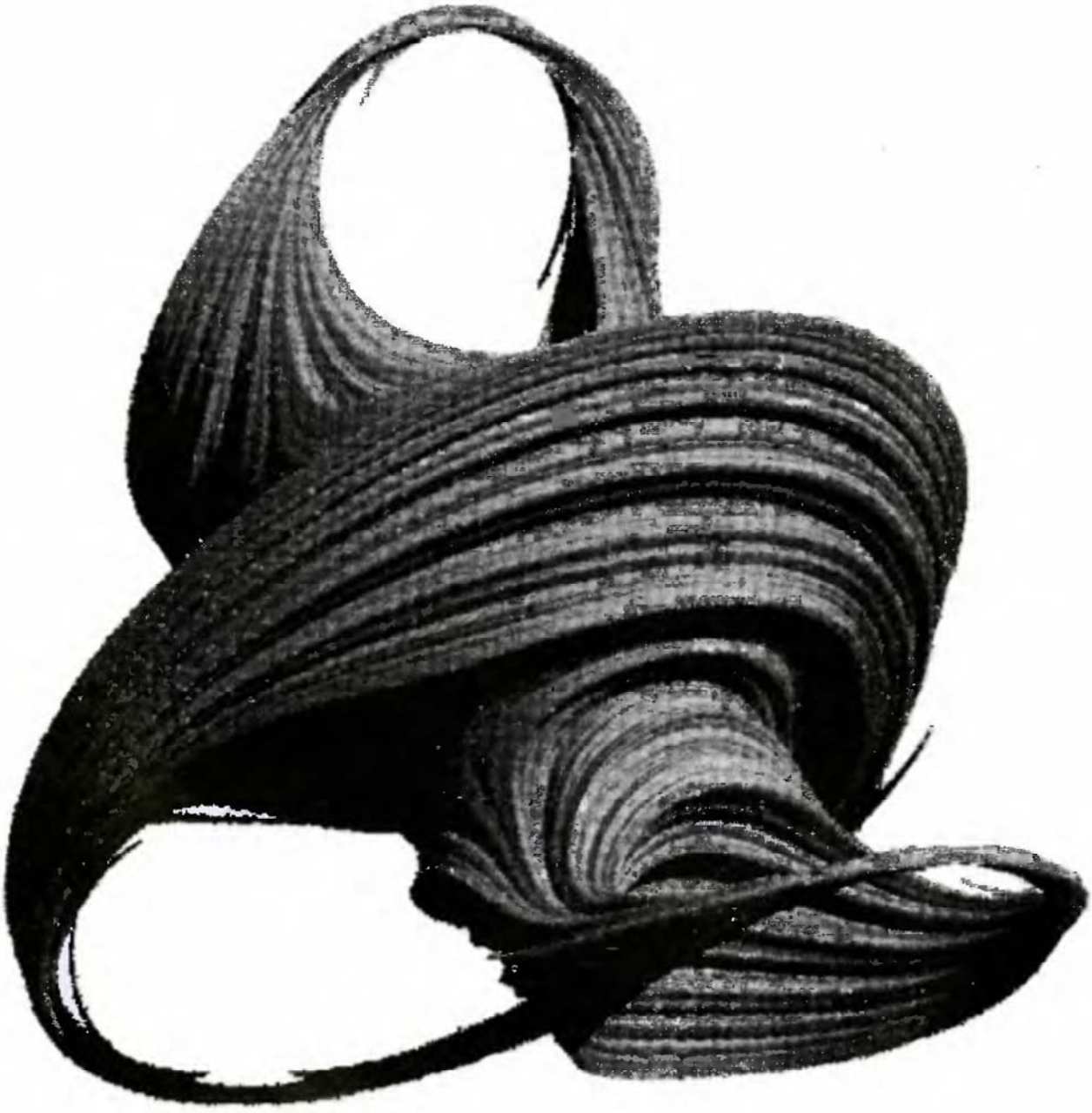


Plate II ¤ ILLUSTRATES PAGE 464

THE FRACTAL GEOMETRY OF NATURE

Updated and Augmented

Benoit B. Mandelbrot

INTERNATIONAL BUSINESS MACHINES
THOMAS J. WATSON RESEARCH CENTER

W. H. FREEMAN AND COMPANY
New York

ABOUT THE AUTHOR ¤¤ IBM Fellow at the IBM Thomas J. Watson Research Center. A Fellow of the American Academy of Arts and Sciences. Graduate, Ecole Polytechnique; M.S. and Ae.E. in Aeronautics, Caltech; Docteur ès Sciences Mathématiques, University of Paris. Dr. Mandelbrot's first positions were with the French Research Council (CNRS), the School of Mathematics at the Institute for Advanced Study (under J. von Neumann), and the University of Geneva. Immediately before joining IBM, he was a junior Professor of Applied Mathematics at the University of Lille and of Mathematical Analysis at Ecole Polytechnique. On leave from IBM, he has been a Visiting Professor of Economics, later of Applied Mathematics, and then of Mathematics at Harvard, of Engineering at Yale, of Physiology at the Albert Einstein College of Medicine, and of Mathematics at the University of Paris-Sud. He has visited M.I.T. several times, first in the Electrical Engineering Department, and most recently as an Institute Lecturer. He was a Visitor at the Institut des Hautes Etudes Scientifiques (Bures). He has been a Fellow of the Guggenheim Foundation, Trumbull Lecturer at Yale, Samuel Wilks Lecturer at Princeton, Abraham Wald Lecturer at Columbia, Goodwin-Richards Lecturer at the University of Pennsylvania, and National Lecturer of Sigma Xi, the Scientific Research Society. Many times a Lecturer at Collège de France since 1973, he delivered there the *leçons* that eventually developed into the present Essay.

AMS Classifications: Primary, 5001, 2875, 60D05;
Secondary, 42A36, 54F45, 60J65, 76F05, 85A35, 86A05.

Current Physics Index Classification (ICSU):
o5.40.+j, 45.30.−b, 98.20.Qw, 91.65.Br.

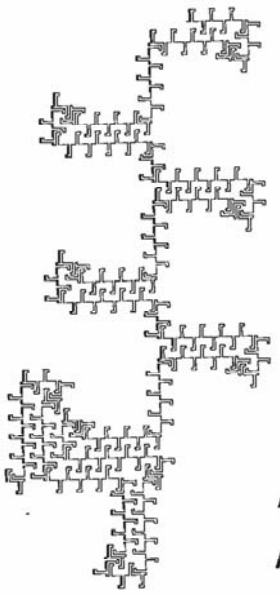
Library of Congress Cataloging in Publication Data

Mandelbrot, Benoit B.
The fractal geometry of nature.

Rev. ed. of: Fractals. c1977.
Bibliography: p.
Includes index.
1. Geometry. 2. Mathematical models. 3. Stochastic processes. I. Title.
QA447.M357 1982 516'.15 81-15085
ISBN 0-7167-1186-9 AACR2

Printed in the United States of America

5 6 7 8 9 KP 1 0 8 9 8 7 6 5 4



In Memoriam, E

Pour Aliette

CONTENTS

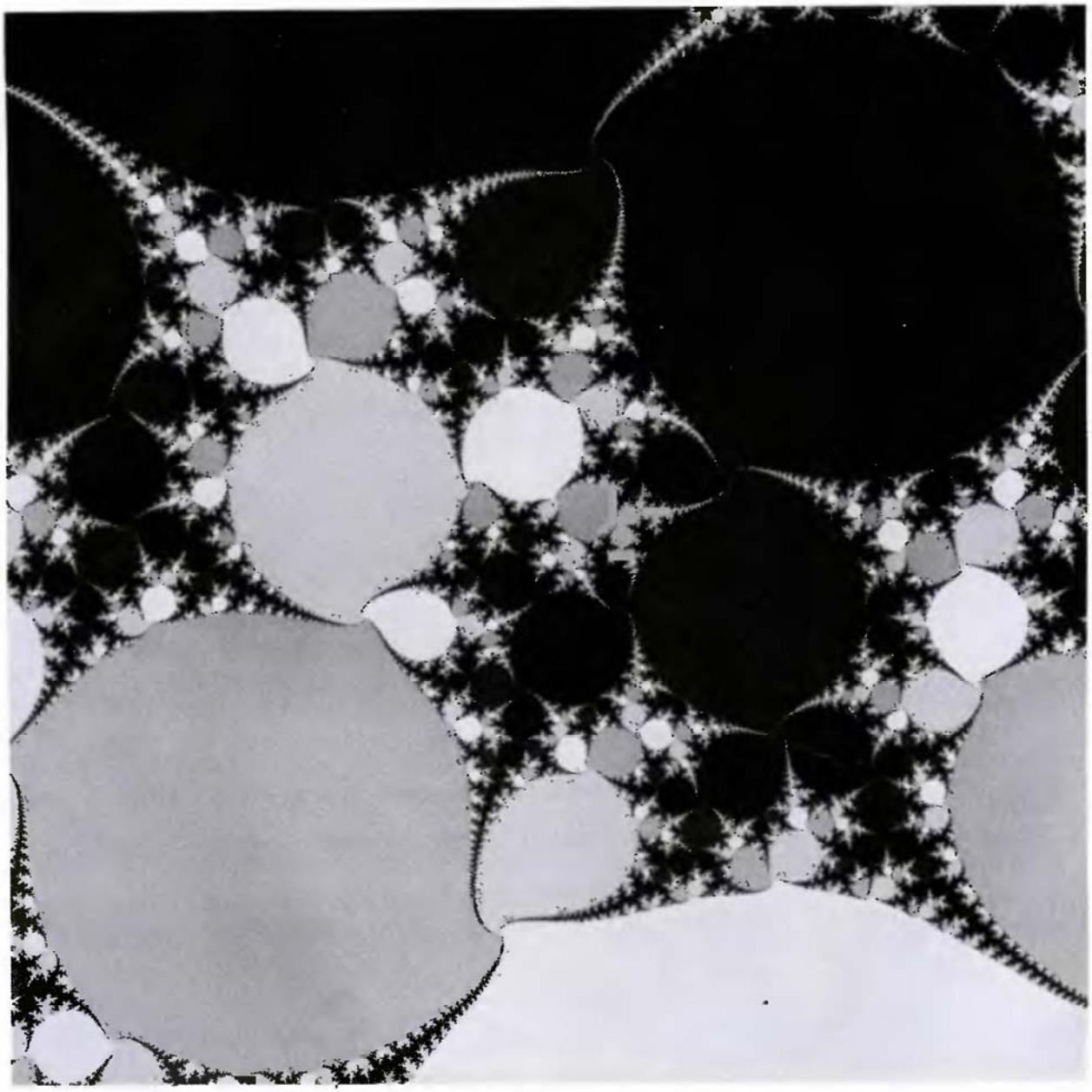


Plate VIII ¤ ILLUSTRATES PAGE 465

FOREWORD

This work follows and largely replaces my 1977 Essay, FRACTALS: FORM, CHANCE AND DIMENSION, which had followed and largely replaced my 1975 Essay in French, LES OBJETS FRACTALS: FORME, HASARD ET DIMENSION. Each stage involved new art, a few deletions, extensive rewriting that affected nearly every section, additions devoted to my older work, and—most important—extensive additions devoted to new developments.

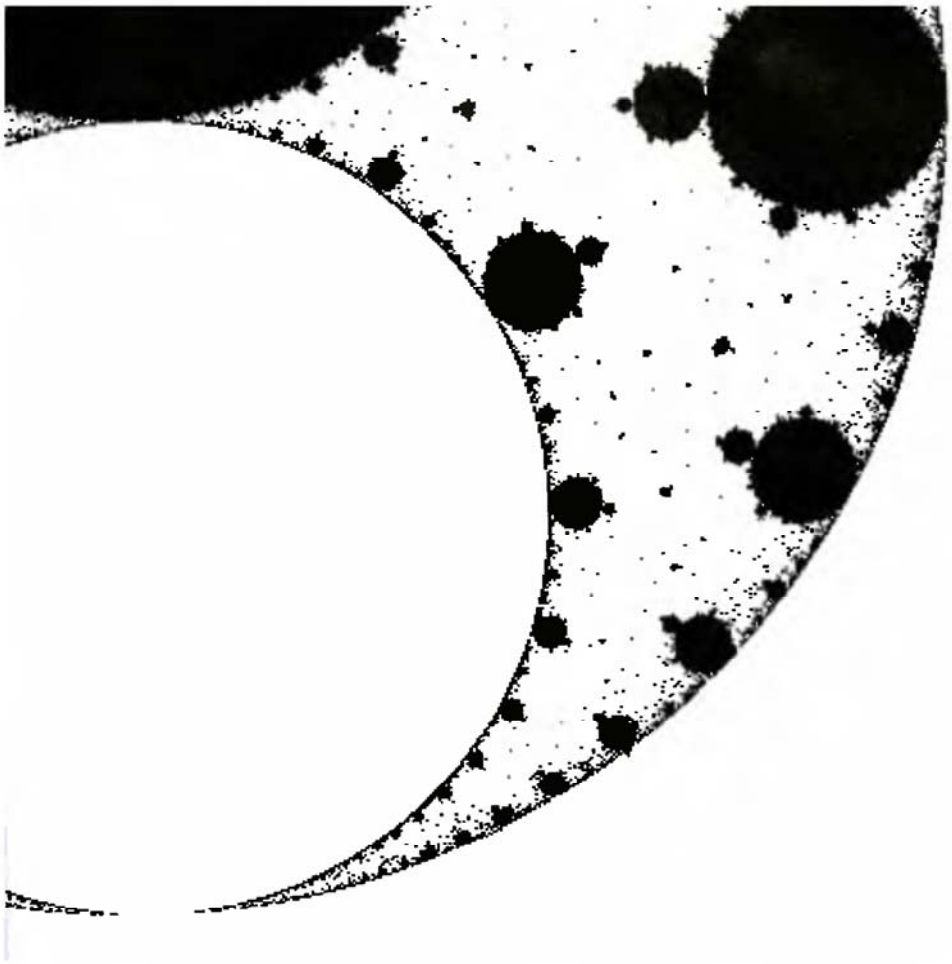
Richard F. Voss made an essential contribution to the 1977 Essay and to this work, especially by designing and now redesigning the fractal flakes, most landscapes, and the planets. The programs for many striking illustrations new to this Essay are by V. Alan Norton.

Other invaluable, close, long-term associates were Sigmund W. Handelman, then Mark R. Laff, for computation and graphics, and H. Catharine Dietrich, then Janis T. Riznychok, for editing and typing.

Individual acknowledgments for the programs behind the illustrations and for other specific assistance are found after the list of references at the end of the volume.

For their backing of my research and my books, I am deeply indebted to the Thomas J. Watson Research Center of the International Business Machines Corporation. As Group Manager, Department Director, and now Director of Research, IBM Vice President Ralph E. Gomory imagined ways of sheltering and underwriting my work when it was a gamble, and now of giving it all the support I could use.

My first scientific publication came out on April 30, 1951. Over the years, it had seemed to many that each of my investigations was aimed in a different direction. But this apparent disorder was misleading: it hid a strong unity of purpose, which the present Essay, like its two predecessors, is intended to reveal. Against odds, most of my works turn out to have been the birth pangs of a new scientific discipline.



THE FRACTAL GEOMETRY OF NATURE

I ¤¤ INTRODUCTION

1 ¤ Theme

Why is geometry often described as "cold" and "dry?" One reason lies in its inability to describe the shape of a cloud, a mountain, a coastline, or a tree. Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.

More generally, I claim that many patterns of Nature are so irregular and fragmented, that, compared with *Euclid*—a term used in this work to denote all of standard geometry—Nature exhibits not simply a higher degree but an altogether different level of complexity. The number of distinct scales of length of natural patterns is for all practical purposes infinite.

The existence of these patterns challenges us to study those forms that Euclid leaves aside as being "formless," to investigate the morphology of the "amorphous." Mathematicians have disdained this challenge, however, and have increasingly chosen to flee from nature by devising theories unrelated to anything we can see or feel.

Responding to this challenge, I conceived and developed a new geometry of nature and implemented its use in a number of diverse fields. It describes many of the irregular and fragmented patterns around us, and leads to full-fledged theories, by identifying a family of shapes I call *fractals*. The most useful fractals involve *chance* and both their regularities and their irregularities are statistical. Also, the shapes described here tend to be *scaling*, implying that the degree of their irregularity and/or fragmentation is identical at all scales. The concept of *fractal* (Hausdorff) *dimension* plays a central role in this work.

Some fractal sets are curves or surfaces, others are disconnected "dusts," and yet others are so oddly shaped that there are no good terms for them in either the sciences or the arts. The reader is urged to sample them now, by browsing through the book's illustrations.

Many of these illustrations are of shapes that had never been considered previously, but others represent known constructs, often for the first time. Indeed, while fractal geometry as such dates from 1975, many of its tools and concepts had been previously developed, for diverse purposes altogether different from mine. Through old stones inserted in the newly built structure, fractal geometry was able to "borrow" exceptionally extensive rigorous foundations, and soon led to many compelling new questions in mathematics.

Nevertheless, this work pursues neither abstraction nor generality for its own sake, and is neither a textbook nor a treatise in mathematics. Despite its length, I describe it as a scientific Essay because it is written from a personal point of view and without attempting completeness. Also, like many Essays, it tends to digressions and interruptions.

This informality should help the reader avoid the portions lying outside his interest or beyond his competence. There are many mathematically "easy" portions throughout, especially toward the very end. *Browse and skip*, at least at first and second reading.

PRESENTATION OF GOALS

This Essay brings together a number of analyses in diverse sciences, and it promotes a new mathematical and philosophical synthesis. Thus, it serves as both a *casebook* and a *manifesto*. Furthermore, it reveals a totally new world of plastic beauty.

A SCIENTIFIC CASEBOOK

Physicians and lawyers use "casebook" to denote a compilation concerning actual cases linked by a common theme. This term has no counterpart in science, and I suggest we appropriate it. The major cases require repeated attention, but less important cases also deserve comment; often, their discussion is shortened by the availability of "precedents."

One case study concerns the widely known application of widely known mathematics to a widely known natural problem: Wiener's geometric model of physical Brownian motion. Surprisingly, we encounter no fresh direct application of Wiener's process, which suggests that, among the phenomena of higher complexity with which we deal, Brownian motion is a special case, an exceptionally simple and unstructured one. Nevertheless, it is included because many useful fractals are careful modifications of Brownian motion.

The other case studies report primarily upon my own work, its pre-fractal antecedents, and its extensions due to scholars who reacted to this Essay's 1975 and 1977 predecessors. Some cases relate to the highly visible worlds of mountains and the like, thus fulfilling at long last the promise of the term *geometry*. But other cases concern submicroscopic assemblies, the prime object of physics.

The substantive topic is occasionally esoteric. In other instances, the topic is a familiar one, but its geometric aspects had not been attacked adequately. One is reminded on this account of Poincaré's remark that there are

questions that one chooses to ask and other questions that ask themselves. And a question that had long asked itself without response tends to be abandoned to children.

Due to this difficulty, my previous Essays stressed relentlessly the fact that the fractal approach is both effective and "natural." Not only should it not be resisted, but one ought to wonder how one could have gone so long without it. Also, in order to avoid needless controversy, these earlier texts minimized the discontinuities between exposition of standard and other published material, exposition with a new twist, and presentation of my own ideas and results. In the present Essay, to the contrary, I am precise in claiming credit.

Most emphatically, I do not consider the fractal point of view as a panacea, and each case analysis should be assessed by the criteria holding in its field, that is, mostly upon the basis of its powers of organization, explanation, and prediction, and not as example of a mathematical structure. Since each case study must be cut short before it becomes truly technical, the reader is referred elsewhere for detailed developments. As a result (to echo d'Arcy Thompson 1917), this Essay is preface from beginning to end. Any specialist who expects more will be disappointed.

A MANIFESTO: THERE IS A FRACTAL FACE TO THE GEOMETRY OF NATURE

Now, the reason for bringing these prefaces together is that each helps one to understand the others because they share a common mathematical structure. F. J. Dyson has given an eloquent summary of this theme of mine.

"*Fràctal* is a word invented by Mandelbrot to bring together under one heading a large class of objects that have [played]...an historical role...in the development of pure mathematics. A great revolution of ideas separates the classical mathematics of the 19th century from the modern mathematics of the 20th. Classical mathematics had its roots in the regular geometric structures of Euclid and the continuously evolving dynamics of Newton. Modern mathematics began with Cantor's set theory and Peano's space-filling curve. Historically, the revolution was forced by the discovery of mathematical structures that did not fit the patterns of Euclid and Newton. These new structures were regarded...as 'pathological,'... as a 'gallery of monsters,' kin to the cubist painting and atonal music that were upsetting established standards of taste in the arts at about the same time. The mathematicians who created the monsters regarded them as important in showing that the world of pure mathematics contains a richness of possibilities going far beyond the simple structures that they saw in Nature. Twentieth-century mathematics flowered in the belief that it had transcended completely the limitations imposed by its natural origins.

"Now, as Mandelbrot points out,...Nature has played a joke on the mathematicians. The 19th-century mathematicians may have been lacking in imagination, but Nature was not. The same pathological structures that the

mathematicians invented to break loose from 19th-century naturalism turn out to be inherent in familiar objects all around us."*

In brief, I have confirmed Blaise Pascal's observation that imagination tires before Nature. ("L'imagination se lassera plutôt de concevoir que la nature de fournir.")

Nevertheless, fractal geometry is *not* a straight "application" of 20th century mathematics. It is a new branch born belatedly of the crisis of mathematics that started when duBois Reymond 1875 first reported on a continuous nondifferentiable function constructed by Weierstrass (Chapters 3, 39, and 41). The crisis lasted approximately to 1925, major actors being Cantor, Peano, Lebesgue, and Hausdorff. These names, and those of Besicovitch, Bolzano, Cesàro, Koch, Osgood, Sierpiński, and Urysohn, are not ordinarily encountered in the empirical study of Nature, but I claim that the impact of the work of these giants far transcends its intended scope.

I show that behind their very wildest creations, and unknown to them and to several generations of followers, lie worlds of interest to all those who celebrate Nature by trying to imitate it.

Once again, we are surprised by what several past occurrences should have led us to expect, that "the language of mathematics reveals itself unreasonably effective in the natural sciences..., a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure even though perhaps also to our bafflement, to wide branches of learning" (Wigner 1960).

*From "Characterizing Irregularity" by Freeman Dyson, *Science*, May 12, 1978, vol. 200, no. 4342, pp. 677–678.

MATHEMATICS, NATURE, ESTHETICS

In addition, fractal geometry reveals that some of the most austerely formal chapters of mathematics had a hidden face: a world of pure plastic beauty unsuspected till now.

"FRACTAL" AND OTHER NEOLOGISMS

There is a saying in Latin that "to name is to know:" *Nomen est numen.* Until I took up their study, the sets alluded to in the preceding sections were not important enough to require a term to denote them. However, as the classical monsters were defanged and harnessed through my efforts, and as many new "monsters" began to arise, the need for a term became increasingly apparent. It became acute when the first predecessor of this Essay had to be given a title.

I coined *fractal* from the Latin adjective *fractus*. The corresponding Latin verb *frangere* means "to break:" to create irregular fragments. It is therefore sensible—and how appropriate for our needs!—that, in addition to "fragmented" (as in *fraction* or *refraction*), *fractus* should also mean "irregular," both meanings being preserved in *fragment*.

The proper pronunciation is *frac'tal*, the stress being placed as in *frac'tion*.

The combination *fractal set* will be defined

rigorously, but the combination *natural fractal* will serve loosely to designate a natural pattern that is usefully representable by a fractal set. For example, Brownian curves are fractal sets, and physical Brownian motion is a natural fractal.

(Since *algebra* derives from the Arabic *jabara* = to bind together, *fractal* and *algebra* are etymological opposites!)

More generally, in my travels through newly opened or newly settled territory, I was often moved to exert the right of naming its landmarks. Usually, to coin a careful neologism seemed better than to add a new wrinkle to an already overused term.

And one must remember that a word's common meaning is often so entrenched, that it is not erased by any amount of redefinition. As Voltaire noted in 1730, "if Newton had not used the word *attraction,* everyone in [the French] Academy would have opened his eyes to the light; but unfortunately he used in London a word to which an idea of ridicule was attached in Paris." And phrases like "the probability distribution of the Schwartz distribution in space relative to the distribution of galaxies" are dreadful.

The terms coined in this Essay avoid this pitfall by tapping underutilized Latin or Greek roots, like *trema,* and the rarely borrowed robust vocabularies of the shop, the home, and the farm. Homely names make the monsters easier to tame! For example, I give technical meanings to *dust*, *curd*, and *whey*. I also advocate *pertiling* for a thorough (= *per*) form of tiling.

RESTATEMENT OF GOALS

In sum, the present Essay describes the solutions I propose to a host of concrete problems, including very old ones, with the help of mathematics that is, in part, likewise very old, but that (aside from applications to Brownian motion) had never been used in this fashion. The cases this mathematics allows us to tackle, and the extensions these cases require, lay the foundation of a new discipline.

Scientists will (I am sure) be surprised and delighted to find that not a few shapes they had to call *grainy, hydralike, in between, pimply, pocky, ramified, seaweedy, strange, tangled, tortuous, wiggly, wispy, wrinkled,* and the like, can henceforth be approached in rigorous and vigorous quantitative fashion.

Mathematicians will (I hope) be surprised and delighted to find that sets thus far reputed exceptional (Carleson 1967) should in a sense be the rule, that constructions deemed pathological should evolve naturally from very concrete problems, and that the study of Nature should help solve old problems and yield so many new ones.

Nevertheless, this Essay avoids all purely technical difficulties. It is addressed primarily to a mixed group of scientists. The presentation of each theme begins with concrete and specific cases. The nature of fractals is meant to be gradually discovered by the reader, not revealed in a flash by the author.

And the art can be enjoyed for itself. ■

2 ¤ The Irregular and Fragmented in Nature

"All pulchritude is relative.... We ought not...to believe that the banks of the ocean are really deformed, because they have not the form of a regular bulwark; nor that the mountains are out of shape, because they are not exact pyramids or cones; nor that the stars are unskillfully placed, because they are not all situated at uniform distance. These are not natural irregularities, but with respect to our fancies only; nor are they incommodious to the true uses of life and the designs of man's being on earth." This opinion of the seventeenth century English scholar Richard Bentley (echoed in the opening words of this Essay) shows that to bring coastline, mountain, and sky patterns together, and to contrast them with Euclid, is an ancient idea.

FROM THE PEN OF JEAN PERRIN

Next we tune to a voice nearer in time and profession. To elaborate upon the irregular or fragmented character of coastlines, Brownian trajectories, and other patterns of Nature to be investigated in this Essay, let me present in free translation some excerpts from Perrin 1906. Jean Perrin's subsequent work on Brownian motion won him the Nobel Prize and spurred the development of probability theory. But here I quote from an early philosophical manifesto. Although it was later paraphrased in the preface to Perrin 1913, this text *failed to gain attention until quoted in this Essay's first (French) version.* It had come to my notice too late to have a substantive effect on my work, but it spurred me on at a time of need, and its eloquence remains unmatched.

"It is well known that, before giving a rigorous definition of continuity, a good teacher shows that beginners already possess the idea which underlies this concept. He draws a well-defined curve and says, holding a ruler, 'You see that there is a tangent at every point.' Or again, in order to impart the notion of the true velocity of a moving object at a point in its trajectory, he says, 'You see, of course, that the mean velocity between two neighboring points does not vary appreciably as these points approach infinitely near to each other.' And many minds, aware that for certain fa-

miliar motions this view appears true enough, do not see that it involves considerable difficulties.

"Mathematicians, however, are well aware that it is childish to try to show by drawing curves that every continuous function has a derivative. Though differentiable functions are the simplest and the easiest to deal with, they are exceptional. Using geometrical language, curves that have no tangents are the rule, and regular curves, such as the circle, are interesting but quite special.

"At first sight the consideration of the general case seems merely an intellectual exercise, ingenious but artificial, the desire for absolute accuracy carried to a ridiculous length. Those who hear of curves without tangents, or of functions without derivatives, often think at first that Nature presents no such complications, nor even suggests them.

"The contrary, however, is true, and the logic of the mathematicians has kept them nearer to reality than the practical representations employed by physicists. This assertion may be illustrated by considering certain experimental data without preconception.

"Consider, for instance, one of the white flakes that are obtained by salting a solution of soap. At a distance its contour may appear sharply defined, but as we draw nearer its sharpness disappears. The eye can no longer draw a tangent at any point. A line that at first sight would seem to be satisfactory appears on close scrutiny to be perpendicular or oblique. The use of a magnifying glass or microscope leaves us just as uncertain, for fresh irregularities appear every time we increase the magnification, and we never succeed in getting a sharp, smooth impression, as given, for example, by a steel ball. So, if we accept the latter as illustrating the classical form of continuity, our flake could just as logically suggest the more general notion of a continuous function without a derivative."

An interruption is necessary to draw attention to Plates 10 and 11.

The black-and-white plates first mentioned in a given chapter are collected on pages that follow immediately, and are numbered as the pages on which they occur. The color plates form a special signature, whose captions are written to be fairly independent of the rest of the book.

The quote resumes.

"We must bear in mind that the uncertainty as to the position of the tangent at a point on the contour is by no means the same as the uncertainty observed on a map of Brittany. Although it would differ according to the map's scale, a tangent can always be found, for a map is a conventional diagram. On the contrary, an essential characteristic of our flake and of the coast is that we *suspect,* without seeing them clearly, that any scale involves details that absolutely prohibit the fixing of a tangent.

"We are still in the realm of experimental reality when we observe under the microscope the Brownian motion agitating a small particle suspended in a fluid [this Essay's Plate 13]. The direction of the straight line joining

the positions occupied at two instants very close in time is found to vary absolutely irregularly as the time between the two instants is decreased. An unprejudiced observer would therefore conclude that he is dealing with a function without derivative, instead of a curve to which a tangent could be drawn.

"It must be borne in mind that, although closer observation of any object generally leads to the discovery of a highly irregular structure, we often can with advantage approximate its properties by continuous functions. Although wood may be indefinitely porous, it is useful to speak of a beam that has been sawed and planed as having a finite area. In other words, at certain scales and for certain methods of investigation, many phenomena may be represented by regular continuous functions, somewhat in the same way that a sheet of tinfoil may be wrapped round a sponge without following accurately the latter's complicated contour.

"If, to go further, we... attribute to matter the *infinitely* granular structure that is in the spirit of atomic theory, our power to apply to reality the *rigorous* mathematical concept of continuity will greatly decrease.

"Consider, for instance, the way in which we define the density of air at a given point and at a given moment. We picture a sphere of volume v centered at that point and including the mass m. The quotient m/v is the mean density within the sphere, and by *true* density we denote some limiting value of this quotient. This notion, however, implies that at the given moment the mean density is practically constant for spheres below a certain volume. This mean density may be notably different for spheres containing 1,000 cubic meters and 1 cubic centimeter respectively, but it is expected to vary only by 1 in 1,000,000 when comparing 1 cubic centimeter to one-thousandth of a cubic millimeter.

"Suppose the volume becomes continually smaller. Instead of becoming less and less important, these fluctuations come to increase. For scales at which the Brownian motion shows great activity, fluctuations may attain 1 part in 1,000, and they become of the order of 1 part in 5 when the radius of the hypothetical spherule becomes of the order of a hundredth of a micron.

"One step further and our spherule becomes of the order of a molecule radius. In a gas, it will generally lie in intermolecular space, where its mean density will henceforth *vanish*. At our point the *true* density will also *vanish*. But about once in a thousand times that point will lie within a molecule, and the mean density will be a thousand times higher than the value we usually take to be the true density of the gas.

"Let our spherule grow steadily smaller. Soon, except under exceptional circumstances, it will become empty and remain so henceforth owing to the emptiness of intra-atomic space; the true density *vanishes* almost everywhere, except at an infinite number of isolated points, where it reaches an infinite value.

"Analogous considerations are applicable to properties such as velocity, pressure, or

temperature. We find them growing more and more irregular as we increase the magnification of our necessarily imperfect image of the universe. The function that represents any physical property will form in intermaterial space a *continuum* with an infinite number of singular points.

"Infinitely discontinuous matter, a continuous ether studded with minute stars, also appears in the cosmic universe. Indeed, the conclusion we have reached above can also be arrived at by imagining a sphere that successively embraces planets, solar system, stars, and nebulae....

"Allow us now a hypothesis that is arbitrary but not self-contradictory. One might encounter instances where using a function without a derivative would be simpler than using one that can be differentiated. When this happens, the mathematical study of irregular continua will prove its practical value."

Then, starting a new section for emphasis. "However, this hope is nothing but a daydream, as yet."

WHEN A "GALLERY OF MONSTERS" BECOMES A MUSEUM OF SCIENCE

Part of this daydream, relative to Brownian motion, did become reality in Perrin's own lifetime. Perrin 1909 chanced to catch the attention of Norbert Wiener (Wiener 1956, pp. 38–39, or 1964, pp. 2–3), who, to his own "surprise and delight" was moved to define and study rigorously a nondifferentiable first model of Brownian motion.

This model remains essential, but physicists stress that its nondifferentiability is traceable to abusive idealization, namely the neglect of inertia. In doing so, physicists turn their back to the feature of Wiener's model that is most significant for the present work.

As to the other applications of mathematics to physics that Perrin foresaw, they were not even attempted until the present work. The collection of sets to which Perrin was alluding (Weierstrass curves, Cantor dusts, and the like) continued to remain a part of "pure mathematics."

Some writers, for example Vilenkin 1965, call this collection a "Mathematical Art Museum," without suspecting (I am sure) how accurate those words were to be proven by the present work. We know from Chapter 1 that other writers (beginning with Henri Poincaré) call it a "Gallery of Monsters," echoing the *Treatise of Algebra* of John Wallis (1685), where the fourth dimension is described as "a Monster in Nature, and less possible than a *Chimera* or *Centaure*."

One of the aims of the present Essay is to show, through relentless hammering at diverse explicit "cases," that the same Gallery may also be visited as a "Museum of Science."

Mathematicians are to be praised for having devised the first of these sets long ago, and scolded for having discouraged us from using them. ■

Plates 10 and 11 ¤ ARTIFICIAL FRACTAL FLAKES

In an inspiring text quoted in Chapter 2, Jean Perrin comments on the form of the "white flakes that are obtained by salting a solution of soap." These illustrations are meant to accompany Perrin's remarks.

One must hasten to state that they are neither photographs nor computer reconstitutions of any real object, be it a soap flake, a rain cloud, a volcanic cloud, a small asteroid, or a piece of virgin copper.

Nor do they claim to result from a theory embodying the diverse aspects of a real flake's formation—chemical, physico-chemical, and hydrodynamical.

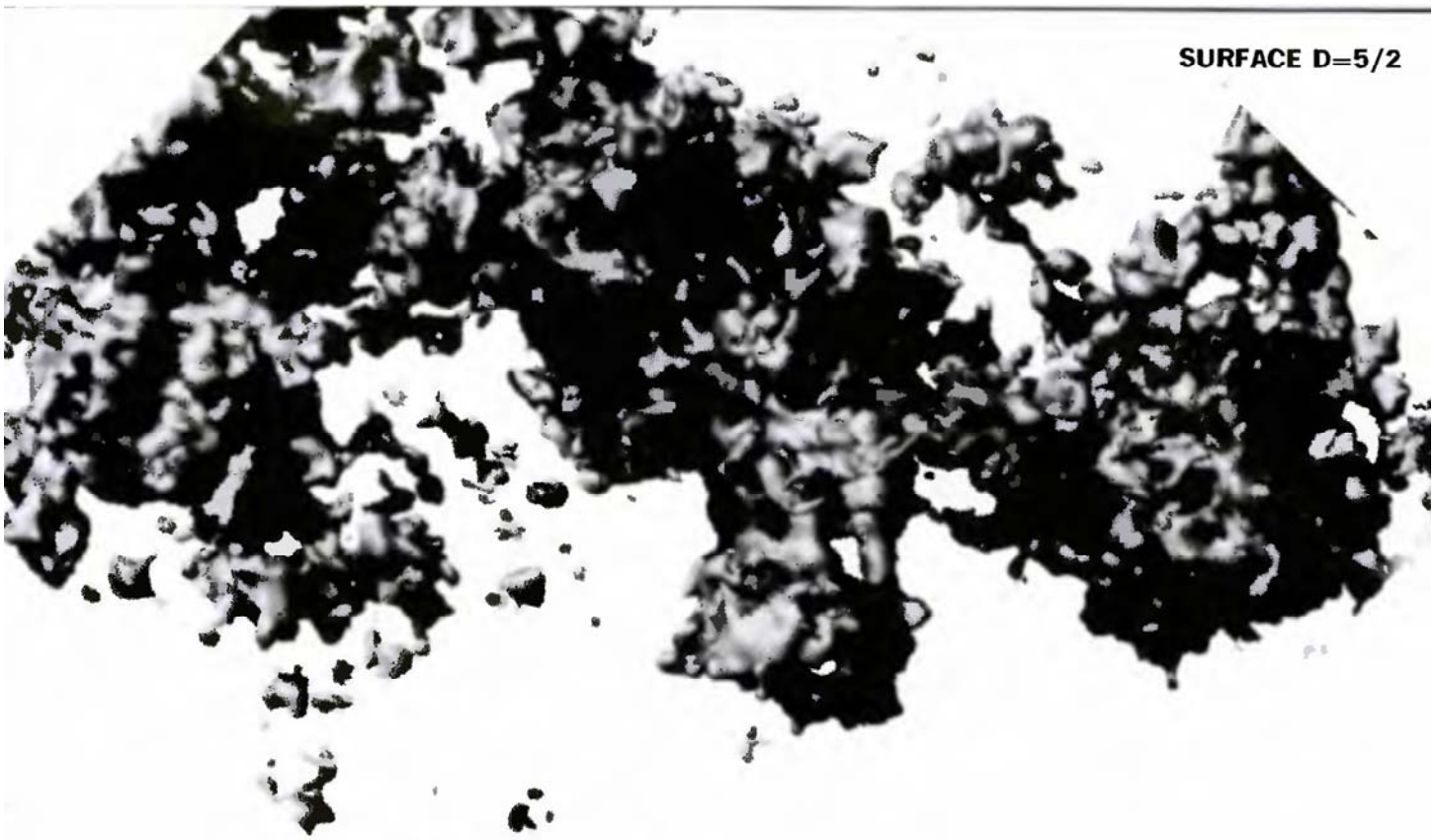
A fortiori, they do not claim to be directly related to scientific principles.

They are computer-generated shapes meant to illustrate as simply as I can manage certain geometric characteristics that seem to be embodied in Perrin's description, and that I propose to model using the notion of fractal.

These flakes exist only in a computer's memory. They were never made into hard models, and the shading too was implemented by computation.

The flakes' construction is explained in Chapter 30. The obvious perceptual differences between them are due to differences in the value of a parameter D written next to each. It is called fractal dimension, is basic to the present work, and is introduced in Chapter 3. The overall shapes being the same in all 3 cases is due to bias introduced by the use of an approximation, and is discussed in the caption of Plates 266 and 267.

An earlier version was oddly reminiscent of a presumed photograph of the Loch Ness monster. Could this convergence of form be coincidental? ■


SURFACE D=5/2


SURFACE D=8/3

Plate 13 ⊔ JEAN PERRIN'S CLASSIC DRAWINGS OF PHYSICAL BROWNIAN MOTION

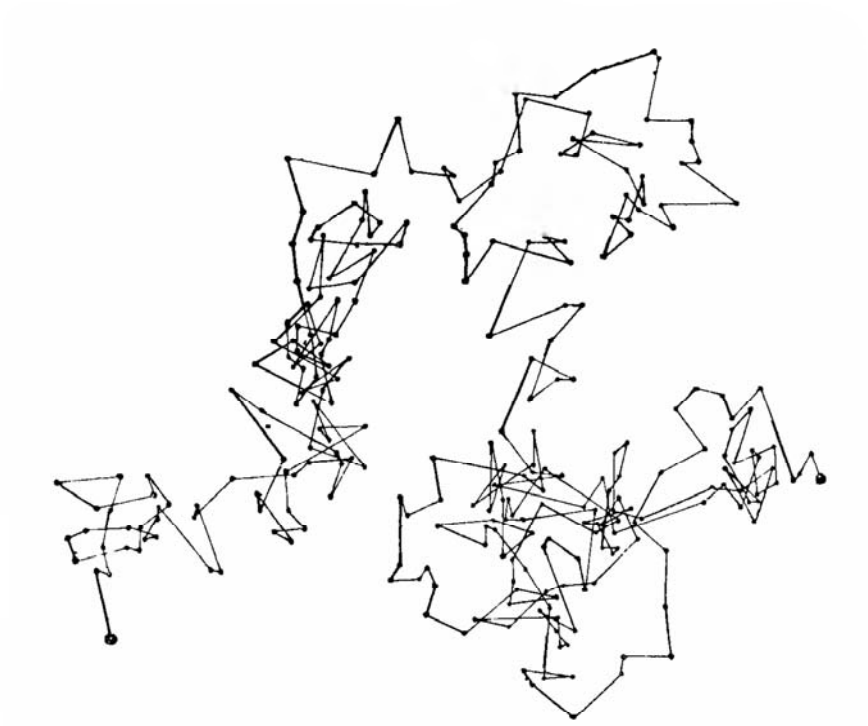
Physical Brownian motion is described in Perrin 1909 as follows: "In a fluid mass in equilibrium, such as water in a glass, all the parts appear completely motionless. If we put into it an object of greater density, it falls. The fall, it is true, is the slower the smaller the object; but a visible object always ends at the bottom of the vessel and does not tend again to rise. However, it would be difficult to examine for long a preparation of very fine particles in a liquid without observing a perfectly irregular motion. They go, stop, start again, *mount*, descend, *mount again*, without in the least tending toward immobility."

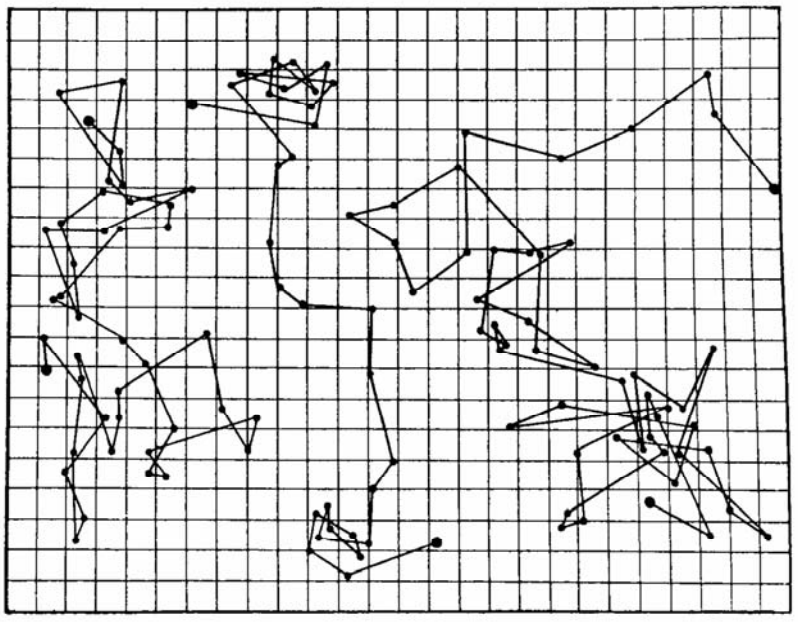
The present plate, the only one in this book to picture a natural phenomenon, is reproduced from Perrin's *Atoms*. We see four separate tracings of the motion of a colloidal particle of radius 0.53μ, as seen under the microscope. The successive positions were marked every 30 seconds (the grid size being 3.2μ), then joined by straight intervals having no physical reality whatsoever.

To resume our free translation from Perrin 1909, "One may be tempted to define an 'average velocity of agitation' by following a particle as accurately as possible. But such evaluations are *grossly wrong*. The apparent average velocity varies crazily in magnitude and direction. This plate gives only a weak idea of the prodigious entanglement of the real trajectory. If indeed this particle's positions were marked down 100 times more frequently, each interval would be replaced by a polygon smaller than the whole drawing but just as complicated, and so on. It is easy to see that in practice the notion of tangent is meaningless for such curves."

This Essay shares Perrin's concern, but attacks irregularity from a different angle. We stress the fact that when a Brownian trajectory is examined increasingly closely, Chapter 25, its length increases without bound.

Furthermore, the trail left behind by Brownian motion ends up by nearly filling the whole plane. Is it not tempting to conclude that in some sense still to be defined, this peculiar curve has the same dimension as the plane? Indeed, it does. A principal aim of this Essay will be to show that the loose notion of dimension splits into several distinct components. The Brownian motion's trail is *topologically* a curve, of dimension 1. However, being practically plane filling, it is *fractally* of dimension 2. The discrepancy between these two values will, in the terminology introduced in this Essay, qualify Brownian motion as being a fractal. ■





3 ¤ Dimension, Symmetry, Divergence

A central role is played in this Essay by the ancient notions of *dimension* (meaning *number of dimensions* or *dimensionality*) and of *symmetry*. Furthermore, we constantly encounter diverse symptoms of *divergence*.

THE IDEA OF DIMENSION

Mathematicians recognized during their 1875-1925 crisis that a proper understanding of irregularity or fragmentation (as of regularity and connectedness) cannot be satisfied with defining dimension as a number of coordinates. The first step of a rigorous analysis is taken by Cantor in his June 20, 1877, letter to Dedekind, the next step by Peano in 1890, and the final steps in the 1920's.

Like all major intellectual developments, the outcome of this story can be interpreted in diverse ways. Anyone who writes a mathematical book on *the* theory of dimension implies that this theory is unique. But to my mind the main fact is that the loose notion of dimension turns out to have many mathematical facets that not only are conceptually distinct but may lead to different numerical values. Just as William of Occam says of entities, dimensions must not be multiplied beyond necessity, but a multiplicity of dimensions is absolutely unavoidable. Euclid is limited to sets for which all the useful dimensions coincide, so that one may call them *dimensionally concordant* sets. On the other hand, the different dimensions of the sets to which the bulk of this Essay is devoted *fail* to coincide; these sets are *dimensionally discordant.*

Moving on from the dimensions of mathematical sets to the "effective" dimensions of the physical objects modeled by these sets, we encounter a different sort of inevitable and concretely essential ambiguity. Both the mathematical and the physical aspects of dimension are previewed in this chapter.

DEFINITION OF THE TERM FRACTAL

The present section uses undefined mathematical terms, but many readers may find it helpful, or at least reassuring, to scan this text, and anybody can skip it.

This and later digressions in this Essay are delimited by the new brackets ◅ and ►. The latter is very bold, so as to be readily found by anyone who becomes lost in a digression and wants to skip ahead. But the "open bracket" symbol avoids attracting attention, so as to prevent digressions from receiving excessive attention. Material discussed later often receives advance mention in digressions.

◅ The fact that the basic fractals are dimensionally discordant can serve to transform the concept of fractal from an intuitive to a mathematical one. I chose to focus on two definitions, each of which assigns to every set $\mathbb{R}^E$ in Euclidean space, no matter how "pathological," a real number which on intuitive and formal grounds strongly deserves to be called its dimension. The more intuitive of the two is the topological dimension according to Brouwer, Lebesgue, Menger, and Urysohn. We denote it by D_T. It is described in an entry in Chapter 41. The second dimension was formulated in Hausdorff 1919 and put in final form by Besicovitch. It is discussed in Chapter 39. We denote it by D.

◅ Whenever (as is usually the case) we work in the Euclidean span $\mathbb{R}^E$, both D_T and D are at least 0 and at most E. But the resemblance ends here. The dimension D_T is *always an integer*, but D *need not be an integer*. And the two dimensions need not coincide; they only satisfy the Szpilrajn inequality (Hurewicz & Wallman 1941, Chapter 4)

$$D \geq D_T.$$

For all of Euclid, $D=D_T$. But nearly all sets in this Essay satisfy $D>D_T$. There was no term to denote such sets, which led me to coin the term *fractal*, and to define it as follows:

◅ *A fractal is by definition a set for which the Hausdorff Besicovitch dimension strictly exceeds the topological dimension.*

◅ Every set with a noninteger D is a fractal. For example, the original Cantor set is a fractal because we see in Chapter 8 that

$$D=\log 2/\log 3 \sim 0.6309 > 0, \text{ while } D_T=0.$$

And a Cantor set in $\mathbb{R}^E$ can be tailored and generalized so that $D_T=0$, while D takes on any desired value between 0 and E (included).

◅ Furthermore, the original Koch curve is a fractal because we see in Chapter 6 that

$$D=\log 4/\log 3 \sim 1.2618 > 1, \text{ while } D_T=1.$$

◅ However, a fractal may have an integer D. For example, Chapter 25 shows that the trail of Brownian motion is a fractal because

$$D=2, \text{ while } D_T=1.$$

◅ The striking fact that D need not be an integer deserves a terminological aside. If one uses *fraction* broadly, as synonymous with a noninteger real number, several of the above-listed values of D are fractional, and indeed the Hausdorff Besicovitch dimension is often called *fractional dimension*. But D may be an integer (not greater than E but strictly greater than D_T). I call D a *fractal dimension*. ►

FRACTALS IN HARMONIC ANALYSIS

◅ Part of the study of fractals is the geometric face of harmonic analysis, but this fact is not stressed in the present work. Harmonic (= spectral or Fourier) analysis is not known to most readers, and many who use it effectively are not acquainted with its basic structures.

Also, both the fractal and the spectral approach have their own strong flavor and personality, which are better appreciated by first investigating each for its own sake. Finally, compared to harmonic analysis, the study of fractals is easy and intuitive. ►

OF "NOTIONS THAT ARE NEW,... BUT"

Lebesgue made fun of certain "notions that are new, to be sure, but of which no use can be made after they have been defined." This comment never applied to D, but the use of D remained concentrated in few areas, all of them in pure mathematics. I was the first to use D successfully in the description of Nature. And one of the central goals of this work is to establish D in a central position in empirical science, thereby showing it to be of far broader import than anyone imagined.

Several areas of physics accepted my claim concerning D with exceptional promptness. In fact, having recognized the inadequacies of standard dimension, numerous scholars in these areas had already been groping towards *broken, anomalous* or *continuous* dimensions of all kind. These approaches had remained mutually unrelated, however. Furthermore, few definitions of dimension were used more than once, none had the backing of a mathematical theory, and none was developed far enough for the lack of mathematical backing to make a difference. For the developments to be described here, to the contrary, the existence of a mathematical theory is vital.

A MATHEMATICAL STUDY OF FORM MUST GO BEYOND TOPOLOGY

A mathematician, if asked which well-defined branch of mathematics studies form, is very likely to mention topology. This field is important to our purposes and is referred to in the preceding section, but the present Essay advances and defends the claim that the loose notion of form possesses mathematical aspects other than topological ones.

Topology, which used to be called *geometry of situation* or *analysis situs* (Τοπος means position, situation in Greek), considers that all pots with two handles are of the same form because, if both are infinitely flexible and compressible, they can be molded into any other continuously, without tearing any new opening or closing up any old one. It also teaches that all single island coastlines are of the same form, because they are topologically identical to a circle. And that the topological dimension is the same for coastlines and circles: equal to 1. If one adds offshore "satellite islands," the cumulative coastline is topologically identical to "many" circles. Thus, topol-

ogy *fails* to discriminate between different coastlines.

By way of contrast, Chapter 5 shows that different coastlines tend to have different fractal dimensions. Differences in fractal dimension express differences in a *nontopological aspect of form,* which I propose to call *fractal form.*

Most problems of real interest combine fractal and topological features in increasingly subtle fashion.

Observe that in the case of topology, the definitions of the field itself and of D_T were refined in parallel, while the notion of D preceded the present study of fractal form by half a century.

Incidentally, Felix Hausdorff's name being given to a class of topological spaces, the widely used contracted term for D, Hausdorff dimension, seems to have undertones of "dimension of a Hausdorff space," thus suggesting it is a topological concept—which emphatically is *not* the case. This is yet another reason for preferring *fractal dimension.*

EFFECTIVE DIMENSION

In addition to the mathematical notions underlying D_T and D, this Essay often invokes *effective dimension,* a notion that *should not* be defined precisely. It is an intuitive and potent throwback to the Pythagoreans' archaic Greek geometry. A novelty of this Essay is that it allows the value of effective dimension to be a fraction.

Effective dimension concerns the relation between mathematical sets and natural objects. Strictly speaking, physical objects such as a veil, a thread, or a tiny ball should all be represented by three-dimensional shapes. However, physicists prefer to think of a veil, a thread, or a ball—if they are fine enough—as being "in effect" of dimensions 2, 1, and 0, respectively. For example, to describe a thread, the theories relating to sets of dimension 1 or 3 must be modified by corrective terms. And the better geometrical model is determined after the fact, as involving the smaller corrections. If luck holds, this model continues to be helpful even when corrections are omitted. In other words, effective dimension inevitably has a subjective basis. It is a matter of approximation and therefore of degree of resolution.

DIFFERENT EFFECTIVE DIMENSIONS IMPLICIT IN A BALL OF THREAD

To confirm this last hunch, a ball of 10 cm diameter made of a thick thread of 1 mm diameter possesses (in latent fashion) several distinct effective dimensions.

To an observer placed far away, the ball appears as a zero-dimensional figure: a point. (Anyhow, it is asserted by Blaise Pascal and by medieval philosophers that on a cosmic scale our whole world is but a point!) As seen from a distance of 10 cm resolution, the ball of thread is a three-dimensional figure. At 10 mm, it is a mess of one-dimensional threads.

At 0.1 mm, each thread becomes a column and the whole becomes a three-dimensional figure again. At 0.01 mm, each column dissolves into fibers, and the ball again becomes one-dimensional, and so on, with the dimension crossing over repeatedly from one value to another. When the ball is represented by a finite number of atomlike pinpoints, it becomes zero-dimensional again. An analogous sequence of dimensions and crossovers is encountered in a sheet of paper.

The notion that a numerical result should depend on the relation of object to observer is in the spirit of physics in this century and is even an exemplary illustration of it.

Most of the objects considered in this Essay are like our ball of thread: they exhibit a succession of different effective dimensions. But a vital new element is added: certain ill-defined transitions between zones of well-defined dimension are reinterpreted as being fractal zones within which $D > D_T$.

SPATIAL HOMOGENEITY, SCALING, AND SELF-SIMILARITY

Having finished with dimensions for the time being, let us prepare for the theme of symmetry by recalling that Euclid begins with the simplest shapes, such as lines, planes, or spaces. And the simplest physics arises when some quantity such as density, temperature, pressure, or velocity is distributed in a homogeneous manner.

The homogeneous distribution on a line, plane, or space has two very desirable properties. It is *invariant under displacement*, and it is *invariant under change of scale*. When we move on to fractals, either invariance *must* be modified and/or restricted in its scope. Hence, the best fractals are those that exhibit the maximum of invariance.

Concerning displacement: different parts of the trail of Brownian motion can never be precisely superposed on each other—as can be done with equal parts of a straight line. Nevertheless, the parts can be made to be superposable in a statistical sense. Nearly all the fractals in the present Essay are to some extent invariant under displacement.

Furthermore, most fractals in this Essay are invariant under certain transformations of scale. They are called *scaling*. A fractal invariant under ordinary geometric similarity is called *self-similar*.

In the compound term *scaling fractals*, the adjective serves to mitigate the noun. While the primary term *fractal* points to disorder and covers cases of intractable irregularity, the modifier *scaling* points to a kind of order. Alternatively, taking *scaling* as the primary term pointing to strict order, *fractal* is a modifier meant to exclude lines and planes.

The motivation for assuming homogeneity and scaling must not be misinterpreted. Here as in standard geometry of nature, no one believes that the world is strictly homogeneous or scaling. Standard geometry investigates straight lines as a preliminary. And mechanics also views uniform rectilinear motion as merely a first step.

The same is true of the study of scaling fractals, but the first step takes much longer in this case because the straight line is replaced by a wealth of diverse possibilities, which this book can merely sample. One should not be surprised that scaling fractals should be limited to providing first approximations of the natural shapes to be tackled. One must rather marvel that these first approximations are so strikingly reasonable.

It is good to point out that self-similarity is an old idea. In the case of the line, it occurred to Leibniz circa 1700 (see under SCALING IN LEIBNIZ AND LAPLACE in Chapter 41). And its generalization beyond lines and planes is almost a hundred years old in mathematics, though its concrete importance was not appreciated until this Essay. Also, it is not new in science, since Lewis F. Richardson postulated in 1926 that over a wide range of scales turbulence is decomposable into self-similar eddies. Furthermore, striking *analytical* consequences of this idea in mechanics are drawn in Kolmogorov 1941. And the analytic aspects of scaling in physics are associated with the notion of renormalization group, Chapter 36.

However, this Essay's 1975 predecessor was the first to address itself to the *geometric* aspects of nonstandard scaling in *Nature*.

"SYMMETRIES" BEYOND SCALING

After it finishes with lines, Euclid tackles shapes with richer properties of invariance, usually called "symmetries." And this Essay also makes a fairly lengthy excursion into nonscaling fractals, in Chapters 15 to 20.

Self-mapping but nonscaling fractals are intimately linked with some of the most refined and difficult areas of "hard" classical mathematical analysis. Contrary to rumors that analysis is a dry subject, these fractals tend to be astoundingly beautiful.

DIVERGENCE SYNDROMES

Almost every case study we perform involves a divergence syndrome. That is, some quantity that is commonly expected to be positive and finite turns out either to be infinite or to vanish. At first blush, such misbehavior looks most bizarre and even terrifying, but a careful reexamination shows it to be quite acceptable..., as long as one is willing to use new methods of thought.

Cases where a symmetry is accompanied by a divergence are also a familiar fixture of quantum physics, within which diverse divergence eliminating arguments take a place of honor. Luckily, the various fractal divergences are much easier to handle. ► ■

4 ¤ Variations and Disclaimers

Now that the diverse objectives of this Essay are outlined, we examine its manner. It too attempts to integrate several distinct facets.

OBSCURITY IS NOT A VIRTUE

To be accessible to scholars and students not necessarily specializing in the various subjects tackled, many of which are esoteric, this work incorporates much exposition.

But exposition is ***not*** *its principal purpose.*

Further, an attempt is made not to frighten away those who are not interested in mathematical precision, but who ought to be interested in my main conclusions. Rigorous mathematical backup is available throughout (and is sounder than in much of physics), but the book's style is informal (though precise). All detail is set aside to Chapter 39, to the references, and to diverse works to come.

Since original work is not expected to show such concerns, this Essay is to some extent a *work of popularization.*

But popularization is ***not*** *its main purpose.*

ERUDITION IS GOOD FOR THE SOUL

As exemplified in Chapter 2, this Essay includes many old and obscure references. Most did not attract my attention until well after my own work in related areas was essentially complete. They did not influence my thinking. However, during the long years when my interests were not shared by anyone, I rejoiced in discovering analogous concerns in ancient works, however fleetingly and ineffectually expressed, witness their failure to be developed. In this fashion, an interest in "classics," which the usual practice of science destroys, was nurtured in my case.

In other words, I rejoiced in finding that the stones I needed—as the architect and builder of the theory of fractals—included many that had been considered by others. But why continue to dwell on this fact today? Casual footnotes would satisfy the prevailing custom, while an excessive stress on distant roots or origins risks fostering the absurd impression that my building is largely a pile of old stones with new names on them.

Thus, my antiquarian curiosity would re-

quire a justification, but I shall not attempt one. It is enough to say that, in my opinion, an interest in the history of ideas is good for the scientist's soul.

However, whenever we read a great man's writings in a light with which he was not blessed, we may ponder the delightful preface Lebesgue wrote to a book by Lusin. He disclaimed many profound thoughts with which said book credited him, saying he might have, or should have, had these thoughts, but had not, and that they originated with Lusin. A related item is Whittaker 1953, wherein quotes from Poincaré and Lorentz are marshalled in favor of a thesis both had pointedly disclaimed: that the physical theory of relativity was their creation and not Einstein's.

Also, for each author jotting down years ago an idea which we can now develop but he did not, we run the risk of finding a second author to declare that the idea is absurd. And should we credit the young Henri Poincaré with ideas he failed to develop, and the mature Henri Poincaré rejected? Stent 1972 might lead us to the conclusion that prematurity, being too much ahead of one's time, deserves nothing but compassionate oblivion.

While excessive erudition in relation to the history of ideas is self-defeating. I do wish to assert the echoes from the past, stressing them further in the biographical and historical sketches in Chapters 40 and 41.

Yet, a display of erudition is certainly ***not*** *the main purpose of this Essay.*

"TO SEE IS TO BELIEVE"

In a letter to Dedekind, at the very beginning of the 1875–1925 crisis in mathematics, Cantor is overwhelmed by amazement at his own findings, and slips from German to French to exclaim that "to see is not to believe" ("Je le vois, mais je ne le crois pas"). And, as if on cue, mathematics seeks to avoid being misled by the graven images of monsters. What a contrast between the rococo exuberance of pre- or counterrevolutionary geometry, and the near-total visual barrenness of the works of Weierstrass, Cantor, or Peano! In physics, an analogous development threatened since about 1800, since Laplace's *Celestial Mechanics* avoided all illustration. And it is exemplified by the statement by P. A. M. Dirac (in the preface of his 1930 *Quantum Mechanics*) that nature's "fundamental laws do not govern the world as it appears in our mental picture in any very direct way, but instead they control a substratum of which we cannot form a mental picture without introducing irrelevancies."

The wide and uncritical acceptance of this view has become destructive. In particular, in the theory of fractals "to see *is* to believe." Therefore, before he proceeds further, the reader is again advised to browse through my picture book. This Essay was designed to help make its contents accessible in various degrees to a wide range of readers, and to try and convince even the purest among mathematicians that the understanding of known concepts and the search for new concepts and

conjectures are both helped by fine graphics. Rarely does contemporary scientific literature show such trust in the usefulness of graphics.

However, showing pretty pictures is ***not*** *the main purpose in this Essay; they are an essential tool, but only a tool.*

One must also recognize that any attempt to illustrate geometry involves a basic fallacy. For example, a straight line is unbounded and infinitely thin and smooth, while any illustration is unavoidably of finite length, of positive thickness, and rough edged. Nevertheless, a rough evocative drawing of a line is felt by many to be useful, and by some to be necessary, to develop intuition and help in the search for proof. And a rough drawing is a more adequate geometric model of a thread than the mathematical line itself. In other words, it suffices for all practical purposes that a geometric concept and its image should fit within a certain range of characteristic sizes, ranging between a large but finite size to be called outer cutoff and a small but positive inner cutoff.

Today, thanks to computer-controlled graphics, the same kind of evocative illustration is practical in the case of fractals. For example, all self-similar fractal curves are also unbounded and infinitely thin. Also, each has a very specific lack of smoothness, which makes it more complicated than anything in Euclid. The best representation, therefore, can only hold within a limited range, on the principles we have already encountered. However, cutting off the very large and the very small detail is not only quite acceptable but even eminently appropriate, because both cutoffs are either present or suspected in Nature. Thus the typical fractal curve can be evoked satisfactorily by elementary strokes in large but finite number.

The larger the number of strokes and the greater the accuracy of the process, the more useful the representation, because fractal concepts refer to the mutual placement of strokes in space, and it is vital in illustrating them to keep to precise scale. Hand drawing would be prohibitive, but computer graphics serves beautifully. My successive Essays have been very much influenced by the availability of increasingly sophisticated systems—and of increasingly sophisticated programmer-artists to run them! Also, I am fortunate in having access to a device that produces camera ready illustrations. This Essay provides a sample of its output.

Graphics is wonderful for matching models with reality. When a chance mechanism agrees with the data from some analytic viewpoint but simulations of the model do not look at all "real," the analytic agreement should be suspect. A formula can relate to only a small aspect of the relationship between model and reality, while the eye has enormous powers of integration and discrimination. True, the eye sometimes sees spurious relationships which statistical analysis later negates, but this problem arises mostly in areas of science where samples are very small. In the areas we shall explore, samples are huge.

In addition, graphics helps find new uses for existing models. I first experienced this

possibility with the random walk illustration in Feller 1950—the curve looked like a mountain's profile or cross section, and the points where it intersects the time axis reminded me of certain records I was then investigating, relative to telephone errors. The ensuing hunches eventually led to the theories presented in Chapters 28 and 31, respectively. My own computer-generated illustrations provided similar inspiration, both to me and to others kind enough to "scout" for me in more sciences than I knew existed.

Naturally, graphics is extended by cinematography: films concerned with some classical fractals have been provided by Max 1971.

THE STANDARD FORM, AND THE NEW FRACTAL FORM, OF GEOMETRIC "ART"

As to this book's endpapers and diverse patterns scattered around, they were the unintended result of faulty computer programming. I hear and read of both the intended and the unintended illustrations being described as a "New Form of Art."

Clearly, competing with artists is ***not at all*** *a purpose of this Essay.* Nevertheless, one must address this issue. The question is *not* whether the illustrations are neatly drawn and printed, and the originals being drawn by computer is not essential either, except in terms of economics. But we do deal with a new form of the controversial but ancient theme that all graphical representations of mathematical concepts are a form of art, one that is best when it is simplest, when (to borrow a painter's term) it can be called "minimal art."

It is widely held that minimal art is restricted to limited combinations of standard shapes: lines, circles, spirals, and the like. But such need not be the case. The fractals used in scientific models are also very simple (because science puts a premium on simplicity). And I agree that many may be viewed as a new form of minimal geometric art.

Is some of it reminiscent of M. C. Escher? It should be, because Escher had the merit of letting himself be inspired by the hyperbolic tilings in Fricke & Klein 1897, which (see Chapter 18) relate closely to shapes that are being incorporated into the fractal realm.

The fractal "new geometric art" shows surprising kinship to Grand Masters paintings or Beaux Arts architecture. An obvious reason is that classical visual arts, like fractals, involve very many scales of length and favor self-similarity (Mandelbrot 1981*l*). For all these reasons, and also because it came in through an effort to imitate Nature in order to guess its laws, it may well be that fractal art is readily accepted because it is not truly unfamiliar. Abstract paintings vary on this account: those I like also tend to be close to fractal geometric art, but many are closer to standard geometric art—too close for my own comfort and enjoyment.

A paradox emerges here: As observed in Dyson's quote in Chapter 1, modern mathematics, music, painting, and architecture may seem to be related to one another. But this is

a superficial impression, notably in the context of architecture: A Mies van der Rohe building is a scalebound throwback to Euclid, while a high period Beaux Arts building is rich in fractal aspects.

POINTS OF LOGISTICS

Successive chapters take up diverse topics by increasing complexity, in order to introduce the basic ideas gradually. The fact that this approach seems feasible is a great asset for the theory of fractals. The amount of built-in repetition is such that the reader is unlikely to lose the main thrust of the argument if he skips the passages he feels to be either repetitious or too complicated (in particular, those that go beyond the most elementary mathematics). Much information is included in the captions of the plates.

As already mentioned, the plates are grouped after the chapters where they are first examined. Also this writer feels every so often the need to engage in private conversation, so to speak, with specific groups of readers who might be overly troubled if some point were left unmentioned or unexplained. The digressions are left in the text but marked by the newfangled brackets ◅ and ►, which should make them easier to skip. Other digressions are devoted to incidental remarks I have no time to explore fully. But this Essay is less digressive than the 1977 *Fractals*.

An attempt is made to show at a glance whether the discussion is concerned with theoretical or empirical dimensions D. The latter are mostly known to one or two decimals, and are therefore written as 1.2 or 1.37. The former are written as integers, ratios of integers, ratios of logarithms of integers, or in decimal form to at least *four* decimals.

BACK TO THE BASIC THEME

Having disclaimed diverse goals that are peripheral to this Essay, let me echo Chapter 1. This work is *a manifesto and a casebook*, devoted nearly exclusively to theories and theses which I initiated but which often led to the revival and the reinterpretation of diverse old works.

None of these theories stopped growing, and a few are still at the seed stage. Some are published here for the first time, while others had been described in my earlier articles. In addition, I mention numerous developments my earlier Essays had inspired, and which in turn stimulated me. However, I do not attempt to list all the fields where fractals prove useful, for fear of destroying the style of an Essay and the flavor of a manifesto.

Last reminder: I do not propose to develop any case study in the full detail desired by the specialists. But many topics are touched upon repeatedly; don't forget to use the index. ■

II ¤¤ THREE CLASSIC FRACTALS, TAMED

5 ¤ How Long Is the Coast of Britain?

To introduce a first category of fractals, namely curves whose fractal dimension is greater than 1, consider a stretch of coastline. It is evident that its length is at least equal to the distance measured along a straight line between its beginning and its end. However, the typical coastline is irregular and winding, and there is no question it is much longer than the straight line between its end points.

There are various ways of evaluating its length more accurately, and this chapter analyzes several of them. The result is most peculiar: coastline length turns out to be an elusive notion that slips between the fingers of one who wants to grasp it. All measurement methods ultimately lead to the conclusion that the typical coastline's length is very large and so ill determined that it is best considered infinite. Hence, if one wishes to compare different coastlines from the viewpoint of their "extent," length is an inadequate concept.

This chapter seeks an improved substitute, and in doing so finds it impossible to avoid introducing various forms of the fractal concepts of dimension, measure, and curve.

MULTIPLICITY OF ALTERNATIVE METHODS OF MEASUREMENT

METHOD A: Set dividers to a prescribed opening ϵ, to be called the yardstick length, and walk these dividers along the coastline, each new step starting where the previous step leaves off. The number of steps multiplied by ϵ is an approximate length $L(\epsilon)$. As the dividers' opening becomes smaller and smaller, and as we repeat the operation, we have been taught to expect $L(\epsilon)$ to settle rapidly to a well-defined value called the *true length*. But in fact what we expect does not happen. In the typical case, the observed $L(\epsilon)$ tends to increase without limit.

The reason for this behavior is obvious:

When a bay or peninsula noticed on a map scaled to 1/100,000 is reexamined on a map at 1/10,000, subbays and subpeninsulas become visible. On a 1/1,000 scale map, sub-subbays and sub-subpeninsulas appear, and so forth. Each adds to the measured length.

Our procedure acknowledges that a coastline is too irregular to be measured directly by reading it off in a catalog of lengths of simple geometric curves. Therefore, METHOD A replaces the coastline by a sequence of broken lines made of straight intervals, which are curves we know how to handle.

METHOD B: Such "smoothing out" can also be accomplished in other ways. Imagine a man walking along the coastline, taking the shortest path that stays no farther from the water than the prescribed distance ϵ. Then he resumes his walk after reducing his yardstick, then again, after another reduction; and so on, until ϵ reaches, say, 50 cm. Man is too big and clumsy to follow any finer detail. One may further argue that this unreachable fine detail (a) is of no direct interest to Man and (b) varies with the seasons and the tides so much that it is altogether meaningless. We take up argument (a) later on in this chapter. In the meantime, we can neutralize argument (b) by restricting our attention to a rocky coastline observed when the tide is low and the waves are negligible. In principle, Man could follow such a curve down to finer details by harnessing a mouse, then an ant, and so forth. Again, as our walker stays increasingly closer to the coastline, the distance to be covered continues to increase with no limit.

METHOD C: An asymmetry between land and water is implied in METHOD B. To avoid it, Cantor suggests, in effect, that one should view the coastline with an out-of-focus camera that transforms every point into a circular blotch of radius ϵ. In other words, Cantor considers all the points of both land and water for which the distance to the coastline is no more than ϵ. These points form a kind of sausage or tape of width 2ϵ, as seen in a different context on Plate 32. Measure the area of the tape and divide it by 2ϵ. If the coastline were straight, the tape would be a rectangle, and the above quotient would be the actual length. With actual coastlines, we have an estimated length $L(\epsilon)$. As ϵ decreases, this estimate increases without limit.

METHOD D: Imagine a map drawn in the manner of pointillist painters using circular blotches of radius ϵ. Instead of using circles centered on the coastline, as in METHOD C, let us require that the blotches that cover the entire coastline be as few in number as possible. As a result, they may well lie mostly inland near the capes and mostly in the sea near the bays. Such a map's area, divided by 2ϵ, is an estimate of the length. This estimate also "misbehaves."

ARBITRARINESS OF THE RESULTS OF MEASUREMENT

To summarize the preceding section, the main finding is always the same. As ϵ is made smaller and smaller, every approximate length

tends to increase steadily without bound.

In order to ascertain the meaning of this result, let us perform analogous measurements on a standard curve from Euclid. For an interval of straight line, the approximate measurements are essentially identical and define the length. For a circle, the approximate measurements increase but converge rapidly to a limit. The curves for which a length is thus defined are called *rectifiable.*

An even more interesting contrast is provided by the results of measurement on a coastline that Man has tamed, say the coast at Chelsea as it is today. Since very large features are unaffected by Man, a very large yardstick again yields results that increase as ϵ decreases.

However, there is an intermediate zone of ϵ's in which $L(\epsilon)$ varies little. This zone may go from 20 meters down to 20 centimeters (but do not take these values too strictly). But $L(\epsilon)$ increases again after ϵ becomes less than 20 centimeters and measurements become affected by the irregularity of the stones. Thus, if we trace the curves representing $L(\epsilon)$ as a function of ϵ, there is little doubt that the length exhibits, in the zone of ϵ's between ϵ=20 meters and ϵ=20 centimeters, a flat portion that was not observable before the coast was tamed.

Measurements made in this zone are obviously of great practical use. Since boundaries between different scientific disciplines are largely a matter of conventional division of labor between scientists, one might restrict geography to phenomena above Man's reach, for example, on scales above 20 meters. This restriction would yield a well-defined value of geographical length. The Coast Guard may well choose to use the same ϵ for untamed coasts, and encyclopedias and almanacs could adopt the corresponding $L(\epsilon)$.

However, the adoption of the same ϵ by all the agencies of a government is hard to imagine, and its adoption by all countries is all but inconceivable. For example, Richardson 1961, the lengths of the common frontiers between Spain and Portugal, or Belgium and Netherlands, as reported in these neighbors' encyclopedias, differ by 20%. The discrepancy must in part result from different choices of ϵ. An empirical finding to be discussed soon shows that it suffices that the ϵ differ by a factor of 2, and one should not be surprised that a small country (Portugal) measures its borders more accurately than its big neighbor.

The second and more significant reason against deciding on an arbitrary ϵ is philosophical and scientific. Nature does exist apart from Man, and anyone who gives too much weight to any specific ϵ and $L(\epsilon)$ lets the study of Nature be dominated by Man, either through his typical yardstick size or his highly variable technical reach. If coastlines are ever to become an object of scientific inquiry, the uncertainty concerning their lengths cannot be legislated away. In one manner or another, the concept of geographic length is not as inoffensive as it seems. It is not entirely "objective." The observer inevitably intervenes in its definition.

IS THIS ARBITRARINESS GENERALLY RECOGNIZED, AND DOES IT MATTER?

The view that coastline lengths are nonrectifiable is doubtless held true by many people, and I for one do not recall ever thinking otherwise. But my search for written statements to this effect is a near fiasco. Aside from the Perrin quote in Chapter 2, there is the observation in Steinhaus 1954 that "the left bank of the Vistula, when measured with increasing precision, would furnish lengths ten, hundred or even thousand times as great as the length read off the school map...[A] statement nearly approaching reality would be to call most arcs encountered in nature nonrectifiable. This statement is contrary to the belief that nonrectifiable arcs are an invention of mathematicians and that natural arcs are rectifiable: it is the opposite that is true." But neither Perrin nor Steinhaus follow up on this insight.

Let me also retell a story reported by C. Fadiman. His friend Edward Kasner would ask small tots "to guess the length of the eastern coast line of the United States. After a 'sensible' guess had been made...he would...point out that this figure increased enormously if you measured the perimeter of each bay and inlet, then that of every projection and curve of each of these, then the distance separating every small particle of coastline matter, each molecule, atom, etc. Obviously the coast line is as long as you want to make it. The children understood this at once; Kasner had more trouble with grownups." The story is nice, but it is not relevant here: Kasner's goal was *not* to point out an aspect of Nature worthy of further exploration.

Therefore, Mandelbrot 1967s and the present Essay are effectively the first works on this subject.

One is reminded of William James writing in *The Will to Believe* that "The great field for new discoveries...is always the unclassified residuum. Round about the accredited and orderly facts of every science there ever floats a sort of dust-cloud of exceptional observations, of occurrences minute and irregular and seldom met with, which it always proves more easy to ignore than to attend to. The ideal of every science is that of a closed and completed system of truth... Phenomena unclassifiable within the system are paradoxical absurdities, and must be held untrue...—one neglects or denies them with the best of scientific consciences... Any one will renovate his science who will steadily look after the irregular phenomena. And when the science is renewed, its new formulas often have more of the voice of the exception in them than of what were supposed to be the rules."

This Essay, whose ambition is indeed to renew the Geometry of Nature, relies upon many puzzles so unclassified that they are only published when the censors nod. The next section discusses a first example.

THE RICHARDSON EFFECT

The variation of the approximate length $L(\epsilon)$ obtained by Method A has been studied em-

pirically in Richardson 1961, a reference that chance (or fate) put in my way. I paid attention because (Chapter 40) I knew of Lewis Fry Richardson as a great scientist whose originality mixed with eccentricity. As we shall learn in Chapter 10, we are indebted to him for some of the most profound and most durable ideas regarding the nature of turbulence, notably the notion that turbulence involves a self-similar cascade. He also concerned himself with other difficult problems, such as the nature of armed conflict between states. His experiments were of classic simplicity, but he never hesitated to use refined concepts when he deemed them necessary.

The diagrams reproduced in Plate 33, found among his papers after he died, were published in a near confidential (and totally inappropriate) *Yearbook*. They all lead to the conclusion that there are two constants, which we shall call λ and D, such that—to approximate a coastline by a broken line—one needs roughly $F\epsilon^{-D}$ intervals of length ϵ, adding up to the length

$$L(\epsilon) \sim F\epsilon^{1-D}.$$

The value of the exponent D seems to depend upon the coastline that is chosen, and different pieces of the same coastline, if considered separately, may produce different values of D. To Richardson, the D in question was a simple exponent of no particular significance. However, its value seems to be independent of the method chosen to estimate the length of a coastline. Thus D seems to warrant attention.

A COASTLINE'S FRACTAL DIMENSION (MANDELBROT 1967s)

Having unearthed Richardson's work, I proposed (Mandelbrot 1967s) that, despite the fact that the exponent D is not an integer, it can and should be interpreted as a dimension, namely, as a fractal dimension. Indeed, I recognized that all the above listed methods of measuring $L(\epsilon)$ correspond to nonstandard generalized definitions of dimension already used in pure mathematics. The definition of length based on the coastline being covered by the smallest number of blotches of radius ϵ is used in Pontrjagin & Schnirelman 1932 to define the covering dimension. The definition of length based on the coastline being covered by a tape of width 2ϵ implements an idea of Cantor and Minkowski (Plate 32), and the corresponding dimension is due to Bouligand. Yet these two examples only hint at the many dimensions (most of them known only to a few specialists) that star in diverse specialized chapters of mathematics. A certain number of them are discussed further in Chapter 39.

Why did mathematicians introduce this plethora of distinct definitions? Because in some cases they yield distinct values. Luckily, however, such cases are never encountered in this Essay, and the list of possible alternative dimensions can be reduced to two that I have not yet mentioned. The older and best investigated one dates back to Hausdorff and serves to define fractal dimension; we come to it momentarily. The simpler one is similarity dimension: it is less general, but in many cases

is more than adequate; it is explored in the following chapter.

Clearly, I do not propose to present a mathematical proof that Richardson's D is a dimension. No such proof is conceivable in any natural science. The goal is merely to convince the reader that the notion of length poses a conceptual problem, and that D provides a manageable and convenient answer. Now that fractal dimension is injected into the study of coastlines, even if specific reasons come to be challenged, I think we shall never return to the stage when $D=1$ was accepted thoughtlessly and naively. He who continues to think that $D=1$ has to argue his case.

The next step, to explain the shape of the coastlines and to deduce the value of D from other more basic considerations, is put off until Chapter 28. Suffice at this point to announce that to a first approximation $D=3/2$. This value is much too large to describe the facts but more than sufficient to establish that it is natural, proper, and expected for a coastline's dimension to exceed the standard Euclidean value $D=1$.

HAUSDORFF FRACTAL DIMENSION

If we accept that various natural coasts are really of infinite length and that the length based on an anthropocentric value of ϵ gives only a partial idea of reality, how can different coastlines be compared to each other? Since infinity equals four times infinity, every coastline is four times longer than each of its quarters, but this is not a useful conclusion. We need a better way to express the sound idea that the entire curve must have a "measure" that is four times greater than each of its fourths.

A most ingenious method of reaching this goal has been provided by Felix Hausdorff. It is intuitively motivated by the fact that the linear measure of a polygon is calculated by adding its sides' lengths without transforming them in any way. One may say (the reason for doing so will soon become apparent) that these lengths are raised to the power $D=1$, the Euclidean dimension of a straight line. The surface measure of a closed polygon's interior is similarly calculated by paving it with squares, and adding the squares' sides raised to the power $D=2$, the Euclidean dimension of a plane. When, on the other hand, the "wrong" power is used, the result gives no specific information: the area of every closed polygon is zero, and the length of its interior is infinite.

Let us proceed likewise for a polygonal approximation of a coastline made up of small intervals of length ϵ. If their lengths are raised to the power D, we obtain a quantity we may call tentatively an "approximate measure in the dimension D." Since according to Richardson the number of sides is $N=F\epsilon^{-D}$, said approximate measure takes the value $F\epsilon^{D}\epsilon^{-D}=F$.

Thus, *the approximate measure in the dimension* D *is independent of* ϵ. With actual data, we simply find that this approximate measure varies little with ϵ.

In addition, the fact that the length of a square is infinite has a simple counterpart and generalization: a coastline's approximate measure evaluated in any dimension d smaller than D tends to ∞ as $\epsilon \to 0$. Similarly, the area and the volume of a straight line are zero. And when d takes any value larger than D, the corresponding approximate measure of a coastline tends to 0 as $\epsilon \to 0$. The approximate measure behaves reasonably if and only if d=D.

A CURVE'S FRACTAL DIMENSION MAY EXCEED 1; FRACTAL CURVES

By design, the Hausdorff dimension preserves the ordinary dimension's role as exponent in defining a *measure*.

But from another viewpoint, D is very odd indeed: it is a fraction! In particular, it exceeds 1, which is the intuitive dimension of curves and which may be shown rigorously to be their topological dimension D_T.

I propose that curves for which the fractal dimension exceeds the topological dimension 1 be called *fractal curves*. And the present chapter can be summarized by asserting that, within the scales of interest to the geographer, coastlines can be modeled by fractal curves. Coastlines are *fractal patterns*. ■

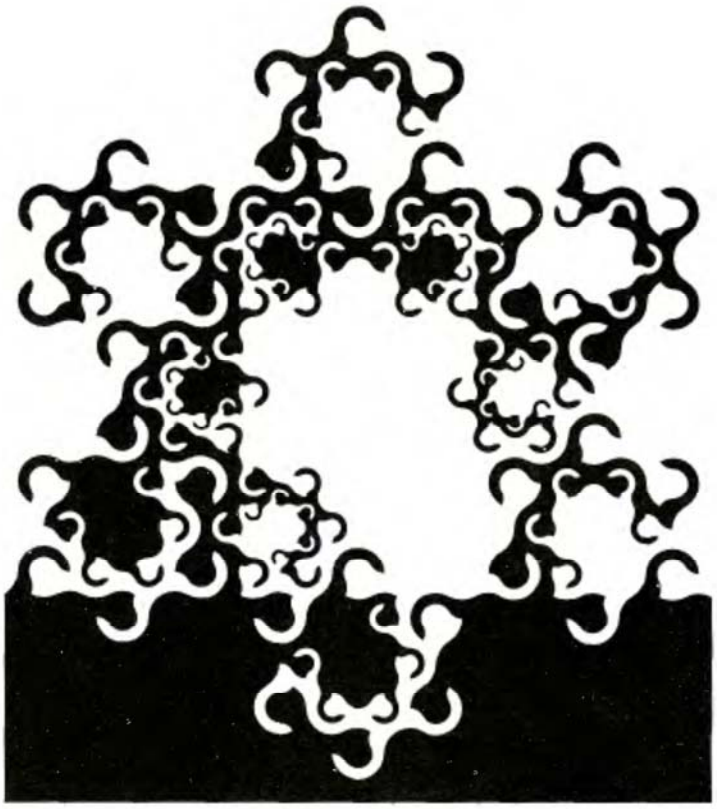
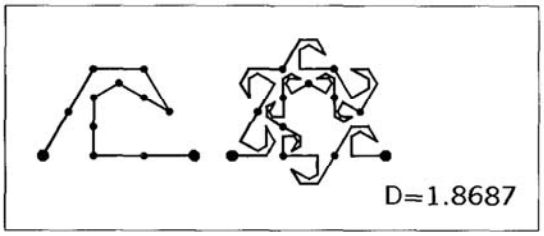


Plate 31 ¤ MONKEYS TREE

At this point, the present small incidental plate should be viewed as merely a decorative drawing, filling a gap.

However, when the reader has finished Chapter 14, he will find in this drawing a hint to help unscramble the "architecture" in Plate 146. A more sober hint resides in the following generator.



■

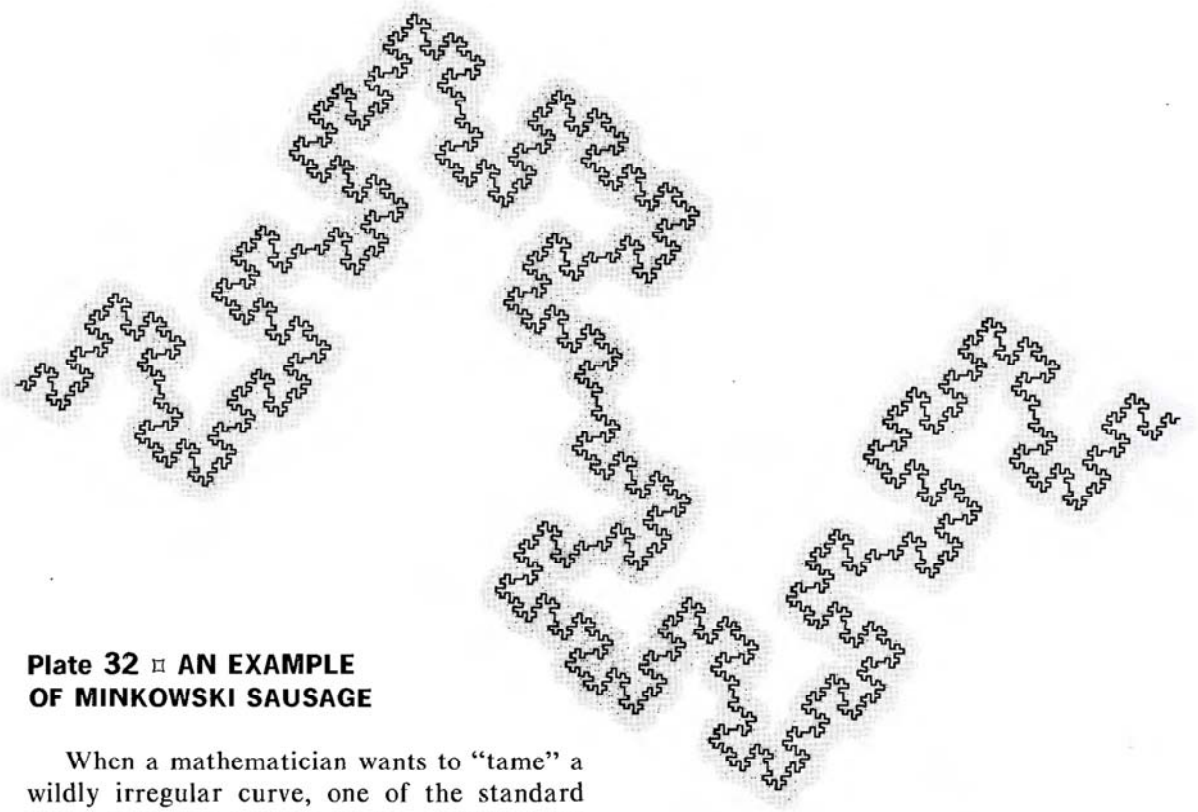
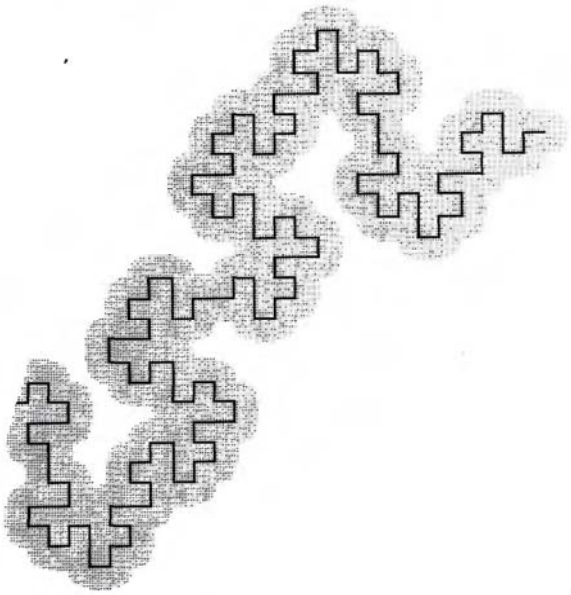


Plate 32 ¤ AN EXAMPLE OF MINKOWSKI SAUSAGE

When a mathematician wants to "tame" a wildly irregular curve, one of the standard procedures is to select a radius ϵ and to draw around each point of the curve a disc of radius ϵ. This procedure, dating back at least to Hermann Minkowski and possibly to Georg Cantor, is brutal but very effective. (As to the term *sausage*, unverifiable rumor claims it is a leftover of an application of this procedure to the Brownian curves of Norbert Wiener.)

In the present illustration such smoothing is not applied to an actual coastline but to a theoretical curve that will be constructed later (Plate 49) by continual addition of ever smaller detail. Comparing the piece of sausage drawn to the right with the rightmost end of the sausage drawn above it, we see that the construction of the curve passes a critical stage when it begins to involve details of size smaller than ϵ. Later stages of construction leave the sausage essentially unaffected. ■



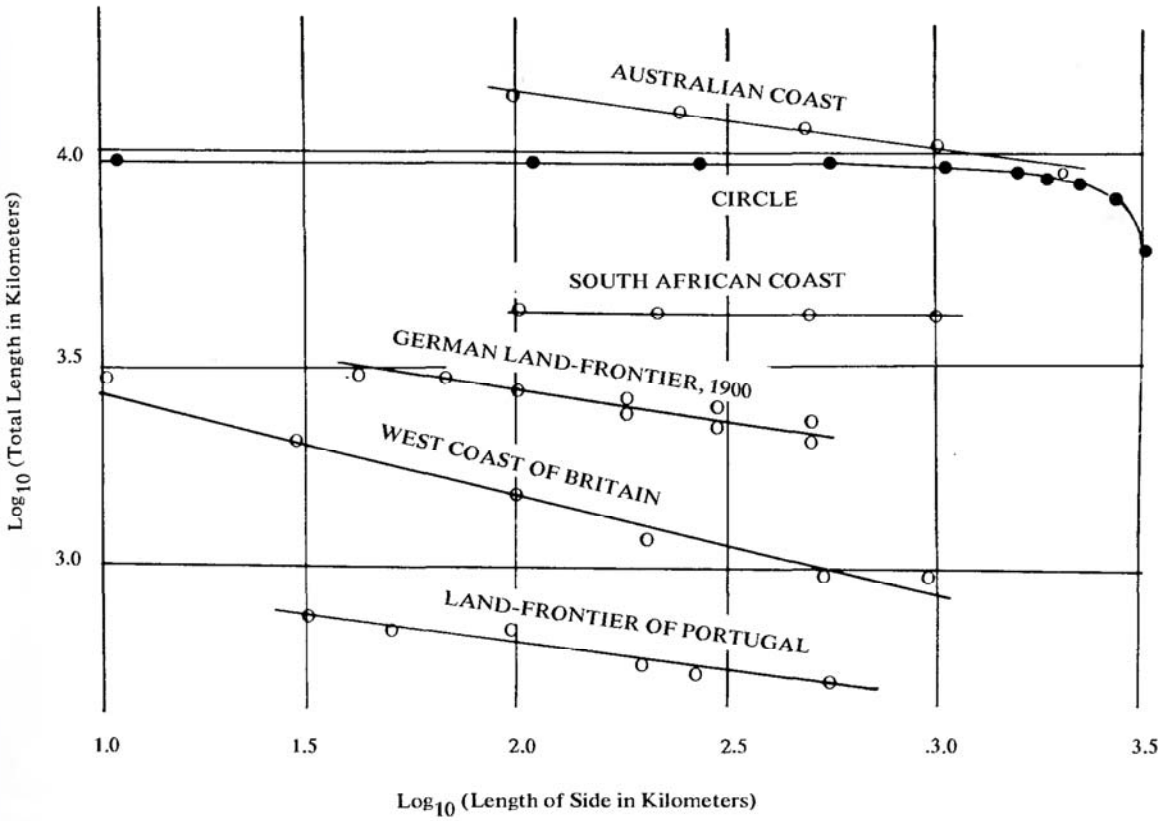


Plate 33 ¤ RICHARDSON'S EMPIRICAL DATA CONCERNING THE RATE OF INCREASE OF COASTLINES' LENGTHS

This Figure reproduces Richardson's experimental measurements of length performed on various curves using equal-sided polygons of increasingly short side ϵ. As expected, increasingly precise measurements made on a circle stabilize very rapidly near a well-determined value.

In the case of coastlines, on the contrary, the approximate lengths do not stabilize at all. As the yardstick length ϵ tends to zero, the approximate lengths, as plotted on doubly logarithmic paper, fall on a straight line of negative slope. The same is true of boundaries between countries. Richardson's search in encyclopedias reveals notable differences in the lengths of the common land frontiers claimed by Spain and Portugal (987 versus 1214 km), and by the Netherlands and Belgium (380 versus 449 km). With a slope of –0.25, the 20% differences between these claims can be accounted for by assuming that the ϵ's differ by a factor of 2, which is not unlikely.

To Richardson, his lines' slopes had no theoretical interpretation. The present Essay, on the other hand, interprets coastlines as approximate fractal curves, and uses the slope of each line as an estimate of 1–D, where D is the fractal dimension. ■

6 ¤ Snowflakes and Other Koch Curves

In order to understand fully my interpretation of Richardson's D as a fractal dimension, we move from natural phenomena over which we have no control, to geometric constructs we can design at will.

SELF-SIMILARITY AND CASCADES

Until now we stressed that coastlines' geometry is complicated, but there is also a great degree of order in their structure.

Although maps drawn at different scales differ in their specific details, they have the same generic features. In a rough approximation, the small and large details of coastlines are geometrically identical except for scale.

One may think of such a shape as drawn by a sort of fireworks, with each stage creating details smaller than those of the preceding stages. However, a better term is suggested by our Lewis Richardson's noted work on turbulence: the generating mechanism may be called a *cascade*.

When each piece of a shape is geometrically similar to the whole, both the shape and the cascade that generate it are called *self-similar*. This chapter probes self-similarity using very regular figures.

The most extreme contrasts to self-similar shapes are provided by curves that (a) have a single scale, like the circle, or (b) have two clearly separated scales, like a circle adorned with "scallops." Such shapes can be described as *scalebound*.

COASTLIKE TERAGONS AND THE TRIADIC KOCH CURVE $\mathcal{K}$

To insure that an infinite number of scales of length are present in a curve, the safest is to put them in deliberately and separately. A regular triangle of side 1 has a single scale, triangles of side ⅓ have a smaller scale, and triangles of side $(⅓)^k$ are of increasingly small scale. And by piling these triangles on top of each other, as in Plate 42, one is left with a shape combining all scales below 1.

In effect, we assume that a bit of coastline drawn to a scale of 1/1,000,000 is a straight interval of length 1, to be called *initiator*.

Then we assume that the detail that becomes visible on a map at 3/1,000,000 replaces the earlier interval's middle third by a promontory in the shape of an equilateral triangle. The resulting second approximation is an broken line formed of four intervals of equal lengths, to be called *generator*. We further assume that the new detail that appears at 9/1,000,000 results from the replacement of each of the generator's four intervals by the generator reduced in a ratio of one-third, forming subpromontories.

Proceeding in this fashion, we break each straight line interval, replacing the initiator by an increasing broken curve. Since we deal with them throughout this Essay, let me coin for such curves the term *teragon*, from the Greek τερας, meaning "monster, strange creature," and γωνια, meaning "corner, angle." Very appropriately, the metric system uses *tera* as prefix for the factor 10^{12}.

And, if the same cascade process is made to continue to infinity, our teragons converge to a limit first considered by von Koch 1904, Plate 45. We must be specific, and shall call it the triadic Koch curve and denote it by $\mathcal{K}$.

This curve's area vanishes, as is obvious on Plate 43. On the other hand, each stage of construction increases its total length in a ratio of 4/3, hence the limit curve is of infinite length. Furthermore, it is continuous, but it has no definite tangent anywhere—like the graph of a continuous function without a derivative.

As a model of a coastline, $\mathcal{K}$ is only a suggestive approximation, but not because it is too irregular, rather because, in comparison with a coastline, its irregularity is far too systematic. Chapters 24 and 28 "loosen it up" to make it fit better.

THE KOCH CURVE AS MONSTER

As introduced in the preceding section, the Koch curve must seem the most intuitive thing in geometry. But the conventional motivation for it is totally different. So is the conventional attitude towards it on the part of mathematicians. They are all but unanimous in proclaiming that $\mathcal{K}$ is a monstrous curve! For elaboration, let us look up *The Crisis of Intuition,* Hahn 1956, which will serve us repeatedly. We read that "the character of [a *non*rectifiable curve or of a curve without a tangent] entirely eludes intuition; indeed after a few repetitions of the segmenting process the evolving figure has grown so intricate that intuition can scarcely follow; and it forsakes us completely as regards the curve that is approached as a limit. Only thought, or logical analysis, can pursue this strange object to its final form. Thus, had we relied on intuition in this instance, we should have remained in error, for intuition seems to force the conclusion that there cannot be curves lacking a tangent at any point. This first example of the failure of intuition involves the fundamental concepts of differentiation."

The best one can say of these words is that they stop short of a celebrated exclamation of Charles Hermite, writing on May 20, 1893, to

T. Stieltjes of "turning away in fear and horror from this lamentable plague of functions with no derivatives." (Hermite & Stieltjes 1905, II, p. 318.) One likes to believe that great men are perfect, and that Hermite was being ironic, but Lebesgue's 1922 *Notice* (Lebesgue 1972–, I) suggests otherwise. Having written a paper concerned with surfaces devoid of tangent planes, "thoroughly crumpled handkerchiefs," Lebesgue wanted it published by the Académie des Sciences, but "Hermite for a moment opposed its inclusion in the *Comptes Rendus*; this was about the time when he wrote to Stieltjes...."

We recall that Perrin and Steinhaus knew better, but the only mathematician to argue otherwise on the basis of intuition alone (Steinhaus argues on the basis of fact) is Paul Lévy (Lévy 1970): "[I have] always been surprised to hear it said that geometric intuition inevitably leads one to think that all continuous functions are differentiable. From my first encounter with the notion of derivative, my experience proved that the contrary is true."

These voices had not been heard, however. Not only near every book but every science museum proclaims that nondifferentiable curves are counter-intuitive, "monstrous," "pathological," or even "psychopathic."

THE KOCH CURVE, TAMED. THE DIMENSION D=log 4/log 3=1.2618

I claim that a Koch curve is a rough but vigorous model of a coastline. For a first quantitative test, let us investigate the length $L(\epsilon)$ of the triadic Koch teragon whose sides are of length ϵ. This lengths can be measured exactly, and the result is extraordinarily satisfying:

$$L(\epsilon)=\epsilon^{1-D}.$$

This *exact* formula is identical with Richardson's *empiric* law relative to the coast of Britain. For the triadic Koch curve,

$$D=\log 4/\log 3\sim 1.2618,$$

hence D lies in the range of values observed by Richardson!

◅ PROOF: Clearly, $L(1)=1$ and

$$L(\epsilon/3)=(4/3)L(\epsilon).$$

This equation has a solution of the form $L(\epsilon)=\epsilon^{1-D}$ if D satisfies

$$3^{D-1}=4/3.$$

Hence $D=\log 4/\log 3$, as asserted. ►

Naturally, the Koch D is not an empirical but a mathematical constant. Therefore the argument for calling D a dimension becomes even more persuasive in the case of the Koch curve than in the case of coastlines.

On the other hand, the approximate Hausdorff measure in the dimension D (a notion introduced in the preceding chapter) equals ϵ^D multiplied by the number of legs of length ϵ, that is, equals $\epsilon^D \cdot \epsilon^{-D}=1$. This is a good indication that the Hausdorff dimension is D. Un-

fortunately, the Hausdorff definition is disappointingly difficult to handle rigorously. Moreover, even if it had been easy to handle, the generalization of dimension beyond integers is so far-reaching an idea that one should welcome further motivation for it.

THE SIMILARITY DIMENSION

It happens that in the case of self-similar shapes a very easy further motivation is available in the notion of *similarity dimension.* One often *hears* mathematicians use the similarity dimension to guess the Hausdorff dimension, and the bulk of the present Essay encounters only cases where this guess is correct. In their context, there can be no harm in thinking of fractal dimension as being synonymous with similarity dimension. ◅ We have here a counterpart to the use of topological dimension as synonymous with "intuitive" dimension. ►

As a motivating prelude, let us examine the standard self-similar shapes: intervals in the line, rectangles in the plane, and the like; see Plate 44. Because a straight line's Euclidean dimension is 1, it follows for every integer "base" b that the "whole" interval $0 \le x < X$ may be "paved" (each point being covered once and only once) by $N=b$ "parts." These "parts" are the intervals $(k-1)X/b \le x < kX/b$, where k goes from 1 to b. Each part can be deduced from the whole by a similarity of ratio $r(N)=1/b=1/N$.

Likewise, because a plane's Euclidean dimension is 2, it follows that whatever the value of b, the "whole" made up of a rectangle $0 \le x < X$; $0 \le y < Y$ can be "paved" exactly by $N=b^2$ parts. These parts are rectangles defined by the combined inequalities

$$(k-1)X/b \le x < kX/b,$$

$$\text{and } (h-1)Y/b \le y < hY/b,$$

wherein k and h go from 1 to b. Each part can now be deduced from the whole by a similarity of ratio $r(N)=1/b=1/N^{1/2}$.

For a rectangular parallelepiped, the same argument gives us $r(N)=1/N^{1/3}$.

And there is no problem in defining spaces whose Euclidean dimension is $E>3$. (The Euclidean—or Cartesian—dimension is denoted by E in this book.) All D-dimensional parallelepipeds defined for $D \le E$ satisfy

$$r(N)=1/N^{1/D}.$$

Thus,

$$Nr^D=1.$$

Equivalent alternative expressions are

$$\log r(N)=\log (1/N^{1/D})=-(\log N)/D,$$

$$D=-\log N/\log r(N)=\log N/\log (1/r).$$

Now let us move on to nonstandard shapes. In order for the exponent of self-similarity to have formal meaning, the sole requirement is that the shape be self-similar, i.e., that the

whole may be split up into N parts, obtainable from it by a similarity of ratio r (followed by displacement or by symmetry). The D obtained in this fashion always satisfies

$$0 \leq D \leq E.$$

In the example of the triadic Koch curve, $N=4$ and $r=1/3$, hence $D=\log 4/\log 3$, identical to the Hausdorff dimension.

CURVES; TOPOLOGICAL DIMENSION

Thus far, we have been casual in calling Koch's $\mathcal{K}$ a curve, but we must return to this notion. Intuitively, a standard arc is a connected set that becomes disconnected if any single point is removed. And a closed standard curve is a connected set that separates into standard arcs if 2 points are removed. For the same reason, Koch's $\mathcal{K}$ is a curve.

The mathematician says that all the shapes with the above property, e.g., $\mathcal{K}$, [0,1] or a circle, are of topological dimension $D_T=1$. Thus, yet another notion of dimension has to be considered! Being disciples of William of Ockham, all scientists know that "entities must not be multiplied beyond necessity." It must therefore be confessed that our switching back and forth between several near equivalent forms of fractal dimension is a matter of convenience. However, the coexistence of a fractal and a topological dimension *is a matter of necessity*. Readers who skipped the digressive definition of fractal in Chapter 3 are advised to scan it now, and everyone is advised to read the entry devoted to DIMENSION in Chapter 41.

INTUITIVE MEANING OF D IN THE PRESENCE OF CUTOFFS Λ AND λ

Cesàro 1905 begins with the motto,

The will is infinite
and the execution confined,
the desire is boundless
and the act a slave to limit.

Indeed, limits apply to scientists no less than to Shakespeare's Troilus and Cressida. To obtain a Koch curve, the cascade of smaller and smaller new promontories is pushed to infinity, but in Nature every cascade must stop or change character. While endless promontories may exist, the notion that they are self-similar can only apply between certain limits. Below the lower limit, the concept of coastline ceases to belong to geography.

It is therefore reasonable to view the real coastline as involving two *cutoff scales*. Its *outer cutoff* Ω might be the diameter of the smallest circle encompassing an island, or perhaps a continent, and the *inner cutoff* ϵ might be the 20 meters mentioned in Chapter 5. Actual numerical values are hard to pinpoint, but the need for cutoffs is unquestionable.

Yet, after the very big and the very small details are cut off, D continues to stand for an *effective dimension* as described in Chapter 3.

Strictly speaking, the triangle, the Star of David, and the finite Koch teragons are of dimension 1. However, both intuitively and from the pragmatic point of view of the simplicity and naturalness of the corrective terms required, it is reasonable to consider an advanced Koch teragon as being closer to a curve of dimension $\log 4/\log 3$ than to a curve of dimension 1.

As for a coastline, it is likely to have several separate dimensions (remember the balls of thread in Chapter 3). Its geographic dimension is Richardson's D. But in the range of sizes of interest in physics, the coastline may have a different dimension—associated with the concept of interface between water, air, and sand.

ALTERNATIVE KOCH GENERATORS AND SELF-AVOIDING KOCH CURVES

Let us restate the basic principle of construction of the triadic Koch curve: One begins with *two shapes,* an *initiator* and a *generator.* The latter is an oriented broken line made up of N equal sides of length r. Thus each stage of the construction begins with a broken line and consists in replacing each straight interval with a copy of the generator, reduced and displaced so as to have the same end points as those of the interval being replaced. In all cases, $D=\log N/\log(1/r)$.

It is easy to change this construction by modifying the generator, in particular by combining promontories with bays, as exemplified in upcoming plates. In this way we obtain Koch teragons that converge to curves whose dimensions are between 1 and 2.

All these Koch curves are self-avoiding: have no self-intersection. This is why their wholes can be divided into disjoint parts with no ambiguity, in order to define D. However, a Koch construction using carelessly chosen generators risks self-contact or self-intersection, or even self-overlap. When the desired D is small, it is easy to avoid double points by careful choice of the generator. The task becomes increasingly difficult as D increases, but remains possible as long as $D<2$.

However, any Koch construction that attempts to reach a dimension $D>2$ leads inevitably to curves that cover the plane infinitely many times. The case $D=2$ deserves a special discussion to be provided in Chapter 7.

KOCH ARCS AND HALF LINES

In some cases, the term *Koch curve* must be replaced by more precise, and pedantic, terminology. The shape at the bottom of Plate 44 is technically the *Koch map* of a line interval, and can be called a *Koch arc.* Thus the boundary in Plate 45 is made of three Koch arcs. And it is often useful to extrapolate an arc into a *Koch half line*: The extrapolation enlarges the original arc, using its left end point as focus, in the ratio $1/r=3$, then in the ratio 3^2 and so on. Each successive extrapolate contains the preceding one, and the limit curve contain all the intermediate finite

stages.

DEPENDENCE OF MEASURE ON THE RADIUS, WHEN D IS A FRACTION

Let us now extend from Euclidean to fractal dimensions another standard result in Euclid. For idealized physical objects of uniform density ρ, the weight $M(R)$ of a rod of length $2R$, of a disc of radius R or of a ball of radius R is proportional to ρR^E. For $E=1$, 2, and 3, the proportionality constants are respectively equal to 2, 2π, and $4\pi/3$.

The rule $M(R) \propto R^D$ *also applies to fractals when they are self-similar.*

In the triadic Koch case, the proof is easiest when the origin is the end point of a Koch half line. When a circle of radius $R_0=3^k$ (with $k \geq 0$) contains the mass $M(R_0)$, the circle of radius $R=R_0/3$ contains the mass $M(R)=M(R_0)/4$. Hence,

$$M(R)=M(R_0)(R/R_0)^D = [M(R_0)R_0^{-D}]R^D.$$

Consequently, the ratio $M(R)/R^D$ is independent of R, and can serve to define a "density" ρ.

KOCH MOTION

Imagine a point moving along a Koch half line, taking equal time to cover arcs of equal measure. If we then invert the function giving time as function of position, we obtain a position as function of time, that is, a motion. Of course its velocity is infinite.

PREVIEW OF RANDOM COASTLINES

The Koch curve reminds us of real maps, but has major defects one encounters almost unchanged in the early models of every case study in this Essay. Its parts are identical to each other, and the self-similarity ratio r must be part of a strict scale of the form b^{-k}, where b is an integer, namely, ⅓, $(⅓)^2$, and so on. Thus, a Koch curve is a very preliminary model of a coastline.

I have developed diverse ways of avoiding both defects, but all involve probabilistic complications which are better tackled after we settle many issues concerning nonrandom fractals. However, curious readers familiar with probability may peek ahead to the models based on my "squig curves" (Chapter 24), and, more important, on level curves of fractional Brown surfaces (Chapter 28).

The same method of exposition is followed later in this Part. Numerous patterns of Nature are discussed against the background of systematic fractals that provide a very preliminary model, while the random models I advocate are postponed to later chapters.

REMINDER. In all cases where D is known precisely, is not an integer, and is written in decimal form to enable comparisons, it is carried to *four* decimals. This number 4 is chosen to make obvious that D is *neither* an empirical value (all empirical values are known

at present to 1 or 2 decimals), *nor* an incompletely determined geometric value (at present, the latter are known either to 1 or 2 decimals, or to 6 decimals and more.)

COMPLEX, OR SIMPLE AND REGULAR?

Koch curves exhibit a novel and most interesting combination of complexity and simplicity. At first blush, they are enormously more complicated than the standard curves of Euclid. However, the Kolmogorov and Chaitin theory of mathematical algorithms suggests the contrary conclusion, that a Koch curve is *not* significantly more complicated than a circle! This theory starts with a collection of "letters" or "atomic operations," and takes the length of the shortest known algorithm that yields a desired function as an objective upper bound to the function's complexity.

To apply this way of thinking to the construction of curves, let the letters or "atoms" of the graphic process be straight "strokes." In this alphabet, tracing a regular polygon requires a finite number of strokes, each described by a finite number of lines of instruction, hence it is a task of finite complexity. By contrast, a circle involves an "infinite number of infinitely short strokes," hence seems a curve of infinite complexity. However, if the construction of the circle is made to proceed recursively, it is seen to involve only a finite number of instructions, hence to be also a task of finite complexity. For example, starting with a regular polygon of 2^m sides ($m>2$), one replaces each stroke of length $2\sin(\pi/2^m)$ by two strokes of length $2\sin(\pi/2^{m+1})$; then the loop starts again. To construct Koch curves, the same approach is used, but with simpler operations, since the stroke length has simply to be multiplied by r, and the replacement strokes' relative positions are the same throughout. Hence this punchline: When complexity is measured by the presently best algorithm's length in this particular alphabet, a *Koch curve is actually simpler than a circle.*

This peculiar ranking of curves by relative simplicity should not be taken seriously. Most notably, the contrary conclusion is reached if the alphabet is based on the compass and ruler—meaning that the circle is relabeled as "atomic." Nevertheless, as long as a sensible alphabet is used, any Koch curve is not only of finite complexity but simpler than most curves in Euclid.

Being fascinated with etymology, I cannot leave this discussion without confessing that I hate to call a Koch curve "irregular." This term is akin to *ruler,* and is satisfactory as long as one keeps to the meaning of *ruler* as an instrument used to trace straight lines: Koch curves are far from straight. But when thinking of a ruler as a king (= *rex,* same Latin root), that is, as one who hands down a set of detailed rules to be followed slavishly, I protest silently that nothing is more "regular" than a Koch curve. ■

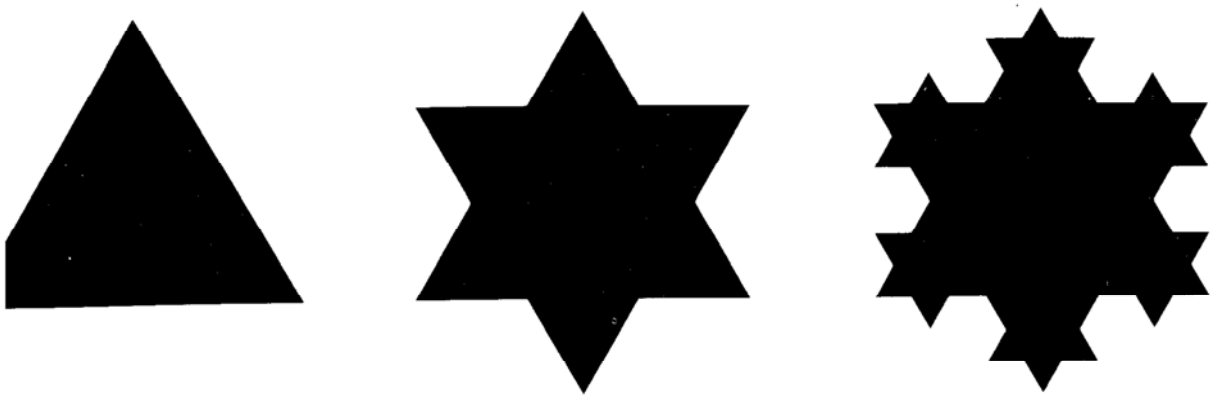


Plate 42 ¤ TRIADIC KOCH ISLAND OR SNOWFLAKE $\mathcal{K}$. ORIGINAL CONSTRUCTION BY HELGE VON KOCH (COASTLINE DIMENSION D=log 4/log 3~1.2618)

The construction begins with an "initiator," namely, a black △ (equilateral triangle) with sides of unit length. Then one pastes upon the midthird of each side a △-shaped peninsula with sides of length ⅓. This second stage ends with a star hexagon, or Star of David. The same process of addition of peninsulas is repeated with the Star's sides, and then again and again, ad infinitum.

Each addition displaces the points in an interval's midthird in a perpendicular direction. The triangular initiator vertices never move. The other 9 vertices of the Star of David achieve their final positions after a finite number of stages. Still other points are displaced without end, but move by decreasing amounts and eventually converge to limits, which define the coastline.

The island itself is the limit of a sequence of domains bounded by polygons, each of which contains the domain bounded by the preceding polygon. A photographic negative of this limit is part of Plate 45.

Observe that this and many other plates in the book represent islands or lakes rather than coastlines, and in general represent "solid areas" rather than their contours. This method takes fullest advantage of the fine resolution of our graphics system.

WHY A TANGENT CANNOT BE DEFINED HERE. Take as fixed point a vertex of the original △ and draw a cord to a point on the limit coastline. As this point converges clockwise to the vertex, the connecting cord oscillates within a 30° angle, and never tends to a limit one could call a clockwise tangent. The counterclockwise tangent is not defined either. A point where there is no tangent because clockwise and counterclockwise chords oscillate in well-defined angles is called *hyperbolic*. The points that $\mathcal{K}$ attains asymptotically fail to have a tangent for a different reason. ■

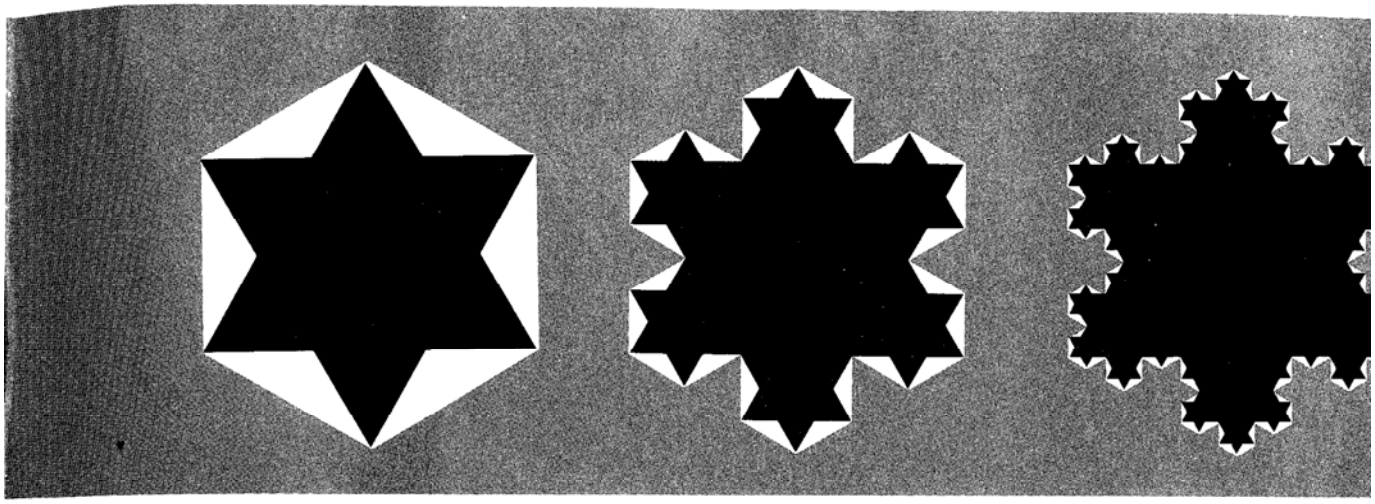


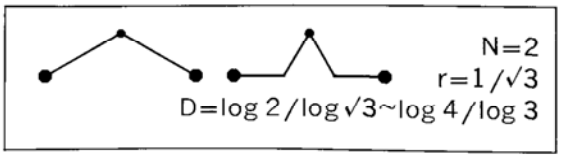
Plate 43 ¤ TRIADIC KOCH ISLAND OR SNOWFLAKE $\mathcal{K}$. ALTERNATIVE CONSTRUCTION BY ERNEST CESÀRO (COASTLINE DIMENSION D=log 4/log 3~1.2618)

An alternative construction of the Koch island is given in Cesàro 1905, a work of such charm as to make me forget the hard search for the original (and the irritation at later finding it reprinted in Cesàro 1964). Here is a free translation of a few ecstatic lines. "This endless imbedding of this shape into itself gives us an idea of what Tennyson describes somewhere as the *inner* infinity, which is after all the only one we could conceive in Nature. Such similarity between the whole and its parts, even its infinitesimal parts, leads us to consider the triadic Koch curve as truly marvelous. Had it been given life, it would not be possible to do away with it without destroying it altogether for it would rise again and again from the depths of its triangles, as life does in the Universe."

Cesàro's initiator is a regular hexagon with sides of length $\sqrt{3}/3$. The surrounding ocean is in gray. Increasingly small Δ-shaped bays are squeezed in ad infinitum, the Koch island being the limit of *decreasing* approximations.

This method of construction and Koch's method described in Plate 42 are carried out in parallel in the present plate. In this way, the Koch coastline is squeezed between an inner and an outer teragon that grow increasingly close to each other. One can think of a cascade process starting with three successive rings: solid land (in black), swamp (in white), and water (in gray). Each cascade stage transfers chunks of swamp to either solid land or water. At the limit the swamp exhausts itself from a "surface" down to a curve.

MIDPOINT DISPLACEMENT INTERPRETATION. It involves the following generator and next step (the angle here is 120°)



When placed outside the inner kth teragon, it yields the outer kth teragon; when placed inside the outer kth teragon, it yields the inner (k+1)st teragon. This approach is useful in Plates 64 and 65, and in Chapter 25. ■

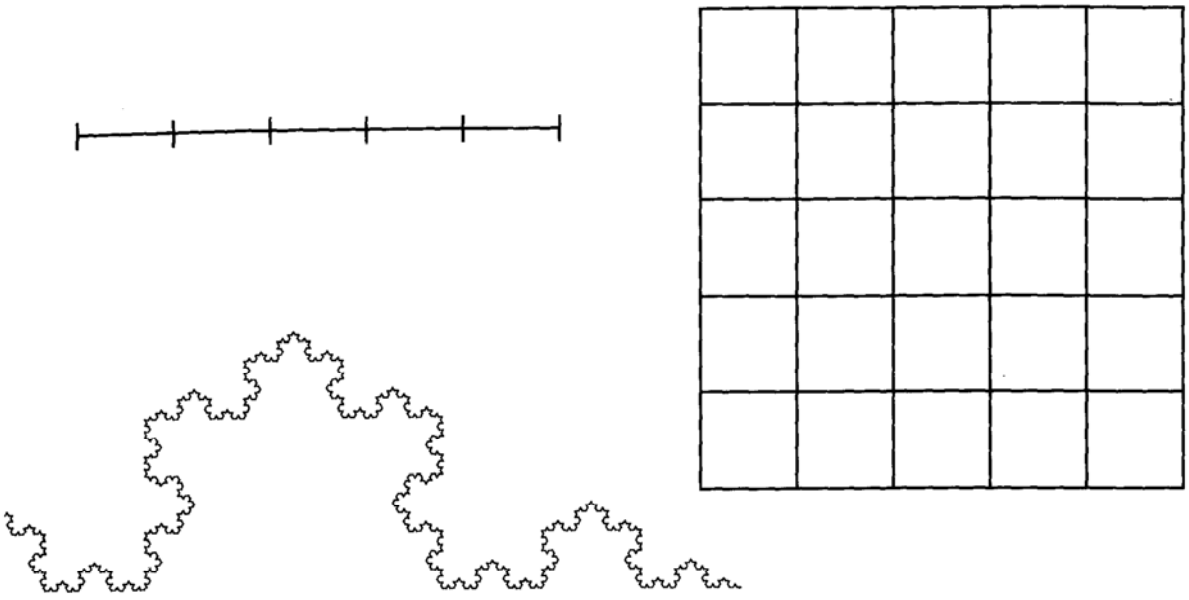


Plate 44 ¤ TWO KINDS OF SELF-SIMILARITY: STANDARD AND FRACTAL

The top Figures recall how, given an integer (here, $b=5$), a straight interval of unit length may be divided into $N=b$ subintervals of length $r=1/b$. Similarly, a unit square can be divided into $N=b^2$ squares of side $r=1/b$. In either case, $\log N/\log(1/r)$ is the shape's similarity dimension—a notion school geometry feels no need of pinpointing, since its value reduces to the Euclidean dimension.

The bottom Figure is a triadic Koch curve, one-third of a Koch coastline. It too can be decomposed into reduced-size pieces, with $N=4$ and $r=\frac{1}{3}$. The resulting similarity dimension $D=\log N/\log(1/r)$ is not an integer (its value is ~ 1.2618), and it corresponds to nothing in standard geometry.

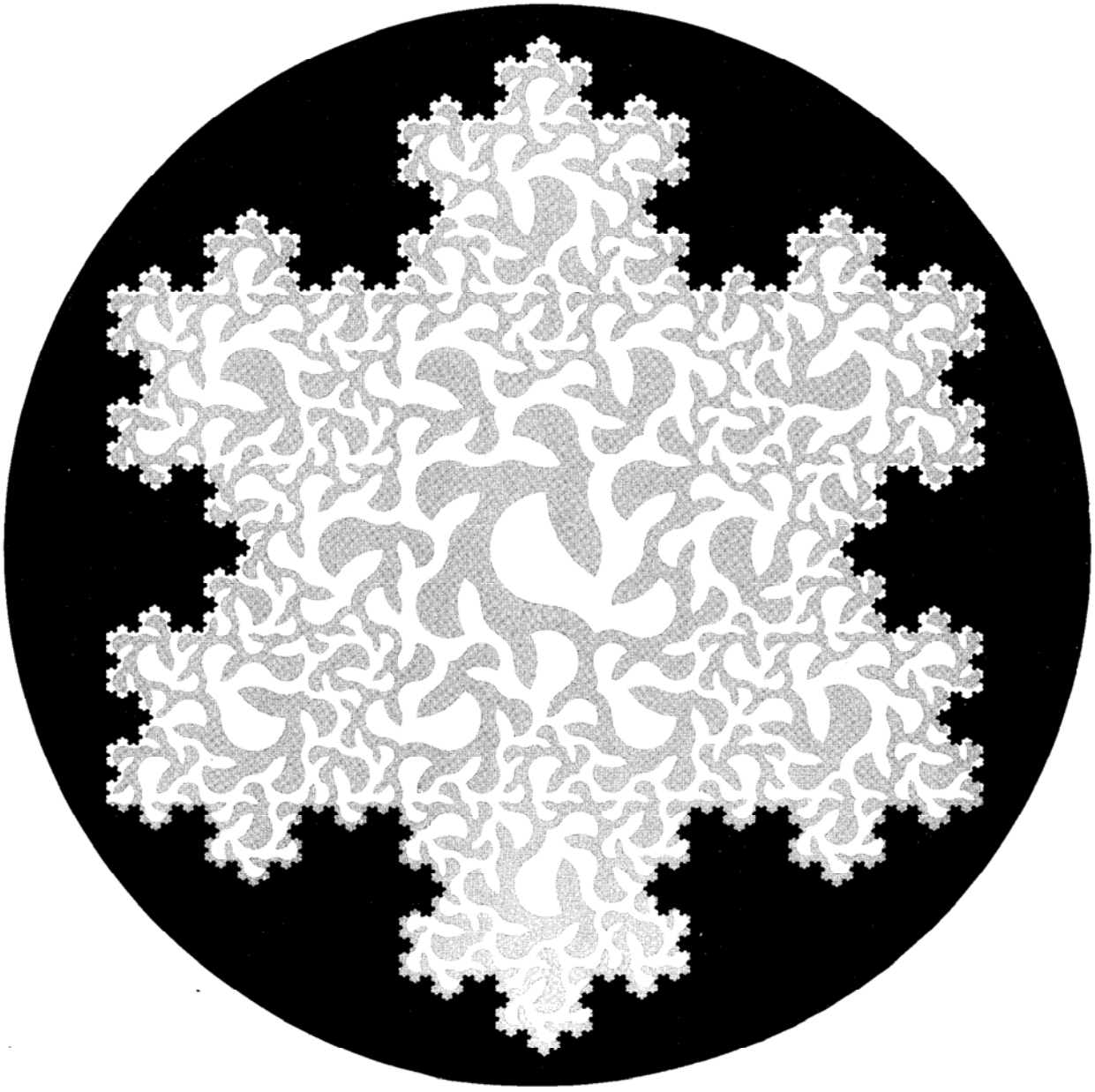
Hausdorff showed that D is of use in mathematics, and that it is identical to the Hausdorff, or fractal, dimension. My claim is that D is also vital in natural science. ■

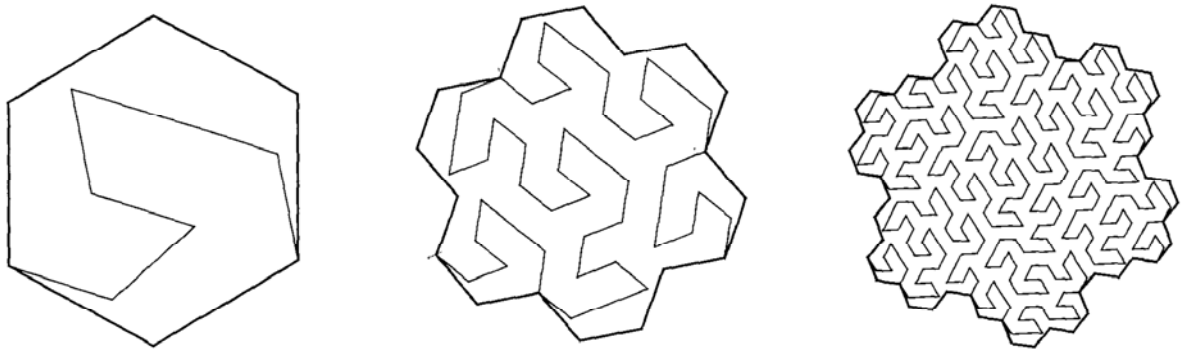
Plate 45 ¤ TRIADIC KOCH LAKE $\mathcal{K}$ (COASTLINE DIMENSION $D=\log 4/\log 3\sim 1.2618$)

The construction described in the captions of Plates 42 and 43 has been carried much further, and a photographic negative taken, yielding a lake rather than an island.

The peculiar pattern of gray "waves" that fills this lake is not haphazard. It is explained in Plates 68 and 69.

The coastline on this Plate is *not* self-similar, because a loop cannot be decomposed into the union of other loops. ◅ However, Chapter 13 uses the notion of self-similarity within an infinite collection of islands. ► ■





Plates 46 and 47 ¤ ALTERNATIVE KOCH ISLAND AND LAKE (COASTLINE DIMENSION D=log 9/log 7~1.1291)

This variant of the Koch island is due to W. Gosper (Gardner 1976): the initiator is a regular hexagon, and the generator is

N=3
1/r=√7
D=log 3/log (√7)~1.1291

PLATE 46. In this plate, several stages of construction of the "Gosper island" are drawn as a bold line "wrapping." The corresponding thin line "filling" is explained in Plate 70.

PLATE 47. This is an advanced construction stage of the "wrapping." The variable thickness "filling" is, again, explained in Plate 70.

Observe that, contrary to Koch's original, the present generator is symmetric with respect to its center point. It combines peninsulas and bays in such a way that the island's area remains constant throughout the construction. The same is true of the Koch curves up to Plate 57.

TILING. The plane can be covered using Gosper islands. This property is called *tiling*.

PERTILING. Moreover, the present island is self-similar, as is made obvious by using variable-widths hatching. That is, each island divides into seven "provinces" deducible from the whole by a similarity of ratio $r=1/\sqrt{7}$. I denote this property by the neologism *pertiling*, coined with the Latin prefix *per-*, as used for example in "to perfume" = "to penetrate thoroughly with fumes."

Most tiles *cannot* be subdivided into equal tiles similar to the whole. For example, it is a widespread source of irritation that hexagons put together do not quite make up a bigger hexagon. The Gosper flake fudges the hexagon just enough to allow exact subdivision into 7. Other fractal tiles allow subdivision into different numbers of parts.

FRANCE. A geographical outline of unusual regularity often described as *the* Hexagon, namely the outline of France, resembles a hexagon less than it resembles Plate 47 (although Brittany is undernourished here.)

◅ REASON WHY A TANGENT CANNOT BE DEFINED AT ANY POINT OF THESE COASTLINES. Fix any point that the coastline attains after a finite number of stages of construction, and join it by a cord to a moving point on the limit coastline. As the moving point approaches the fixed point along the limit coastline, either clockwise or counter-clockwise, the cord's direction winds without end around the fixed point. Such a point is called *loxodromic*. ► ■

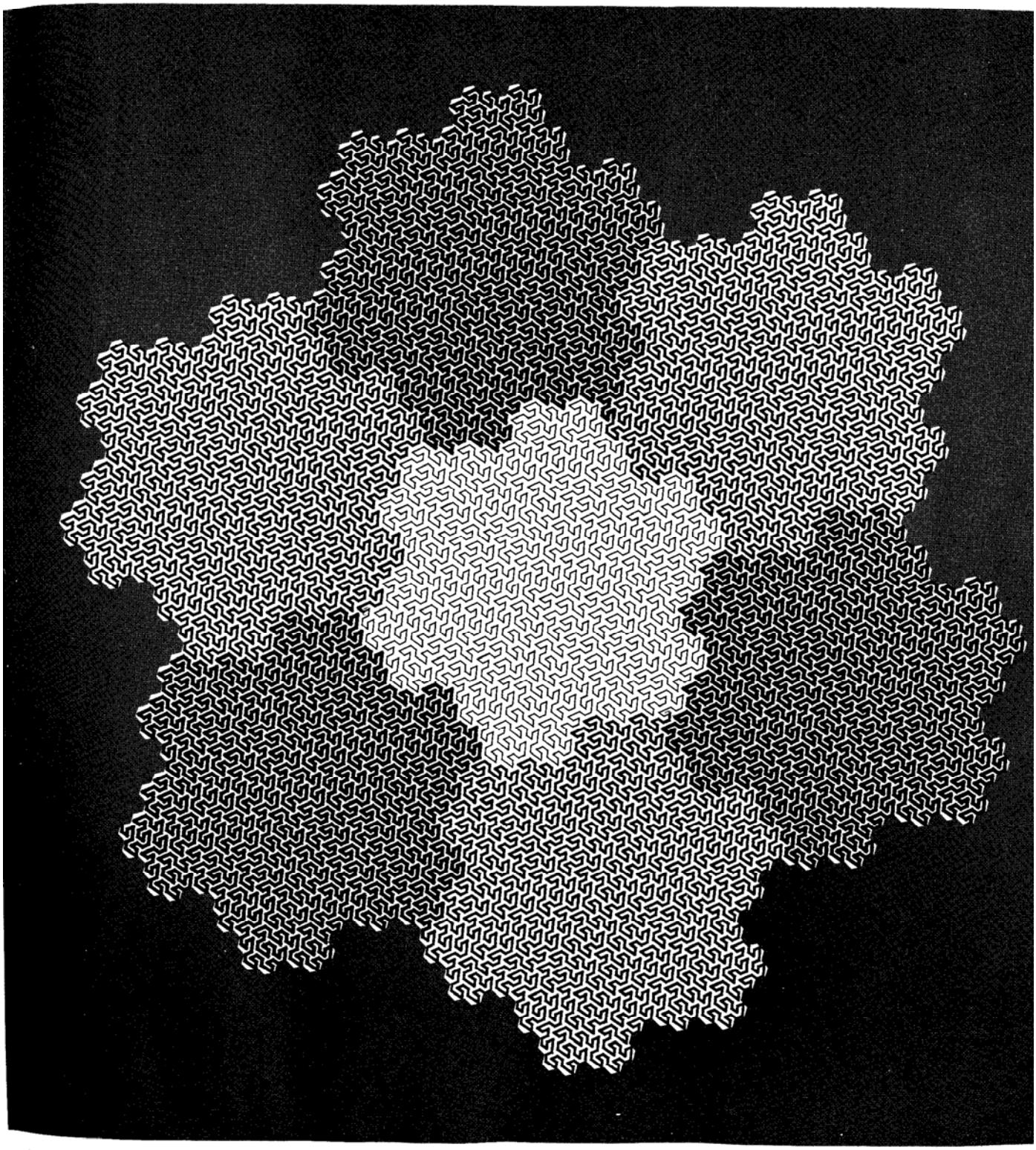


Plate 49 ¤ ALTERNATIVE KOCH ISLANDS AND LAKES (COASTLINE DIMENSIONS FROM 1 TO D=log 3/log √~1.3652)

Throughout this sequence of fractal curves, the initiator is a regular polygon with M sides, and the generator is such that N=3 and that the angles between the first and second and second and third legs are both $\theta=2\pi/M$. Plates 46 and 47 had involved the special value M=6 (not repeated here), and the value M=3 is discussed in Plate 72. The present plate exhibits advanced teragons for the values M=4, 8, 16, and 32, in the form of nested lakes and islands. For example, M=4 corresponds to the generator

N=3
1/r=√5
D=log 3/log √5~1.3652

The shading on the central island (M=4) is explained in Plates 72 and 73.

Were this pattern extended to M=∞, it would converge to a circle. As we move in, the figures "shrivel," first gradually, then by rapid jumps. The next stage of shriveling would lead to M=3, but the corresponding curve is no longer self-avoiding. We meet it later, in Plates 72 and 73.

A CRITICAL DIMENSION. When the initiator is [0,1], the angle θ may take any value from 180° down to 60°. There is a critical angle θ_{crit}, such that the "coastline" is self-avoiding if, and only if, $\theta>\theta_{crit}$. The corresponding D_{crit} is a *critical dimension* for self-intersection. The angle θ_{crit} is close to 60°.

GENERALIZATION. The constructions of Plates 46 to 57 are easily generalized as follows. Let the generators that are shown be called straight (S), and define the flipped generator (F) as the mirror image of the straight generator in the line y=0. Each stage of the construction must use the same generator throughout, either S or F, but different stages may select different generators. These plates, and more which follow, use S throughout, but other infinite sequences of S and F yield immediate variants.

◅ If F and S alternate, the formerly loxodromic points become hyperbolic, as in the Koch curve. ►■

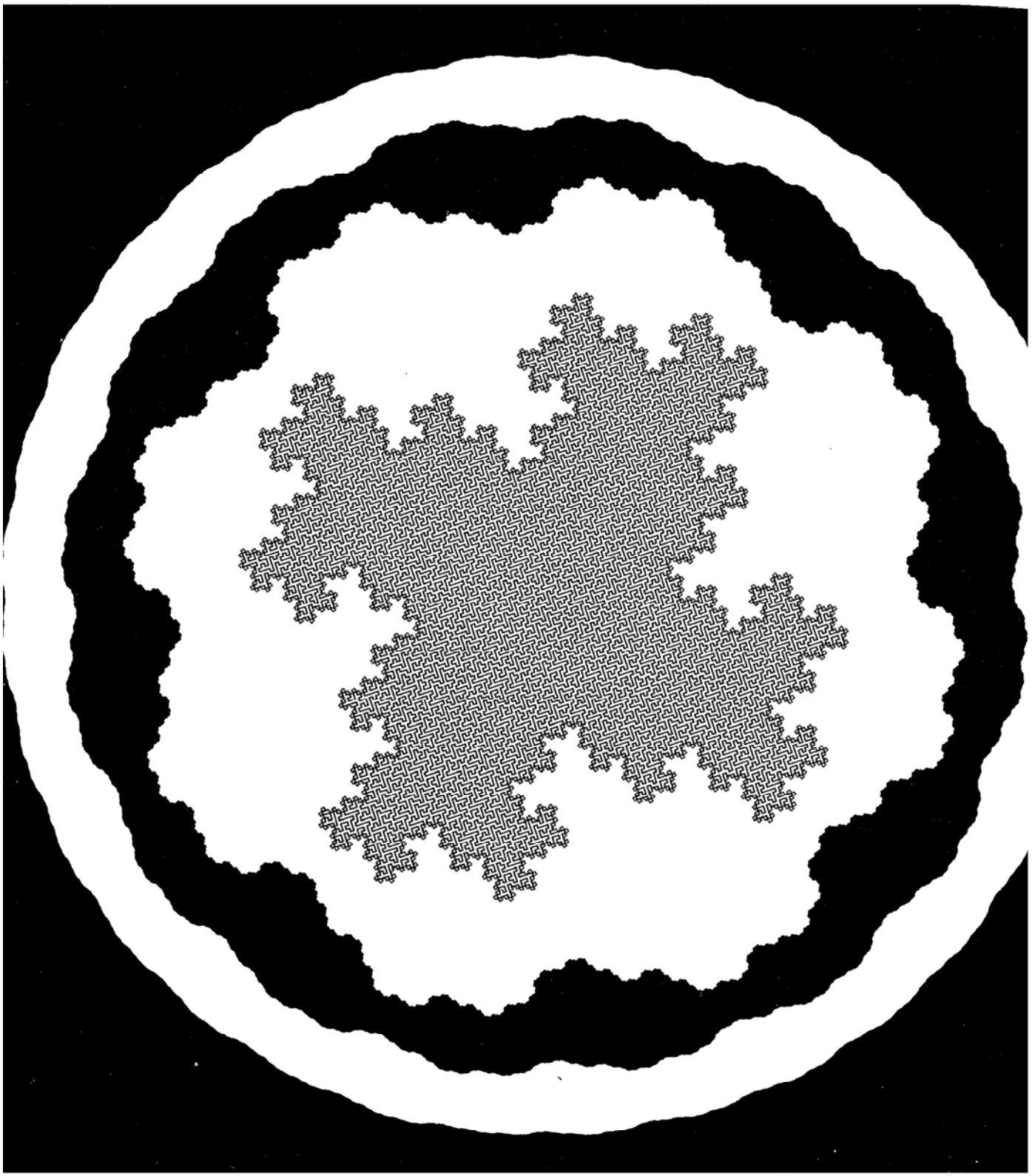
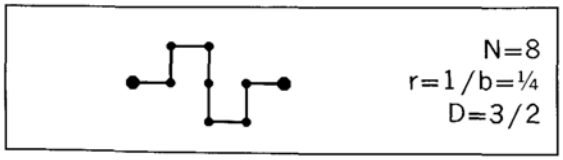


Plate 51 ¤ A QUADRIC KOCH ISLAND (COASTLINE DIMENSION D=3/2=1.5000)

Plates 49 to 55 show several Koch constructions initiated with a square (hence the term *quadric*). One advantage is that one can experiment with these constructions even when the available graphic systems are crude. ◅ Another advantage is that quadric fractal curves lead on directly to the original Peano curve described on Plate 63. ►

PLATE 51. Here, the initiator is a square, and the generator is



As in Plates 46 to 49, the total island area remains constant throughout the succession of stages. Plate 51 shows two stages on a small scale, and the next on a larger scale.

In the last stage, enlarged even further, the detail shows as very thin and barely visible whiskers, but much would be lost perceptually if the graphics were less excellent, forcing us to omit this detail.

Both the teragons and the limit curve involve no self-overlap, no self-intersection, and no self-contact. The same is true through Plate 55.

◅ One should not forget that the fractal in Plates 51 to 55 is the coastline; the land and sea are conventional shapes that have positive and finite areas. Page 144 mentions a case in which the "sea" alone has a well-defined area, being again the union of simple-shaped tremas, while the land has no interior point. ►

TILING AND PERTILING. The present island is decomposable into 16 islands reduced by the ratio of r=¼. Each is the Koch island built on one of the 16 squares forming the first stage of the construction.

◅ Chapters 25 and 29 show that D=3/2 is also encountered for various Brown functions. Hence this value is easy to obtain with random curves and surfaces. ■

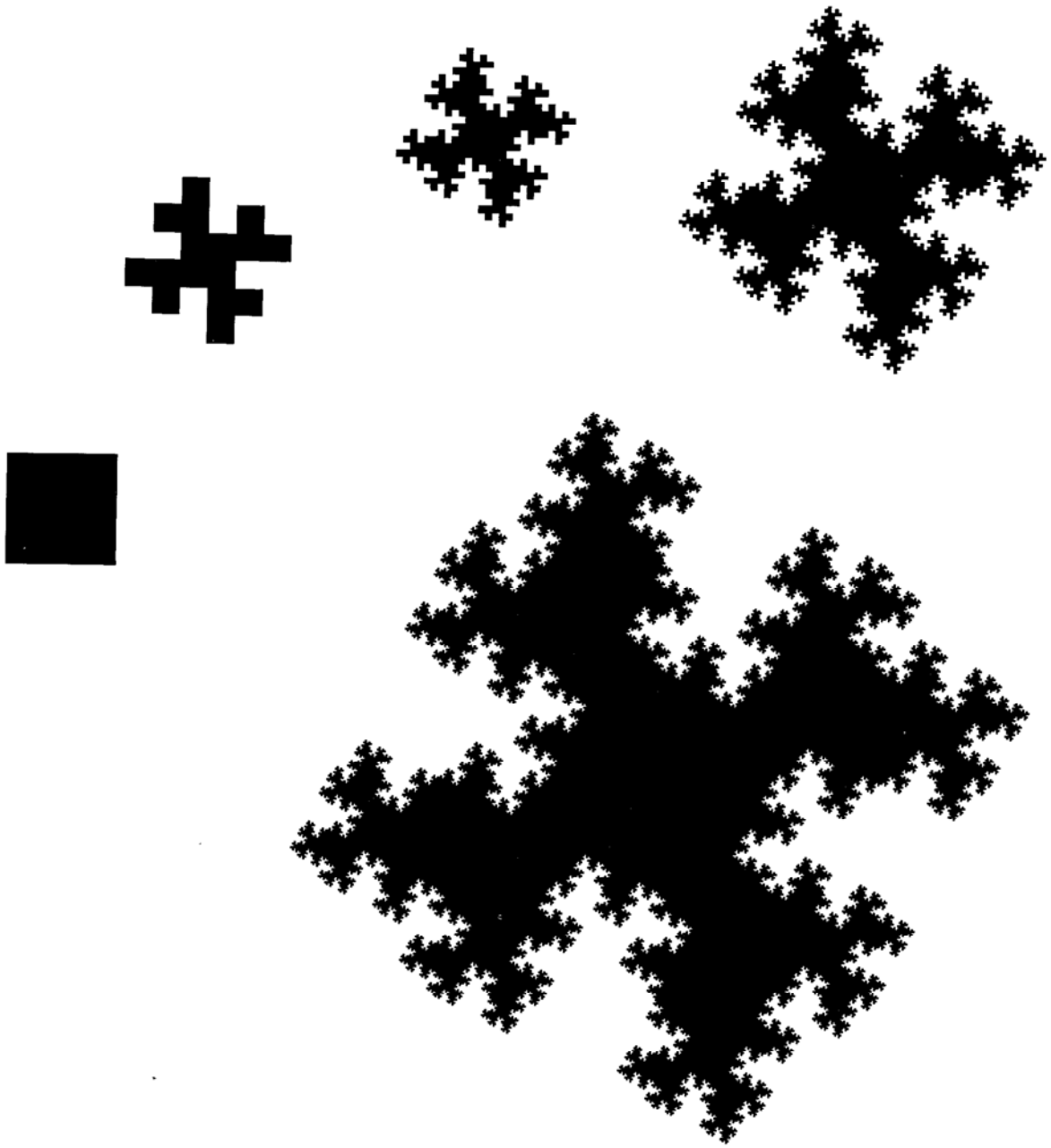
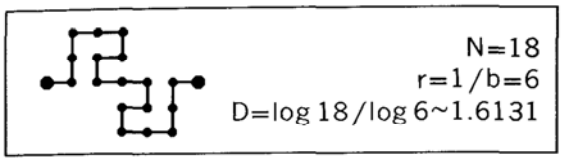


Plate 53 ¤ A QUADRIC KOCH ISLAND (COASTLINE DIMENSION D=log 18/log 6~1.6131)

The initiator is again a square, and the generator is



The fact that the form of the quadric Koch islands in the present portfolio of illustrations depends very markedly upon D is significant. However, their having roughly the same overall outline is due to the initiator's being a square. When the initiator is an M-sided regular polygon (M>4), the overall shape looks smoother, increasingly so as M increases. A genuine link between overall form and the value of D will not enter until Chapter 28, which deals with random coastlines that effectively determine the generator and the initiator at the same time.

◅ MAXIMALITY. Another fact that contributes to the similarity of overall outline is that the quadric Koch curves in Plates 49 to 55 possess an interesting property of maximality. Consider all Koch generators that yield self-avoiding curves are traced on a square lattice made by straight lines parallel and perpendicular to [0,1], and in addition can be used with any initiator on the square lattice. We denote as *maximal* the generators that attain the highest possible value of N and hence of D. One finds that $N_{max}=b^2/2$ when b is even, while $N_{max}=(b^2+1)/2$ when b is odd.

◅ As the value of b increases, so does the maximal N, and so also does the number of alternative maximal polygons. Therefore, the limit Koch curve becomes increasingly influenced by the original generator. It also looks increasingly contrived, because the wish to achieve a maximal dimension without contact points imposes a degree of discipline that increases with D. It reaches its paroxysm in the next chapter, for the Peano limit D=2.

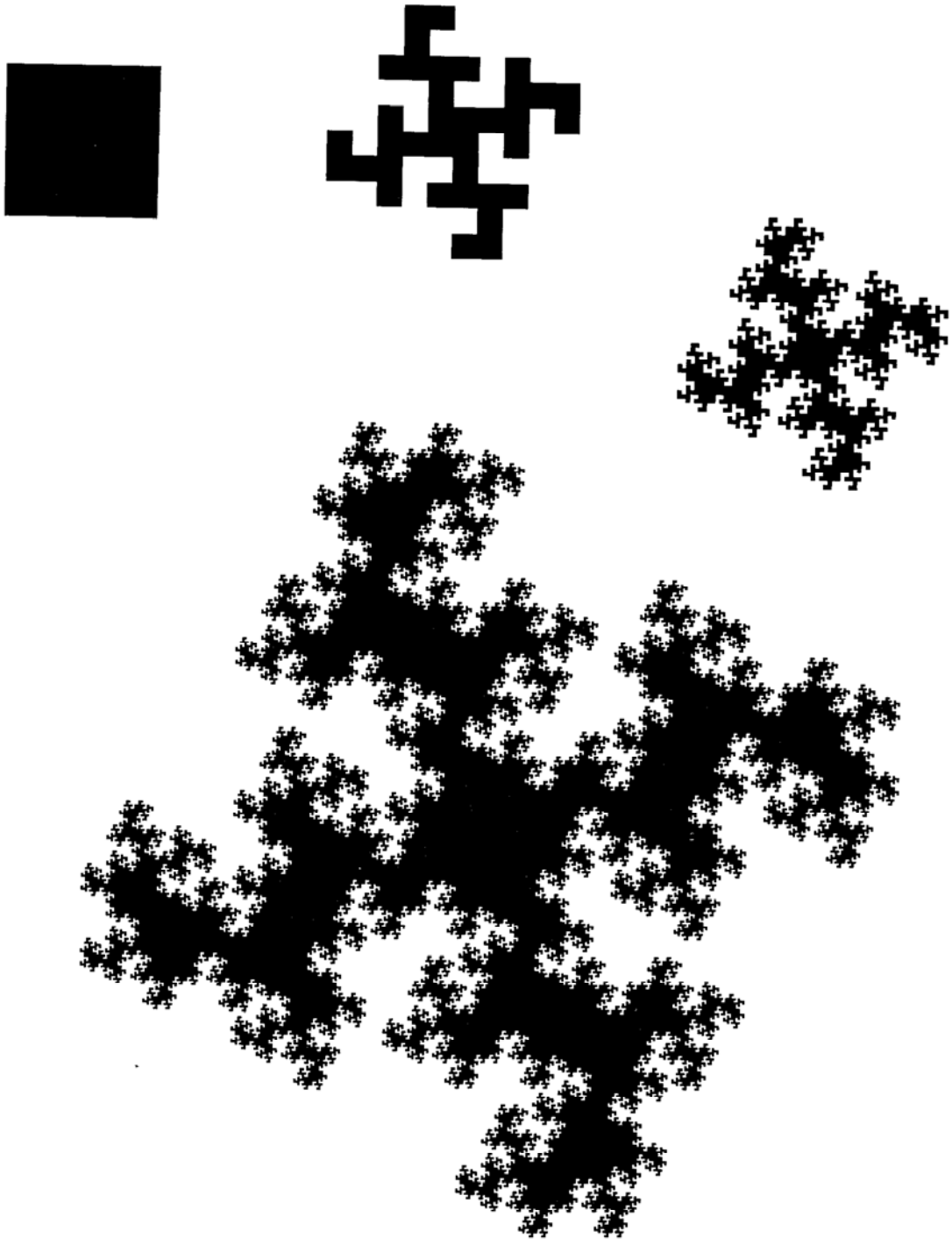
◅ LACUNARITY. Fractal curves sharing D but with different N and r may differ qualitatively from each other. The resulting parameter beyond D is discussed in Chapter 34. ■

CAPTION OF PLATE 55, CONTINUED

◅ In fact, the value of D is likely to depend on the fluid's initial energy, and on the size of the vessel in which dispersion is contained. A low initial energy would wither a disc-shaped blob into a curve with D close to 1 (Plate 49). A high initial energy in a small vessel might lead to more thorough dispersion, with planar sections more reminiscent of Plate 54 (D~1.7373) or even of the dimension D=2 (Chapter 8). See Mandelbrot 1976c.

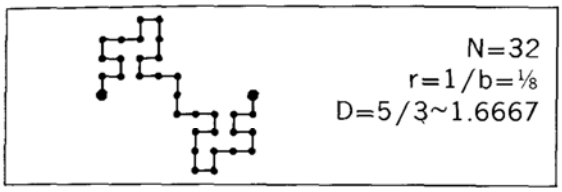
◅ If this last inference is valid, the next step would be to investigate the relation between initial energy and D, and to seek the lowest energy that yields D=2 in the plane, i.e., D=3 in space. When we examine the limit case D=2 (Chapter 7), we shall see that it differs qualitatively from D<2 because it allows ink particles that start far apart to come into asymptotic contact. ◅ Thus, I would not be at all surprised if it turns out that the turbulent dispersion is a single term representing two sharply distinct phenomena.

◅ POSTSCRIPT. Well after this plate had first appeared in the 1977 *Fractals*, Paul Dimotakis photographed thin sections of a turbulent jet dispersing in a laminar medium. The resemblance with the present plate is most gratifying. ► ■

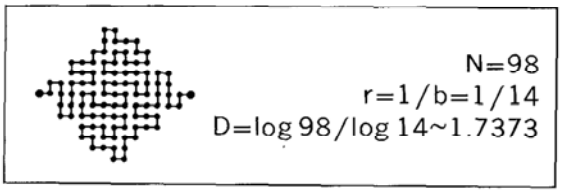


Plates 54 and 55 ¤ A QUADRIC KOCH ISLANDS (COASTLINE DIMENSIONS D=5/3~1.6667 AND D=log 98/log 14~1.7373)

Now the same construction as in Plate 49 is carried out with the following generators. In Plate 55,



and in Plate 54,



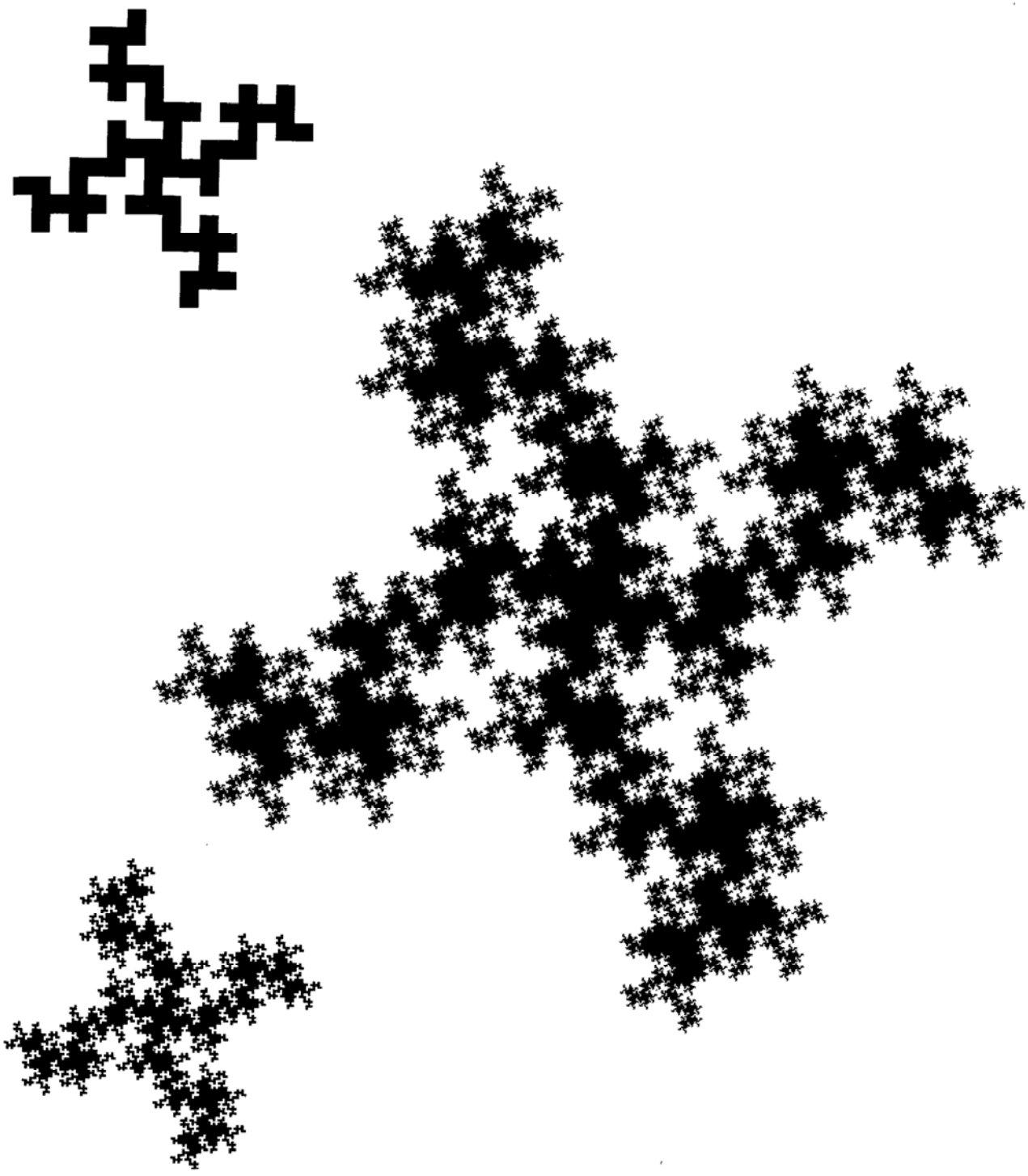
The causeways and the channels in these nightmarish marinas become increasingly narrow as one proceeds toward the peninsulas' tips or the bays' deepest points. In addition, these widths tend to narrow down as the fractal dimension increases, and "wasp waists" appear around D~5/3.

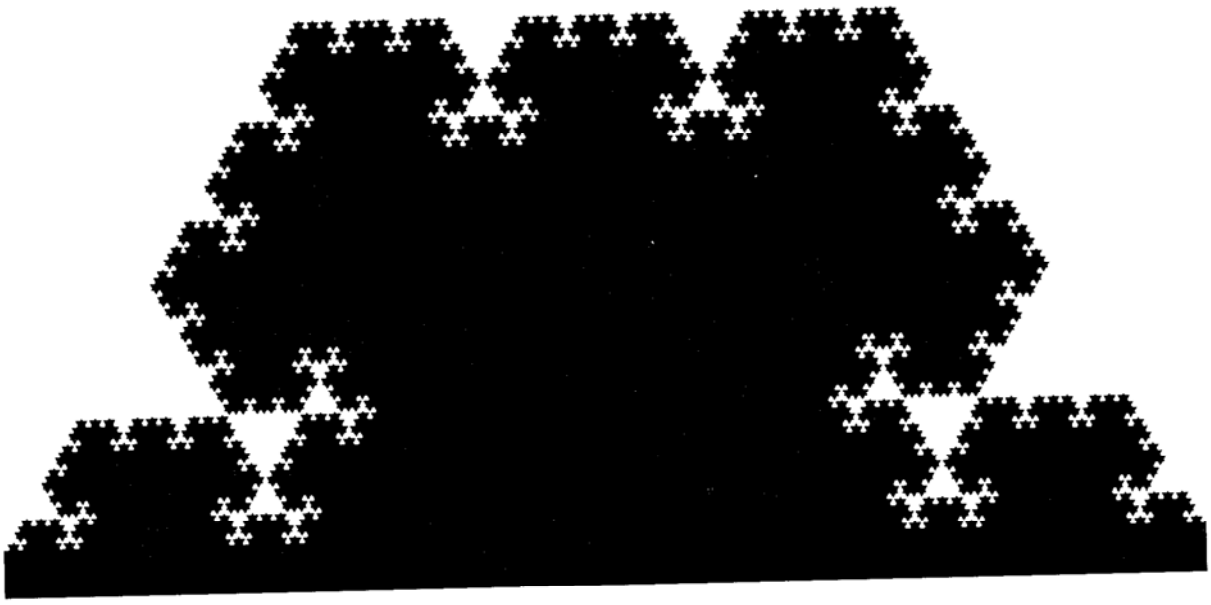
◅ DIGRESSION CONCERNING TURBULENT DISPERSION. I see an uncanny resemblance between the sequence of approximate fractals drawn in Plate 55, and the successive stages of turbulent dispersion of ink in water. Actual dispersion is of course less systematic, a feature one can mimic by invoking chance.

◅ One can almost see a Richardsonian cascade at work. A finite pinch of energy spreads a square ink blob around. Then the original eddy splits into smaller scale eddies, the effects of which are more local. The initial energy cascades down to ever smaller typical sizes, eventually contributing nothing but slight fuzziness to the outline of the final ink blob, just as in the following diagram from Corrsin 1959b.

◅ The conclusion that a Richardsonian cascade leads to a shape bounded by a fractal is inescapable, but the conclusion that D=5/3 is shaky. This value of D corresponds to planar sections of spatial surfaces with D=8/3, which occur often in turbulence. In the case of isosurfaces of scalars (studied in Chapter 30), D=8/3 is reducible to the Kolmogorov theory. Nevertheless, numerological analogies are not to be trusted.

THIS CAPTION CONTINUES ON PAGE 52



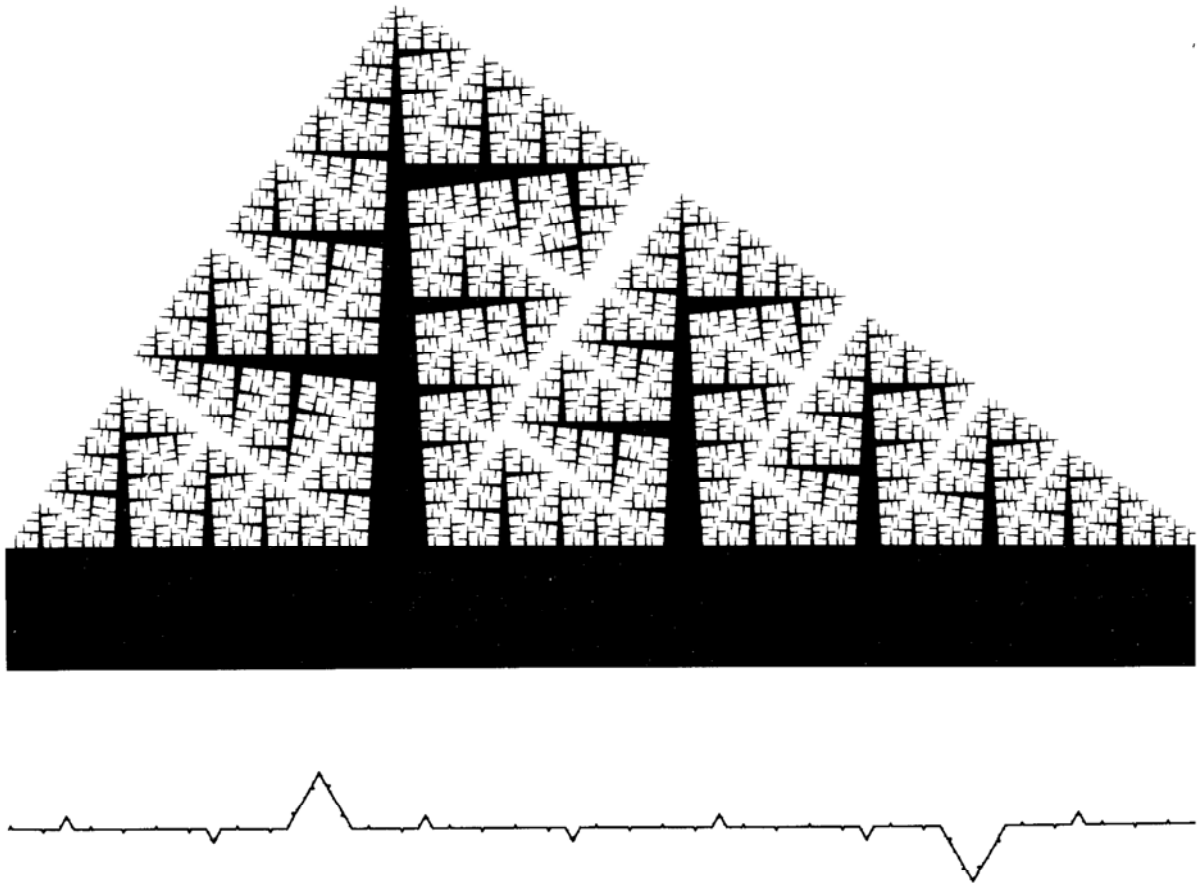


Plates 56 and 57 ¤ GENERALIZED KOCH CURVES AND SELF-SIMILARITY WITH UNEQUAL PARTS (D~1.4490, D~1.8797, D~1+ϵ)

These plates are constructed in the manner of Koch, except that the lengths of the generators' sides take different values r_m. Until now, we assume that the N "parts" into which our "whole" is divided all involve the same similarity ratio r. Using unequal r_m, the Koch curve becomes less relentlessly regular. Thus Plate 56 adds variety to the triadic Koch curve.

Note that in all this series of plates, the construction continues until it reaches details of a predetermined small size. When $r_m \equiv r$, this goal is reached after a predetermined number of construction stages, but here we need a variable number of stages.

The next task is to extend the notion of similarity dimension to this generalization of the Koch recursion. In a search for suggestions, let ordinary Euclidean shapes be paved with parts reduced in the respective ratios r_m. When D=1, the r_m must satisfy $\Sigma r_m = 1$, and, more generally, Euclidean shapes require



$\Sigma r_m{}^D=1$. Furthermore, in the case of fractals that can be split into equal parts, the familiar condition $Nr^D=1$ can be rewritten as $\Sigma r_m{}^D=1$. These precedents suggest forming the dimension-generating function, namely $G(D)=\Sigma r_m{}^D$, and defining D as its unique real root of $G(D)=1$. It remains to investigate whether or not said D coincides with the Hausdorff Besicovitch dimension. In every case I know of, it does.

EXAMPLES. Plate 56 has a D above Koch's original $\log 4/\log 3$. The top of Plate 57 has a D slightly below 2. As $D\to 2$, the coastline on this Figure tends toward the Peano-Pólya curve, a variant of the Peano curves examined in the next chapter. The resemblance between this Figure and a row of trees is not accidental, as seen in Chapter 17. Finally, the bottom of Plate 57 has a D slightly above 1. ■

7 ¤ Harnessing the Peano Monster Curves

When the end of Chapter 6 tackles generalized Koch curves that do not self-intersect, there is good reason for stopping short of D=2. When D reaches D=2, a profound qualitative change occurs.

We shall assume that the teragons do not self-intersect, although they may self-contact. Then one symptom of reaching D=2 is that points of self-contact become *inevitable* asymptotically. The major symptom is that it is *inevitable* that the limit should fill a "domain" of the plane, that is, a set that contains discs (filled in circles).

This double conclusion is *not* due to a corrigible lack of imagination on the part of mathematicians. It involves a fundamental principle, central to the 1875–1925 crisis in mathematics.

PEANO "CURVES," MOTIONS, SWEEPS

The corresponding limits, exemplified in upcoming plates, are called *Peano curves,* because the first is found in Peano 1890. They are also called *plane-filling curves*. For them, the formal definition of dimension by $\log N/\log(1/r)=2$ is justified, but for a disappointing reason. From the mathematical viewpoint, a Peano curve is merely an unusual way of looking at a domain or piece of plane, a set for which all the classical definitions yield the dimension 2. In other words, the term *plane-filling curve* should be avoided by careful writers.

Fortunately, most Peano "curves," including those obtained by a recursive Koch construction, are parametrized naturally by a scalar t, which may be called "time." In their case, we can (with no fear of the guardians of rigor) use the terms *Peano motions*, *plane-filling motions*, *tile sweeping motions*, or *tile sweeps* (tiles are discussed later in the chapter). We shall do so when it seems appropriate, but Essays need not attempt full consistency on any account.

THE PEANO CURVES AS MONSTERS

"Everything had come unstrung! It is difficult to put into words the effect that [Giuseppe]

Peano's result had on the mathematical world. It seemed that everything was in ruins, that all the basic mathematical concepts had lost their meaning" (Vilenkin 1965). "[Peano motion] cannot possibly be grasped by intuition; it can only be understood by logical analysis" (Hahn 1956). "Some mathematical objects, like the Peano curve, are totally non-intuitive..., extravagant" (Dieudonné 1975).

THE PEANO CURVES' TRUE NATURE

I claim that the preceding quotes merely prove that no mathematician ever examined a good Peano graph with care. An unkind observer could say these quotes demonstrate a lack of geometric imagination.

I assert to the contrary that, after Peano teragons are observed attentively, letting one's thoughts wander about, it becomes very difficult *not* to associate them with diverse aspects of Nature. This chapter takes up the self-avoiding curves, those whose teragons *avoid* self-contact. Chapter 13 takes up teragons that self-contact *moderately*. Teragons that fill a lattice (e.g., lines parallel to the axes and having integer coordinates) must first be processed to eliminate the self-contacts.

RIVER AND WATERSHED TREES

Examining diverse Peano teragons, I saw in each case a set of two *trees* (or sets of trees) possessing an endless variety of concrete interpretations. They are particularly conspicuous on the "snowflake sweep" Peano curve I designed, Plate 69. It is, for example, easy to visualize this Plate as a collection of bushes rooted side by side along the bottom third of a Koch snowflake, and creeping up a wall. Alternatively, one may choose to be reminded of the boldly emphasized outline of a collection of rivers meandering around, and eventually flowing into a river that follows the snowflake's bottom. This last interpretation suggests immediately that the curves that separate the rivers from each other combine into watershed trees. And of course, the labels *river* and *watershed* can be interchanged.

This *new* rivers-watersheds analogy is so obvious *after the fact* that it lays to rest any notion that the Peano curve is necessarily pathological. As a matter of fact if a tree made of rivers of vanishing width is to drain an area thoroughly, it *must* penetrate everywhere. One who follows the rivers' combined bank performs a plane-filling motion. Ask any child for confirmation!

Helped by the intuition garnered from Plate 68, it would be difficult not to see analogous conjugate networks in every Peano teragon. Even the crude island of Plate 63 begins to make intuitive sense. The thin fingers of water that penetrate it cannot be viewed as a marina, however exaggerated, but can be viewed as branching rivers.

When rivers give rise to a proper science, it should be called *potamology*—Maurice Pardé's coinage from ποταμος (= river) and λογος. But sober usage merges the study of

rivers into the science of water, hydrology, into which this Essay makes many incursions.

MULTIPLE POINTS ARE UNAVOIDABLE IN TREES, HENCE IN PEANO MOTIONS

Suddenly many mathematical properties of Peano curves become obvious too. To account for double points, assume one starts on a river's shore in a Peano river tree, and moves upstream or downstream, making a detour for the slightest branch (moving ever faster as one gets to finer branches). It is clear that one will eventually face the point of departure from across the river. And since the limit river is infinitely narrow, one will effectively return to the starting point. Thus, double points in a Peano curve are *inevitable*, not only from a logical but also from an intuitive viewpoint. Furthermore, they are *everywhere dense*.

Also, it is inevitable that some points be visited more than twice, because a point where rivers join is one where at least three points of the bank coincide. When all points of confluence involve only two rivers, there is no point of multiplicity above three. On the other hand, one can do without points of multiplicity of three if one agrees to have points of higher multiplicity.

All the assertions in the preceding paragraphs have been proven, and, since the proofs are delicate and led to controversy, the properties themselves seem "technical." But the contrary is the case. Who would continue to argue that a purely logical approach toward them is preferable to my own intuitive one?

Typically, a Peano curve's rivers are not standard shapes but fractal curves. This is fortunate for the needs of modeling, because every argument in Chapter 5 to the effect that geographic curves are nonrectifiable applies equally well to river banks. In fact, the Richardson data include frontiers that follow rivers or watersheds. And rivers are involved in the quote from Steinhaus 1954. As to rivers' drainage basins, they are surrounded by closed curves akin to island coastlines, made of portions of watershed. Each basin is the juxtaposition of partial basins and is crisscrossed by the rivers themselves, but plane-filling curves that are bounded by fractal curves display all the structure we need.

PEANO MOTION AND PERTILING

Taking the original Peano curve (Plate 63), develop t in the counting base N=9, in the form $0.\tau_1\tau_2\ldots$. Times sharing the same first "digit" are mapped on the same ninth of the initial square, those with the same second digit on the same 9^2-th, etc. Thus, the tiling of [0,1] into 9-th maps on a tiling of the square. Successive 9-ths of the linear tiles map on successive planar subtiles. And the interval's property of being pertiling (page 46), i.e. subdivisible recursively and ad infinitum into smaller tiles similar to [0,1], is mapped on the square. Alternative Peano motions, due to E. Cesàro, G. Pólya and others, map this property on diverse pertilings of the triangle.

More generally, most Peano motions generate pertilings of the plane. In the simplest case, there is a base N, and one starts with a linear pertiling that consists of successive divisions into N-th. But the snowflake sweep of Plate 68-69 requires an irregular division of the [0,1] interval of t, into four subintervals of length $1/9$, then four of length $1/9\sqrt{3}$, one of length $1/9$, two of length $1/9\sqrt{3}$, and two of length $1/9$.

ON MEASURING DISTANCE BY AREA

Exquisite relationships, wherein length and area interchange, are a common occurrence in Peano motion, especially if it is *isometric,* meaning that a time interval $[t_1, t_2]$ maps on an *area* equal to the *length* $|t_1-t_2|$. (Most Peano motions are both isometric and pertiling, but these are distinct notions.) Calling the map of the time interval $[t_1,t_2]$ a planar Peano *interval* implies that, instead of measuring distances through a time, one may do so through an area. But we encounter a vital complication, because points that sit across from each other on different banks of a river coincide in space but are visited repeatedly.

The definition of "Peano distance" may involve only the order of the visits. Denoting the instants of first and last visits of P_1 and P_2 by t'_1 and t'_2 and by t''_1 and t''_2, the *left Peano interval* $\mathcal{L}\{P_1, P_2\}$ is defined as the map of $[t'_1, t'_2]$ and the *right Peano interval* $\mathcal{R}\{P_1, P_2\}$ is defined as the map of $[t''_1,t''_2]$. These intervals' lengths define the *left distance* and the *right distance* as $|\mathcal{L}\{P_1, P_2\}| = |t'_1-t'_2|$ and $|\mathcal{R}\{P_1, P_2\}| = |t''_1-t''_2|$. Each of these distances is additive, meaning for example that if three points P_1, P_2, and P_3 are left ordered according to the order of first visits, one has

$$|\mathcal{L}(P_1,P_3)|=|\mathcal{L}(P_1,P_2)|+|\mathcal{L}(P_2,P_3)|.$$

Alternate definitions of interval and distance distinguish between river and watershed points. Denote by t' and t'' the instants of first and last visit of P. P is a *river point* if the map of [t',t''] is bounded by P and watersheds. Successive visits of P face each other across rivers. P is a *watershed point* if the map of [t', t''] is bounded by P and rivers.

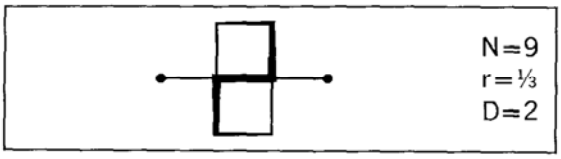
Furthermore, once a Peano curve is represented as the common shore of a river tree and a watershed tree, the paths that link P_1 and P_2 through rivers (resp., along watersheds) include a common minimal path. It is reasonable to follow this path in order to measure the distance between P_1 and P_2. Save for exceptional cases, the rivers' and watersheds' dimension D is strictly below 2 and strictly above 1. Hence the minimal path can be measured neither by length nor by area, but in typical cases it has a nontrivial Hausdorff measure in the dimension D.

MORE. Very important additional considerations on Peano motions are detailed in the captions that follow. ■

Plate 63 ¤ A QUADRIC KOCH CONSTRUCTION OF DIMENSION D=2: THE ORIGINAL PEANO CURVE, A SQUARE SWEEP

The *Peano plane-filling curve* in this plate is the original one. Giuseppe Peano's incredibly terse algorithm was graphically implemented in Moore 1900 (which receives undue credit in my 1977 *Fractals*). The present plate rotates Peano's curve by 45°, and by doing so brings it into the fold of Koch curves in the strict sense: the generator is always placed in the same way on the sides of the teragon obtained at the preceding stage.

The initiator here is the unit square (bounding the black box) and the generator is



Because this generator self-contacts, the resulting finite Koch islands are sets of black squares on a chunk from an infinite chessboard. And the nth Koch teragon is a grid of lines, a distance of $\eta=3^{-n}$ apart; they crisscross a square of area equal to 2 that becomes covered increasingly tightly as $k\rightarrow\infty$. It suffices to show one example of this dull design (next to the initial black box).

Three illustrations on the top of page 63 avoid ambiguity by cutting off the corners while leaving the total area invariant.

On the same scale, the fourth stage of this sequence would merge into 50% gray, but a larger drawing of one-fourth of the coastline can be followed unambiguously (at some risk of becoming seasick). It shows graphically what is meant by saying that the limit Koch curve fills the plane.

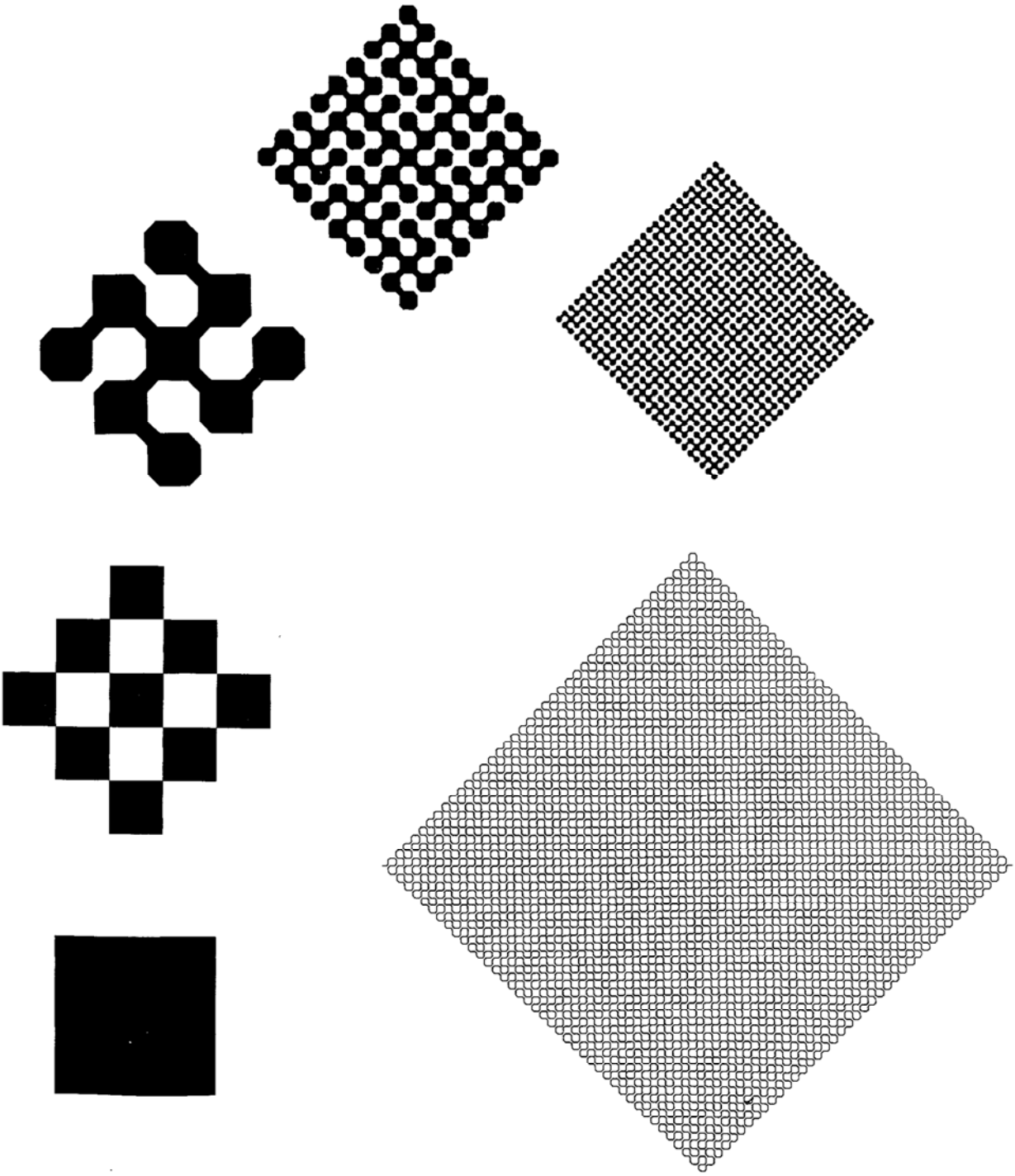
It would have been nice to be able to define a limit island in analogy to the Koch islands of Chapter 6, but in the present case it is impossible. A point chosen at random almost surely flips between being inland and in the ocean, without end. Advanced teragons are penetrated by bays or rivers so deeply and uniformly that a square of middling side x—such that $\eta<<x<<1$—divides between dry land and water in near equal proportions!

INTERPRETATION. The limit Peano curve establishes a continuous correspondence between the straight line and the plane. The fact that self-contacts are mathematically unavoidable is classical. The fact that they are valuable in modeling Nature is new to this work.

LONG-RANGE ORDER. Without knowing of the descending cascades that built our finite Peano curves, one would be baffled by the extraordinary long-range order that allows these curves to avoid not only self-intersection but also self-contact. Any lapse in discipline would make the latter very likely.

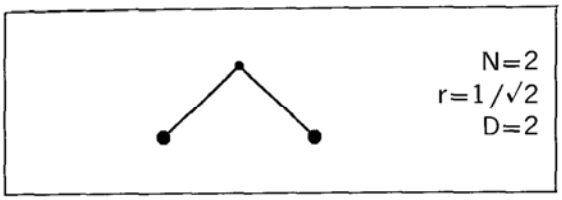
◅ And total breakdown of discipline makes endlessly repeated self-intersection almost certain, since a totally undisciplined Peano curve is Brownian motion, mentioned in Chapter 2 and explored in Chapter 25.

◅ LIOUVILLE THEOREM AND ERGODICITY. Mechanics represents the state of a complex system by a single point in a "phase space." Under the equations of motion, every domain in this space is known to behave as follows: its measure (hyper-volume) remains invariant (Liouville theorem), but its shape changes and it disperses and fills all the space available to it with increasing uniformity. Clearly, both of these characteristics are echoed by the behavior we impose upon the black domain in the present Peano construction. It is interesting, therefore, to dig deeper, by observing that in many simplified "dynamical" systems that allow a detailed study each domain disperses by transforming into an increasingly long and thin ribbon. It would be interesting to see whether other systems' dispersion proceeds through Peano-like trees instead of ribbons. ► ■



Plates 64 and 65 ¤ QUADRIC KOCH CONSTRUCTIONS OF DIMENSION D=2: CESÀRO'S AND POLYA'S TRIANGLE SWEEPS, AND VARIANTS

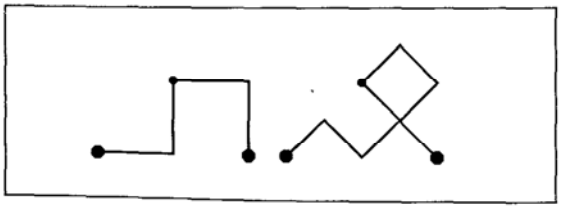
The simplest generator one could imagine is made of N=2 equal intervals making an angle θ that satisfies $90° \leq \theta \leq 180°$. The limit case $\theta=180°$ generates a straight interval; the case $\theta=120°$ (illustrated in the caption of Plate 43) generates the triadic Koch curve (among others). The limit case $\theta=90°$ is



This generator gives rise to an uncanny number of different Peano curves, according to the initiator's shape, and the rule of placement of the generator upon the preceding teragon. Plates 64 to 67 examine a few notable examples.

◅ In addition, Chapter 25 obtains Brownian motion by randomizing the class of all Peano curves with these N and r ►.

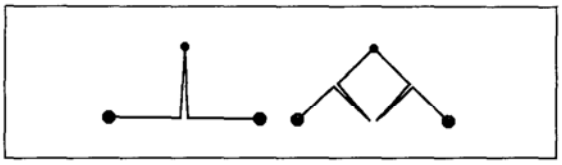
PÓLYA'S TRIANGLE SWEEP. The initiator is [0,1], the generator is as above, and it alternates between the right and the left of the teragon. The first position also alternates. The early construction stages yield the following



The teragons are pieces of square graph paper contained within a right isosceles triangle whose *side* is [0,1]. The limit curve sweeps this triangle.

PLATE 64. PÓLYA SWEEP OVER A RIGHT NONISOSCELES TRIANGLE. The generator is changed to be made of two unequal orthogonal intervals. Guessing the processing chosen to avoid self-contact is left to the reader as an exercise.

CESÀRO'S TRIANGLE SWEEP. The initiator is [1,0], the generator is again as above, and the next two construction stages are as follows (for the sake of clarity, the drawing refers to $\theta=85°$ instead of $\theta=90°$).



Thus, in all the odd-numbered construction stages, the generator is positioned to the right, yielding as teragon a grid of lines parallel to the initiator's diagonals. And in all the even-numbered stages, the generator is positioned to the left, yielding as teragon a grid of lines parallel to the initiator's sides. Asymptotically, this curve fills a right isosceles triangle whose *hypotenuse* is [0,1].

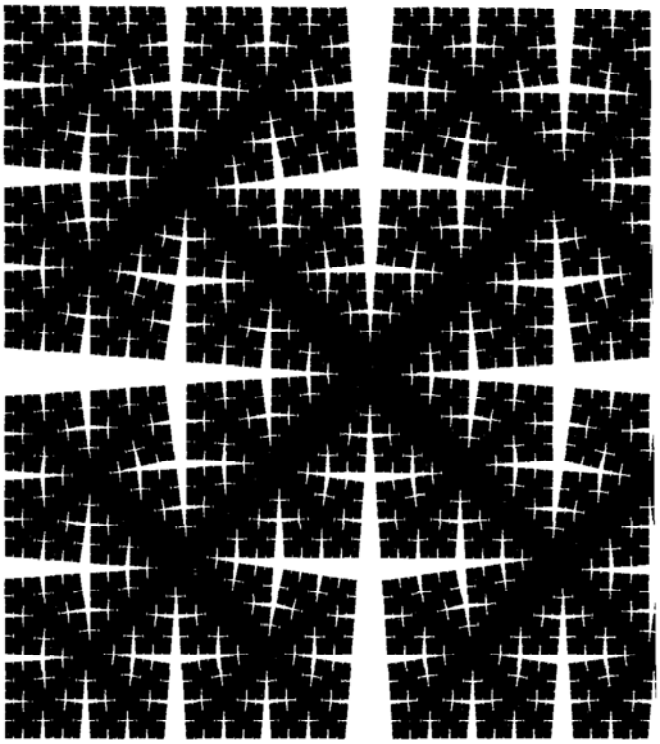


PLATE 65. This plate represents a square sweep obtained by adding the Cesàro sweeps initiated by [0,1] and [1,0]. (Again, $\theta=85°$ instead of 90° for the sake of clarity.)

SELF-OVERLAP. Each interval in the grids covered by the Cesàro teragons is covered *twice*. Not only the construction is self-contacting, but it is self-overlapping.

"EFFICIENCY" OF PLANE FILLING. AN EXTREMAL PROPERTY OF THE PEANO-CESÀRO DISTANCE. The Peano curve of Plate 63 maps [0,1] on the square of diagonal [0,1] and area ½. The same shape is covered by the Pólya curve. But the Cesàro curve fills a right isosceles triangle of hypotenuse [0,1] and area ¼. To cover the whole square, Cesàro must add the maps of [1,0] and [0,1]. Thus, the Cesàro curve in the less "efficient," of the two. As a matter of fact, it is the least efficient non-self-intersecting Peano curve on a square lattice. But this fact endows it with a redeeming virtue: the left or right Peano distance (see p. 61) between two points P_1 and P_2 is at least equal to the square Euclidean distance:

$$|\mathcal{L}(P_1,P_2)| \geq |P_1P_2|^2; |\mathcal{R}(P_1,P_2)| \geq |P_1P_2|^2$$

For other Peano curves, the difference between Peano and Euclid distance may take either sign.

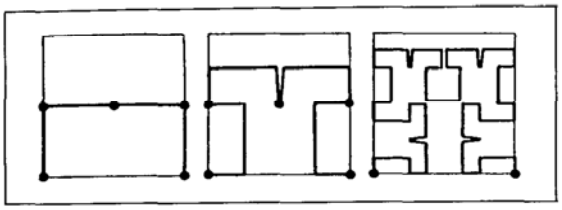
KAKUTANI-GOMORY PROBLEM. After selecting M points P_m in the square $[0,1]^2$, Kakutani (private communication) investigates the expression $\inf\Sigma|P_mP_{m+1}|^2$, where the infinitum is taken over all the chains that join the P_m in sequence. He proves that $\inf\leq 8$, but conjectures that this bound is not the best one. Indeed, R. E. Gomory (private communication) obtains the improved bound $\inf\leq 4$.

The proof uses the Peano-Cesàro curve, as follows. (A) Add the square's corners if they are not already among the P_m. (B) Rank the M points P_m in the order they are first visited by the string of four Peano-Cesàro curves drawn inside the square, along its sides. (C) Observe that, by lengthening the chain in step (A), we did not decrease $\Sigma|P_mP_{m+1}|^2$. (D) Observe that each addend $|P_mP_{m+1}|^2$ is not decreased if replaced by $|\mathcal{L}(Z_m,Z_{m+1})|$. (E) Observe that $\Sigma|\mathcal{L}(Z_m,Z_{m+1})| = 4$. If different Peano curves were used, steps (B) and (D) would be invalid. ■

Plates 66 and 67 ¤ A SQUARE SWEEP AND THE DRAGON SWEEP

The generator is the same here as in Plates 64 and 65, but seemingly slight changes in other rules have lasting consequences.

A LATER SQUARE SWEEP BY PEANO. The initiator is [0,1], but the second, fourth and sixth construction stages are changed to

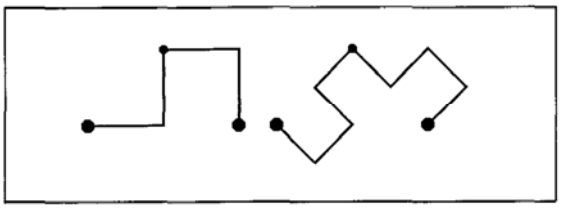


EFFICIENCY. AN EXTREMAL PROPERTY. This curve fills a domain of area equal to 1, while the curves of Plates 64-65 and the dragon curve to be below covers ½ or ¼. When the teragons lie on an orthogonal lattice, the covered area cannot exceed 1. It reaches this maximum whenever the teragons are self-avoiding. In other words, absence of self-contact is more than a matter of esthetics, and a self-contacting curve whose self-contacts are rounded off, as in Plate 63, does *not* become equivalent to a self-avoiding Koch curve.

By taking the odd numbered stages of the present square sweep, then joining the midpoints of the teragons' successive intervals to avoid self-contact, one falls back on a Peano curve due to Hilbert.

PLATE 67. A CURVE SWEEPING A RIGHT TRAPEZOID. The generator is changed to be made of two unequal orthogonal intervals. The processing to avoid self-contact is the same as in the preceding plate.

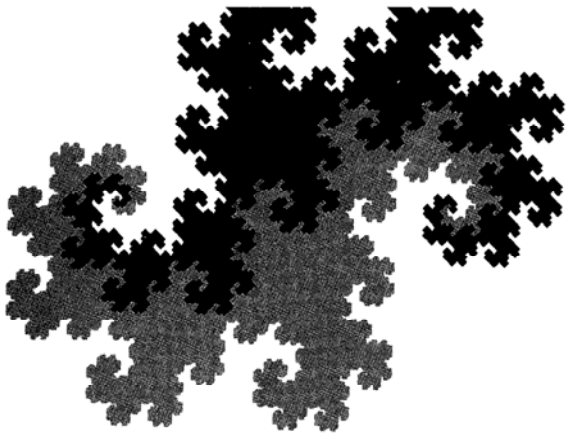
THE HARTER-HEIGHTWAY DRAGON. (See Gardner 1967, Davis & Knuth 1970.) Here the initiator is [1,0], the generator is as above, and it alternates between the right and the left of the teragon. The only difference with the Polya triangle sweep is that the first position is always to the right at every stage of construction, early stages being as follows



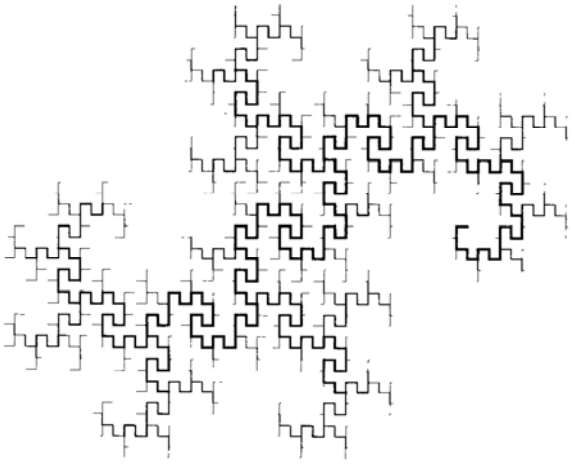
The consequences of this change are dramatic, since a mature stage looks like this

On this illustration, the curve itself has become indistinct, and we see only its boundary, called *dragon curve*. Thus, this Peano curve deserves to be called *dragon sweep*. As any Koch curve initiated by [0,1], the dragon is self-similar. But in addition it is seen to be segmented into portions, which join at wasp waists. The sections are similar to one another, but not to the dragon itself.

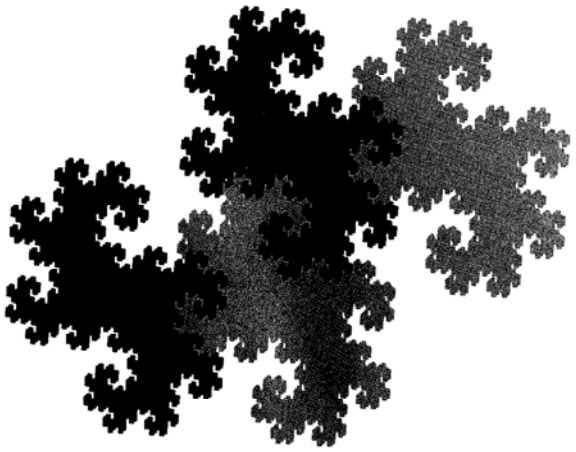
TWINDRAGON. The 1977 *Fractals* points out that, with the dragon's rules of construction, a more natural initiator is [0,1] followed by [1,0], and terms the shape that is swept as a result, a *twindragon*. This shape is encountered number representations, Knuth 1980. It looks like this (one component dragon is in black and the other is in gray).



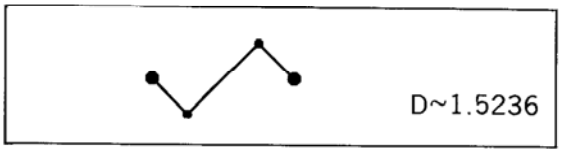
TWINDRAGON RIVER. After the streams near the source are erased (for legibility), the river tree of a twindragon looks like this.



A twindragon can be tiled by reduced size replicas of itself, like this.

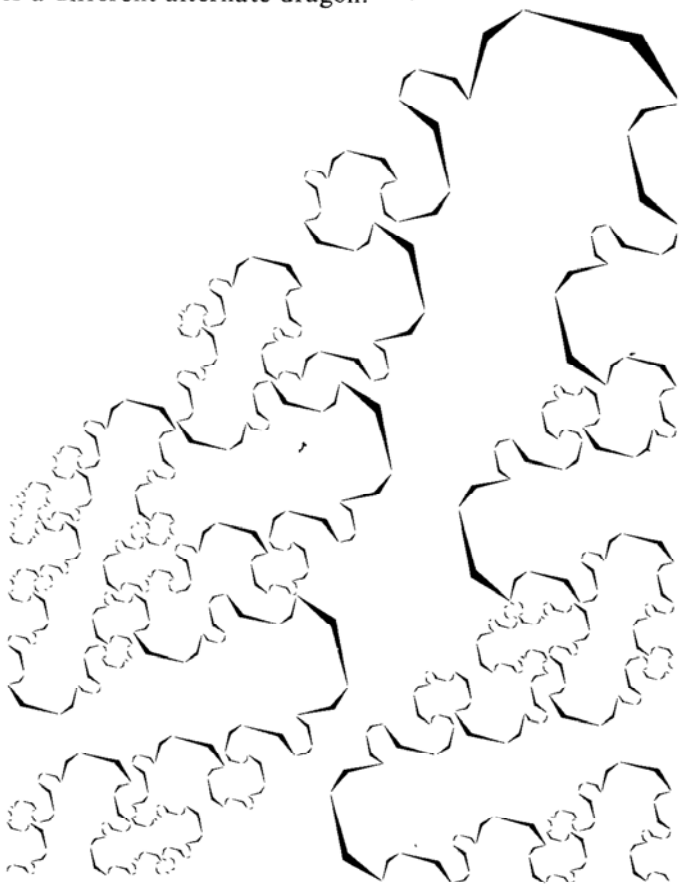


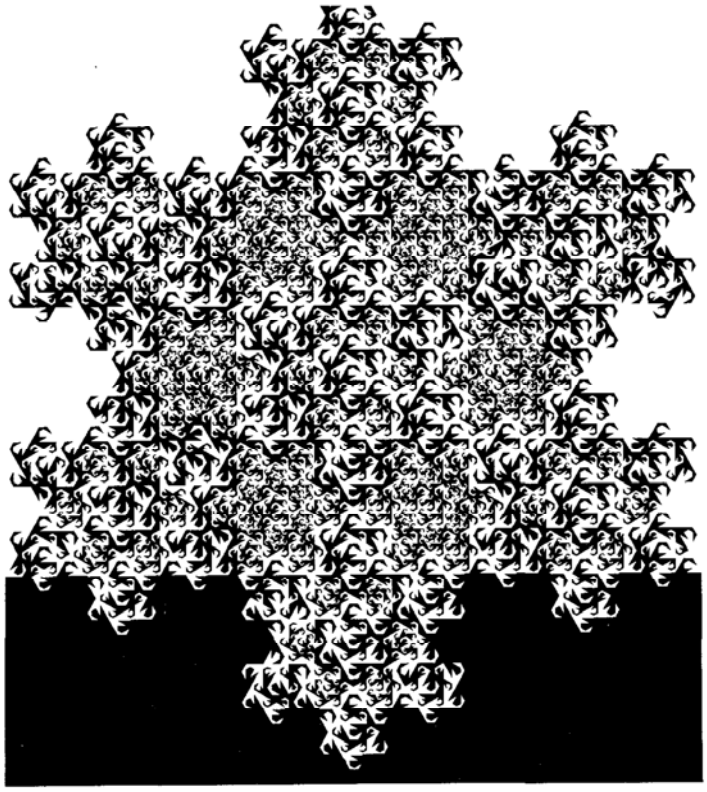
TWINDRAGON SKIN. This is a Koch curve with the following generator



The short and long intervals here are of lengths $r_1=1/\sqrt{2}$ and $r_2 = (\frac{1}{2})(\sqrt{2}) = r_1^3$, respectively. Hence, the dimension generating function is $(1/\sqrt{2})^D+2(2\sqrt{2})^D = 1$, showing that the quantity $2^{D/2}$ satisfies $x^3-x^2-2=0$.

ALTERNATE DRAGONS. (Davis & Knuth 1970.) Pick any infinite sequence x_1, x_2..., where each x_k can be either 0 or 1, and use the value of x_k to determine the first position of the generator during the k-th stage off construction: when $x_k=1$, a generator is first positioned to the right, but when $x_k=0$ it is first positioned to the left. Each sequence generates a different alternate dragon. ■





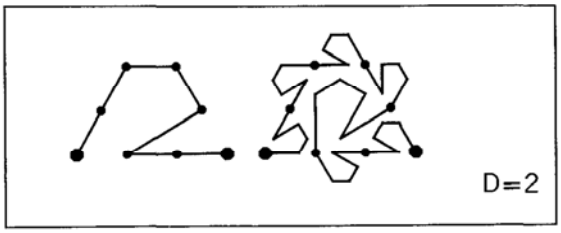
Plates 68 and 69 ¤ THE SNOWFLAKE SWEEPS: NEW PEANO CURVES AND TREES (WATERSHED AND RIVER DIMENSIONS D~1.2618)

These plates illustrate a family of Peano curves I designed. They fill the original Koch snowflake (Plate 45), hence two basic monsters of circa 1900 are brought together.

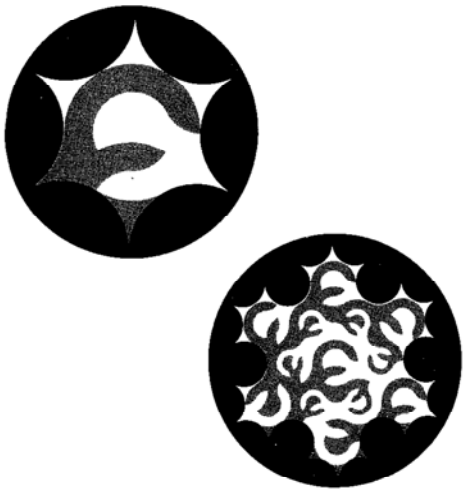
A more important virtue is that a glance suffices here to document a major theme of the present Essay: Peano curves are far from being mathematical monsters with no concrete interpretation. If they fail to self-contact, they involve readily visible and interpretable conjugate trees. These trees are good first-order models of rivers, watersheds, botanical trees, and human vascular systems.

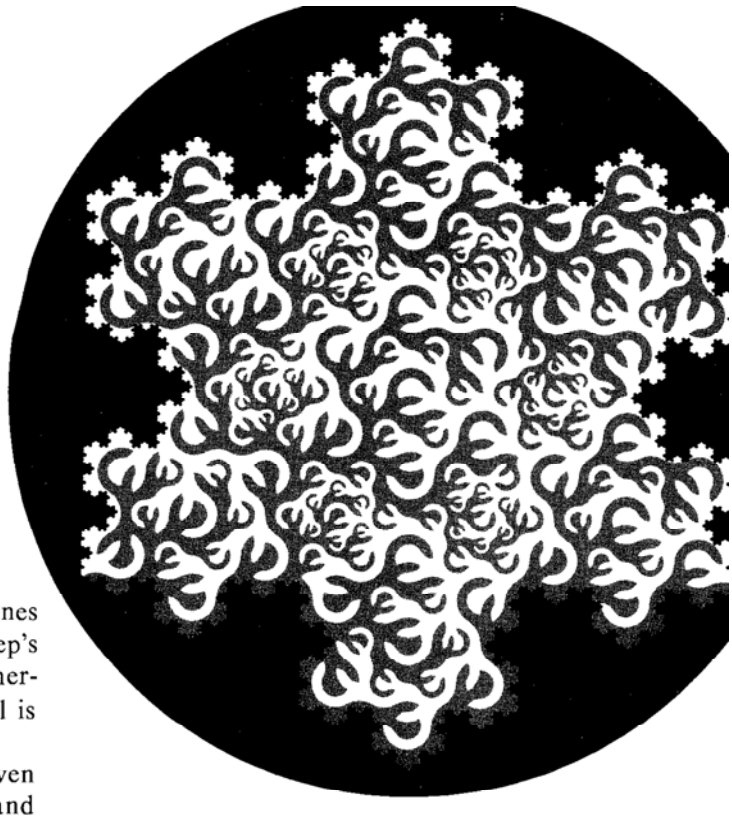
As a by-product, we obtain here a method for tiling the snowflake with unequal snowflakes.

SEVEN INTERVAL GENERATOR. Let the initiator be [0,1], and the generator and the second construction stage be



More precisely, let the above generator be denoted by S and called straight, and define the flipped generator F as the mirror image of S in the line $x=\frac{1}{2}$. At any stage of the construction of the snowflake sweep, one can use either the F, or the S generator, at will. Hence, each infinite sequence of F and S yields a different snowflake sweep.





ROUNDED OFF TERAGONS. Broken lines tend to look raw, and the snowflake sweep's teragons are made to look isotropic and otherwise much more "natural", if each interval is rounded off into one sixth of the circle.

PLATE 45. An advanced teragon of a seven interval snowflake sweep, rounded off, and later filled in, was used long ago in Plate 45 to provide a wavy background shading. Looking at it again, we are reminded of a liquid's flow past a fractal boundary, and of the shear lines between two roughly parallel flows of different velocities.

THIRTEEN INTERVAL GENERATOR. Now change the above 7 interval generator by replacing 5-th leg by a reduced version of the whole. This version can be positioned either in the S or the F position. The latter yields the following generator and second construction stage

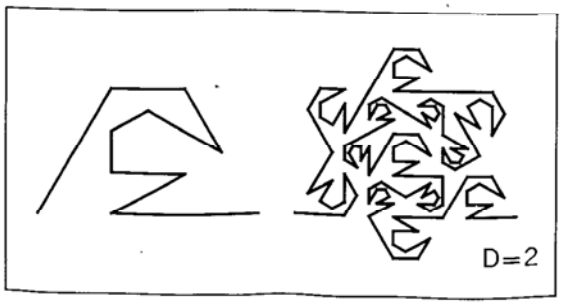
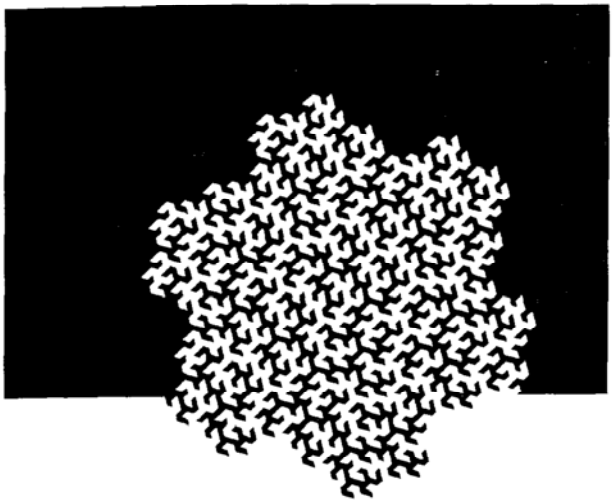
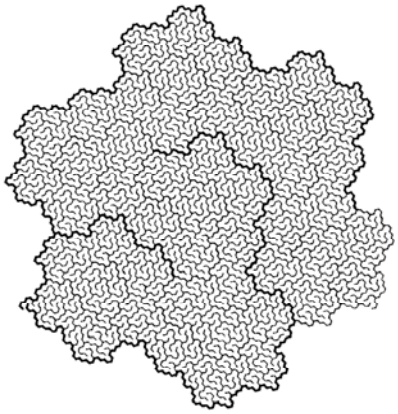


PLATE 68. This advanced teragon, shown as boundary between two fantastically intertwined domains serves better than any number of words to explain what plane-filling means.

PLATE 69. Let the above 13-interval generator be rounded off, and do the same in parallel to the snowflake curve. The resulting first few stages are shown in Plate 69.

RIVER DIMENSIONS. In Peano's original curve, each individual river is of finite length, hence of dimension 1. Here individual rivers are of dimension $\log 4/\log 3$. To achieve the dimension $D=2$, all rivers have to be taken together. ■

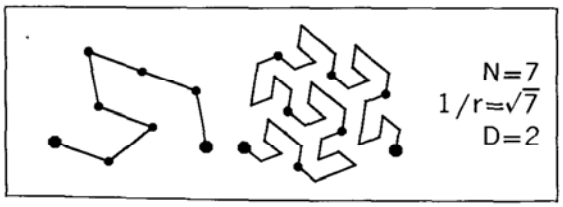




Plates 70 and 71 ¤ THE PEANO-GOSPER CURVE. ITS TREES, AND RELATED KOCH TREES (WATERSHED AND RIVER DIMENSIONS D~1.1291)

BACK TO PLATE 46. The thin broken lines on this plate, unexplained until now, represent the early construction stages 1 to 4 of a curve due to Gosper (Gardner 1976). This was the first self-avoiding Peano curve to be obtained by the Koch method without further processing.

The initiator is [0,1], and the generator is



By turning the generator counterclockwise until its first link becomes horizontal, one sees that it is drawn as a triangular lattice, on which it occupies 7 out of 3×7 links. This feature extends to triangular lattices a property which page 66 discusses for square lattices.

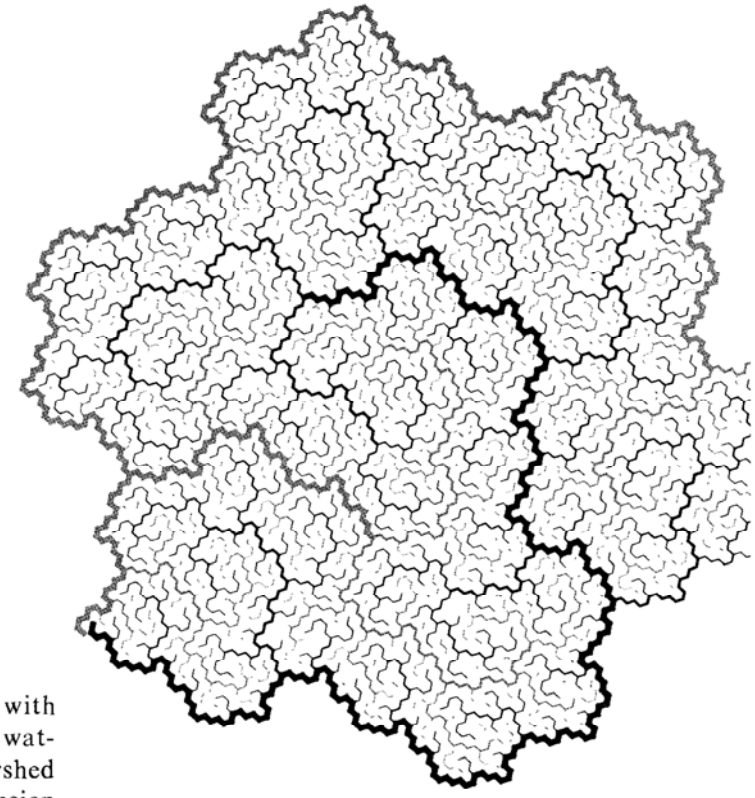
Now we see that the present Peano curve fills the Koch curve of Plate 46. The variable width hatching in Plate 46 can be explained now: it represents the fifth stage of the present construction.

LEFT OF PLATE 70. The fourth teragon of the Gosper curve is redrawn as the boundary between a black and a white region.

RIGHT OF PLATE 70. RIVER AND WATERSHED TREES. Rivers and watersheds are drawn along the midlines of the white and black "fingers" of the figure to the left of Plate 70.

TOP OF PLATE 71. Starting with the river and watershed trees to the right of Plate 70, the widths of the links are redrawn according to their relative importance in the Horton-Strahler scheme (Leopold 1962). In this instance, the river or watershed links are given widths proportional to their lengths as the crow flies. The rivers are in black, and the watersheds in gray.

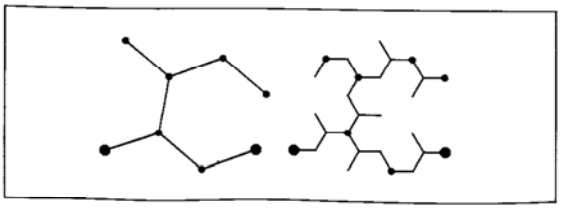
DIMENSIONS. Each Peano curve determines the D of its own boundary. In Plates 63 and 64, said boundary is merely a square. In later plates, it was a dragon's skin, then a snow-



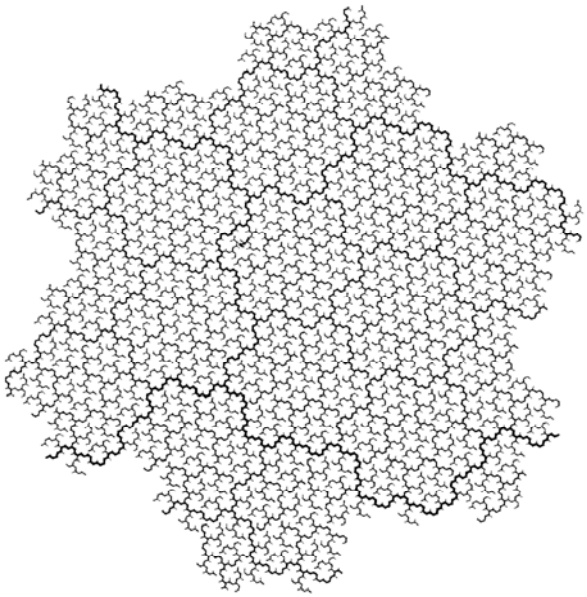
flake curve. Here it is a fractal curve with D~1.1291, which is part river and part watershed. And every other river and watershed also converges to a curve of fractal dimension D~1.1291.

FRANCE. One who as a schoolboy often gazed on a map showing the rivers Loire and Garonne does not feel far from home.

BOTTOM OF PLATE 71. A RIVER TREE CONSTRUCTED DIRECTLY BY A KOCH CASCADE. When the generator is itself tree-shaped, it generates a tree. For example, let the generator be



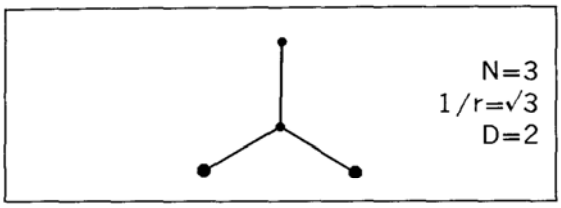
Here we have an alternative method of draining the Koch curve of Plate 46. (The last branches near the "sources" have been clipped off.) ■



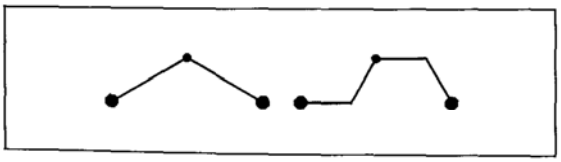
Plates 72 and 73 ¤ PLANE-FILLING FRACTAL TREES, FUDGEFLAKE, AND QUARTET

The plane-filling "river" trees deduced from some Peano curves can also be obtained by a direct recursive construction. The key is a generator that is itself tree shaped. A dull example is obtained if the tree generator is made up of 4 legs forming a + sign. One obtains the river tree of the Peano Cesàro curve (Plate 65).

FUDGEFLAKE. A better example results from taking [0,1] as initiator, and using the following generator



We begin by observing that individual rivers are generated by a midpoint displacement shape like on Plate 43. Hence, every asymptotic river has the dimension $D = \log 2/\log\sqrt{3} = \log 4/\log 3$. This value is very familiar from the snowflake curve, but the curve with which we deal here is not a snowflake, because the positioning of the generator follows a different rule.



In order to leave room for the rivers, the generator must be made to alternate between the right and the left. Therefore, the snowflake's symmetry is fudged, and the domain these rivers drain is to be called *fudgeflake*.

Now, we turn to the river tree. Its teragons do not self-overlap, but they self-contact badly. This feature's asymptotic variant is unavoidable, and it is also unobjectionable, since it expresses quite properly the fact that several rivers can originate at the same point. But we shall see later in this caption that river teragons *may* avoid self-contact. Due to self-contacts, the present river teragon is an illegible chunk of hexagonal graph paper, bounded by an approximate fractal.

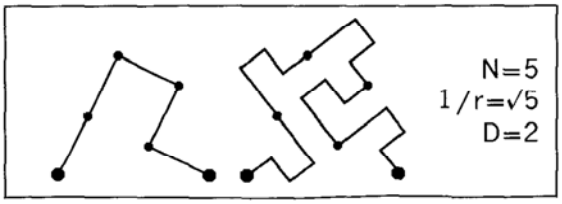
TOP OF PLATE 73. The river tree is made more transparent by erasing all river intervals that touch a source, and using a bolder pen to draw the principal river. The area drained by this tree is $\sqrt{3}/2 \sim .8660$.

FUDGEFLAKE SWEEP. Now draw a Peano curve with a Δ shaped initiator, and a generator in the shape of a Z whose legs are equal and make angles of 60°. This is the extreme case for M=3 of the family of generators used in Plates 46 and 47, but it differs profoundly from all the other cases. It is investigated in Davis & Knuth 1970.

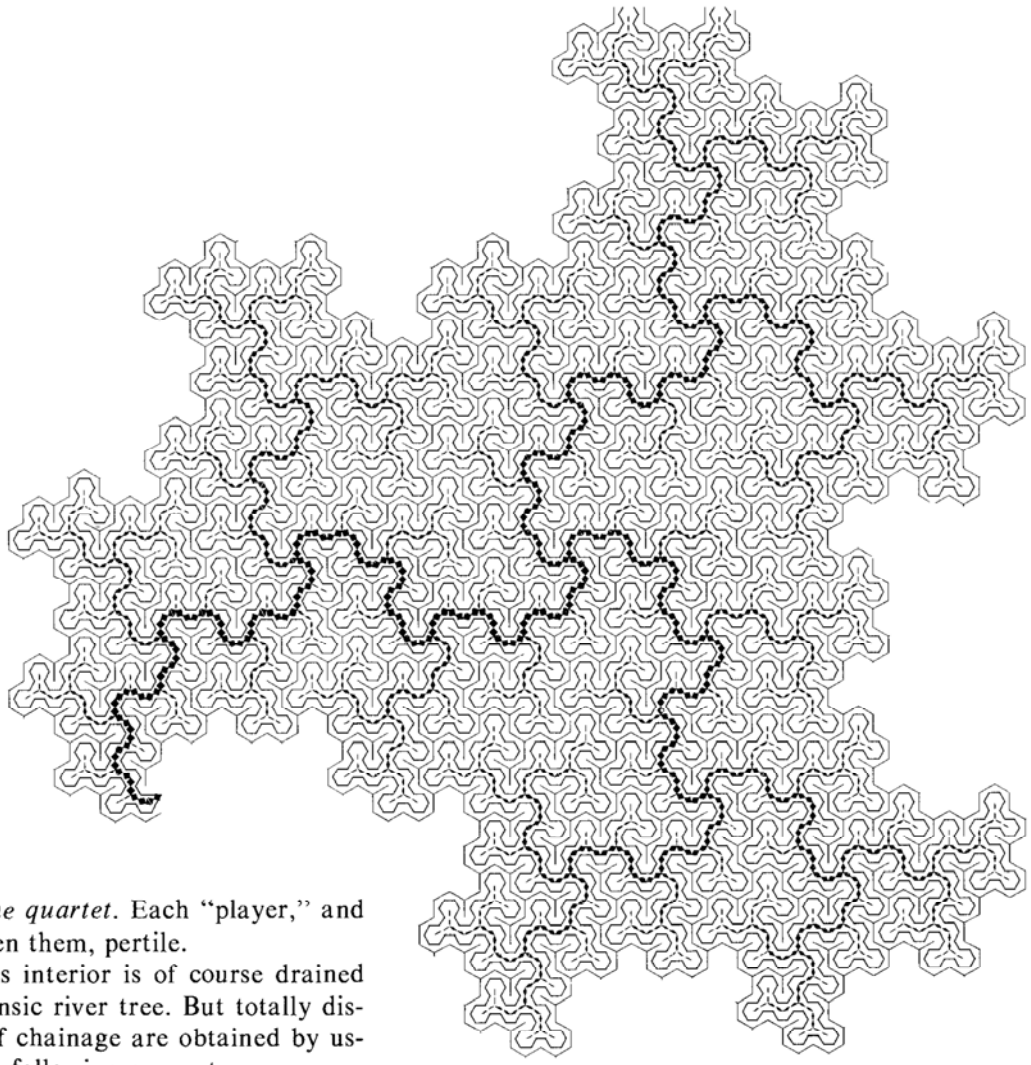
One can verify that this Peano curve's river tree is none else than the tree we just drew directly. The initiator's sides are of length 1, and the corresponding Peano curve sweeps an area equal to $\sqrt{3}/6 \sim .2886$ (how inefficient!).

QUARTET. Next, we consider a different Koch curve, together with three curves that fill it: one Peano curve and two trees. These shapes, which I designed, illustrate a further theme of interest.

Take [0,1] as initiator, and take the following generator

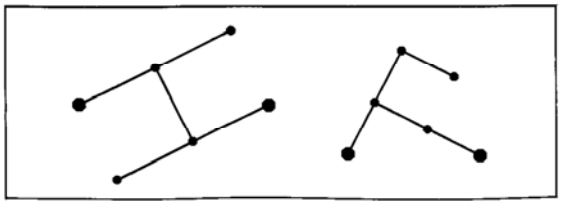


This curves' boundary converges to a Koch curve of dimension $D=\log 3/\log\sqrt{5}=1.3652$. Advanced teragons of the boundary and of the Peano curve are seen in the center of Plate 49,

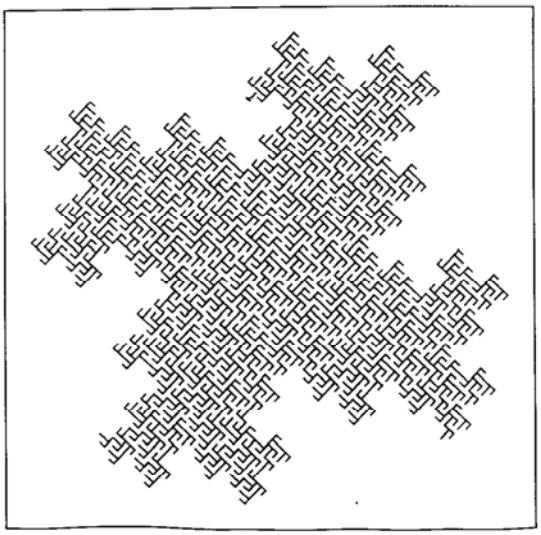


which I term *the quartet*. Each "player," and the table between them, pertile.

The quartet's interior is of course drained by its own intrinsic river tree. But totally distinct patterns of chainage are obtained by using either of the following generators



With the generator to the left, the teragons self-contact, as with the first example in this caption. And the drainage area turns out to be ½. With the generator to the right, the teragons *avoid* self-contact. And the drainage area is 1. An advanced teragon is shown in the bottom figure of Plate 73. ■



8 ¤ Fractal Events and Cantor Dusts

This chapter's principal goal is to acquaint the reader concretely and painlessly with yet another mathematical object ordinarily viewed as pathological, the Cantor dust, $\mathcal{C}$. This and related dusts we shall describe have fractal dimensions between 0 and 1.

Being formed by points on a straight line, they are easy to study. In addition, they help introduce in simplest form several concepts that are central to fractals but that have been so underutilized in the past that no specific terms were required to denote them. First, the term *dust* is given a technical meaning, as an informal equivalent to a *set of topological dimension* $D_T=0$, just as "curve" and "surface" denote sets of topological dimensions $D_T=1$ and $D_T=2$. Other new terms are *curd, gap,* and *trema,* to be explained.

NOISE

For the layman, a noise is a sound that is too strong, has no pleasing rhythm or purpose, or interferes with more desirable sounds. Partridge 1958 proclaims that the term "derives from the Latin *nausea* (related to *nautes* = sailor), the semantic link being afforded by the noise made by an ancient shipful of passengers groaning and vomiting in bad weather." (The *Oxford English Dictionary* is not so sure.) As to contemporary physics, it is less colorful, and not nearly so precise: it uses *noise* as a synonym of chance fluctuation or error, irrespective of origin and manifestation. This chapter introduces $\mathcal{C}$ through the case study of an esoteric but simple noise.

ERRORS IN DATA TRANSMISSION LINES

A transmission channel is a physical system capable of transmitting electricity. However, electric current is subject to spontaneous noise. The quality of transmission depends on the likelihood of error due to noise distortion, which depends, in turn, on the ratio between the intensities of signal and noise.

This chapter is concerned with channels that transmit computer data and involve very strong signals. An interesting fact is that the signal is discrete, hence the distribution of

errors simplifies the distribution of noise to the bone, so to speak. Noise involves a function having several possible values, while errors involve a function that has only two possible values. For example, it may be the indicator function, which is 0 when there is no error at time t, and 1 if there is an error.

Physicists have mastered the structure of the noises that predominate in the case of weak signals, e.g., thermal noise. In the problem just described, however, the signal is so strong that the classical noises are negligible.

The nonnegligible *excess noises* are difficult and fascinating because little is known about them. This chapter examines an excess noise that was, around 1962, of practical importance to electrical engineers, so that diverse talents were called upon to investigate it. My contribution to this effort was the first concrete problem in which I experienced the need to use fractals. No one remotely imagined at that time that a careful study of this apparently modest engineering difficulty would get us so far.

BURSTS AND GAPS

Let us subject the errors to increasingly refined analysis. A rough analysis reveals the presence of periods during which no error is encountered. Let these remission periods be called "gaps of rank 0" if their duration exceeds one hour. By contrast, any time interval flanked by gaps of rank 0 is singled out as being a "burst of errors of rank 0." As the analysis is made three times more accurate, it reveals that the original burst is itself "intermittent." That is, shorter gaps "of rank 1," lasting 20 minutes or more, separate correspondingly shorter bursts "of rank 1." Likewise, each of the latter contains several gaps "of rank 2," lasting 400 seconds, separating bursts "of rank 2," and so on, each stage being based on gaps and bursts that are three times shorter than the previous ones. The process is illustrated very roughly by Plate 80. (Do not pay attention to the caption yet.)

The preceding description suggests something about the relative positions of the bursts of rank k within a burst of rank k–1. The probability distribution of these relative positions seems independent of k. This invariance is obviously an example of self-similarity, and fractal dimension cannot be far behind, but let us not rush. This Essay's diverse case studies are meant, among others, to elicit new themes or refine old ones. With this in mind, it seems best to reverse the historical order, and introduce a new theme through a rough nonrandom variant of the Berger & Mandelbrot stochastic model of errors, Chapter 31.

A ROUGH MODEL OF ERROR BURSTS: THE CANTOR FRACTAL DUST $\mathcal{C}$

The preceding section constructs the set of errors by starting with a straight line, namely the time axis, then cutting out shorter and shorter error-free gaps. This procedure may be unfamiliar in natural science, but pure

mathematics has used it at least since Georg Cantor (Hawkins 1970, especially p. 58).

In Cantor 1883, the initiator is the closed interval [0,1]. The term "closed" and the use of brackets indicate that the extreme points are included; this notation was used in Chapter 6, but there was no need until now to make it explicit. The first construction stage consists in dividing [0,1] into 3 pieces, then removing the middle open third, designated]⅓, ⅔[. The term "open" and the use of reversed brackets indicate that the extreme points are excluded. Next, one removes the open middle of each of N=2 remaining thirds. And so on to infinity.

The remainder set $\mathcal{C}$ is called either *dyadic*, due to the fact that N=2, or *triadic* or *ternary*, due to the fact that [0,1] is subdivided into 3 pieces.

More generally, the number of pieces, called *base*, is denoted by b, the ratio between each N-th of the set and the whole being r=1/b. $\mathcal{C}$ is also called *Cantor discontinuum*, and I shall momentarily suggest the term, *Cantor fractal dust*. Since a point on the time axis marks an "event," $\mathcal{C}$ is a fractal sequence of events.

CURDLING, TREMAS, AND WHEY

Cantor's procedure is a *cascade*, to use a term Lewis Richardson had applied to turbulence, and we first borrowed in Chapter 6 to describe coastlines and the Koch curve. "Stuff" that was uniformly distributed over an initiator [0,1] is subjected to a centrifugal eddy which sweeps it into the extreme thirds.

The middle third portion cut out of [0,1] to form a gap is henceforth denoted as *trema generator*. This neologism is being coined in this section from τρημα meaning hole, whose distant relative is the Latin *termes* = termite. It may be the shortest Greek word that has not yet been put to work with a significant scientific meaning.

In this context, tremas coincide with gaps, but in different instances to be encountered later they do not, which is why two different terms are required.

While a "first-order trema" is emptied, the total stuff is conserved and redistributed with uniform density over the outer thirds, to be called *precurds*. Then two centrifugal eddies come in and repeat the same operation, starting with the two intervals [0,⅓] and [⅔,1]. The process continues as a Richardsonian cascade converging at the limit to a set to be called *curd*. If a stage's duration is proportional to the eddy size, the total process is of finite duration.

In parallel, I propose *whey* (a term Miss Muffet should not mind) to denote the space outside the curd.

It is suggested that the above terms be used not only in a mathematical but also in a physical meaning: *curdling* to denote any cascade of instabilities resulting in contraction, and *curd* to denote a volume within which a physical characteristic becomes increasingly concentrated as a result of curdling.

ETYMOLOGY. *Curd* derives from the old

English *crudan,* 'to press, to push hard.' This erudition from Partridge 1958 is not necessarily irrelevant, since the etymological kin of curd doubtless include fractal kin of interest; see Chapter 23.

Note the following free associations: curds → cheese → milk → Milky Way → Galaxy ($\gamma\alpha\lambda\alpha$ = milk) → galaxies. I coined *curdling* while working on galaxies, and the etymological undertones of "galactic curdling" did not escape my notice.

OUTER CUTOFF AND EXTRAPOLATED CANTOR DUSTS

As a prelude to the extrapolation of $\mathcal{C}$, let us recall a point of history. When Cantor introduced $\mathcal{C}$, he had barely left his original field, the study of trigonometric series. Since such series are concerned with periodic functions, the only extrapolation they involve is endless repetition. Now recall the self-explanatory terms of *inner* and *outer cutoff,* which Chapter 6 borrows from the study of turbulence. These are, respectively, the sizes ϵ and Ω of the smallest and the largest feature present in a set, and one may say that Cantor restricted himself to $\Omega=1$. The k-th construction stage yields $\epsilon=3^{-k}$, but $\epsilon=0$ for $\mathcal{C}$ itself. To achieve any other $\Omega<\infty$, for example the value of 2π appropriate in a Fourier series, one enlarges the periodic Cantor dust in the ratio Ω.

However, self-similarity, which this Essay views as valuable, is destroyed by repetition. But it is readily saved, if the initiator is used only for extrapolation and if extrapolation follows an *inverse* or *upward* cascade. The first stage enlarges $\mathcal{C}$ in the ratio $1/r=3$ and positions it on [0,3]. The result is $\mathcal{C}$ plus a replica translated to the right and separated from $\mathcal{C}$ by a new trema of length 1. The second stage enlarges the outcome of the first stage in the same ratio 3 and positions it on [0,9]. The result is $\mathcal{C}$ plus 3 replicas translated to the right and separated by two new tremas of length 1, and one new trema of length 3. The upward cascade continues to enlarge $\mathcal{C}$ in the successive ratios of the form 3^k.

If one prefers, one may alternate two stages of interpolation, then a stage of extrapolation, etc. In this fashion, each series of three stages multiplies the outer cutoff Ω by 3 and divides the inner cutoff ϵ by 3.

◅ In this extrapolated dust, the negative axis is empty: an infinite trema. The underlying notion is discussed further in Chapter 13, where we tackle the (infinite) continent and the infinite cluster. ►

DIMENSIONS D BETWEEN 0 AND 1

The set yielded by infinite interpolation and extrapolation is self-similar, and

$$D=\log N/\log(1/r)=\log 2/\log 3\sim 0.6309,$$

a fraction between 0 and 1.

By following a different curdling rule, we can achieve other D's, in fact any dimension between 0 and 1. If the first stage trema is of

length $1-2r$, where $0<r<\frac{1}{2}$, the dimension is $\log 2/\log(1/r)$.

Further variety becomes possible if $N\neq 2$. For the sets with $N=3$ and $r=1/5$, we find

$$D=\log 3/\log 5\sim 0.6826.$$

For the sets with $N=2$ and $r=\frac{1}{4}$, we find

$$D=\log 2/\log 4=\frac{1}{2}.$$

For the sets with $N=3$ and $r=1/9$, we also find

$$D=\log 3/\log 9=\frac{1}{2}.$$

Although their D are equal, these last two sets "look" very different. This observation is taken up again and extended in Chapter 34, and leads to the notion of lacunarity.

Observe also that there is at least one Cantor set for every $D<1$, but it follows from $Nr\leq 1$ that $N<1/r$, hence D is never above 1.

$\mathcal{C}$ IS CALLED DUST BECAUSE $D_T=0$

While a Cantor set's D can vary between 0 and 1, from the topological viewpoint all Cantor sets are of dimension $D_T=0$, because any point is by definition cut from the other points, without anything having to be removed to cut it. From this viewpoint, there is no difference between $\mathcal{C}$ and finite sets of points! The fact that $D_T=0$ in this last case is familiar in standard geometry, and Chapter 6 uses it in arguing that Koch's $\mathcal{K}$ is of topological dimension 1. But $D_T=0$ for all totally disconnected sets.

In the absence of accepted colloquial counterparts to "curve" and "surface" (which are connected sets with $D_T=1$ and $D_T=2$), I propose that sets with $D_T=0$ be called *dusts*.

GAPS' LENGTH DISTRIBUTION

In a Cantor dust, let u be a possible value of a gap's length, and denote by U the length when it is unknown, and by $Nr(U>u)$ the number of gaps or tremas of length U greater than u. ◅ This notation is patterned after the notation $Pr(U>u)$ of probability theory. ► One finds there is a constant prefactor F, such that the graph of the function $Nr(U>u)$ constantly crosses the graph of Fu^{-D}. Here comes dimension again. With $\log u$ and $\log Nr$ as coordinates, the steps are uniform.

AVERAGE NUMBERS OF ERRORS

As in the case of a coastline, a rough idea of the sequence of errors is obtained if Cantor curdling stops with intervals equal to $\epsilon=3^{-k}$. The ϵ may be the length of time required to transmit a single symbol. One must also use Cantor's periodic extrapolation with a large but finite Ω.

The number of errors between times 0 and R, denoted by M(R), keeps time by counting only those instants that witness something

noteworthy. It is an example of *fractal time.*

When the sample begins at t=0 (which is the only case to be considered here), the derivation of M(R) proceeds as in the case of the Koch curve. As long as R is smaller than Ω, the number of errors doubles each time R is multiplied by 3. As a result, $M(R) \propto R^D$.

This expression is like the standard expression for the mass of a disc or ball of radius R in D-dimensional Euclidean space. It is also identical to the expression obtained in Chapter 6 for the Koch curve.

As a corollary, the average number of errors per unit length varies roughly like R^{D-1} as long as R lies between the inner and the outer cutoffs. When Ω is finite, the decrease in the average number of errors continues to the final value of Ω^{D-1}, which is reached with $R=\Omega$. Thereafter, the density remains more or less constant. When Ω is infinite, the average number of errors decreases to zero. Finally, the empirical data often suggest that Ω is finite and very large, but fail to determine its value with any accuracy. If this is the case, the average number of errors has a lower limit that does not vanish but that is so ill-determined as to be of no practical use.

TREMA ENDPOINTS AND THEIR LIMITS

◅ The most conspicuous members of $\mathcal{C}$, the trema endpoints, do *not* exhaust $\mathcal{C}$; in fact they constitute but a tiny portion of it. The other points' physical importance is discussed in Chapter 19. ►

THE CANTOR DUSTS' TRUE NATURE

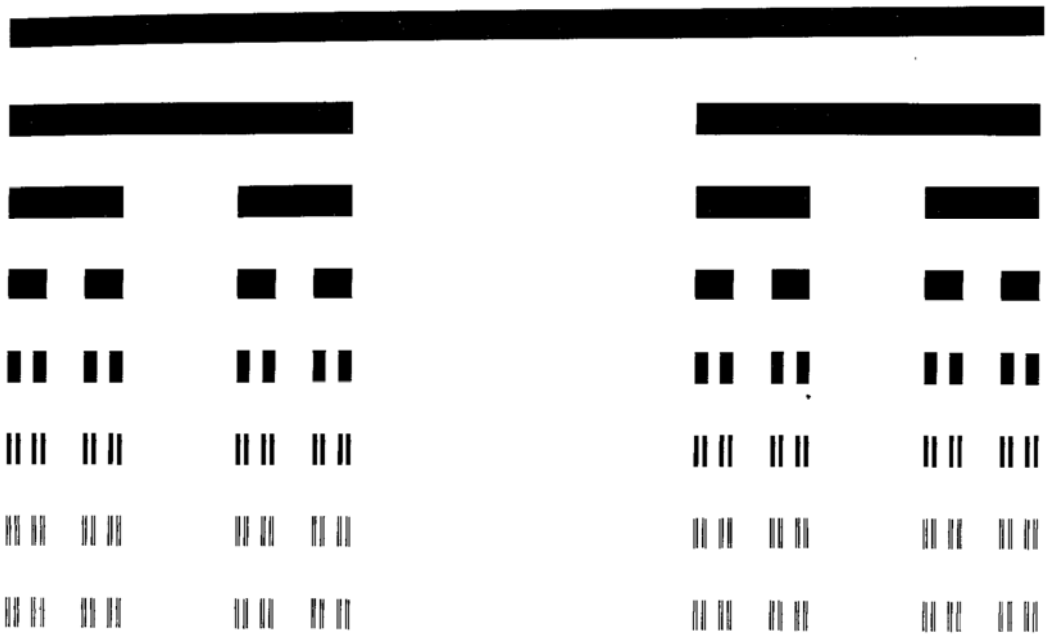
The reader who has followed thus far and/or has heard the echo of the rapidly growing literature on Devil's Staircases (caption of Plate 83) must find it hard to believe that, when I started on this topic in 1962, everyone was agreeing that Cantor dusts are *at least* as monstrous as the Koch and Peano curves.

Every self-respecting physicist was automatically "turned off" by a mention of Cantor, ready to run a mile from anyone claiming $\mathcal{C}$ to be interesting in science, and eager to assert that such claims had been advanced, tested, and found wanting. My sole encouragement came from S. Ulam's suggestions, tantalizing despite their failure to be either developed or accepted, concerning the possible role for Cantor sets in the gravitational equilibrium of star aggregates; see Ulam 1974.

To publish on Cantor dusts, I had to erase every mention of Cantor!

But here we were led to $\mathcal{C}$ by Nature's own peculiarities. And Chapter 19 describes a second, very different, physical role for $\mathcal{C}$. All this must mean that the true nature of the Cantor dust is very different.

It is undeniable that in most cases $\mathcal{C}$ itself a very rough model, requiring many improvements. I contend, however, that *the very same properties that cause Cantor discontinua to be viewed as pathological are indispensable in a model of intermittency*, and must be preserved in more realistic substitutes for $\mathcal{C}$. ■



Plates 80 and 81 ¤ CANTORIAN TRIADIC BAR AND CAKE (HORIZONTAL SECTION DIMENSION D=log 2/log 3=0.6309). SATURN'S RINGS. CANTOR CURTAINS.

The Cantor dust uses [0,1] as initiator, and its generator is

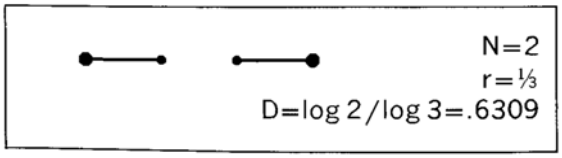


PLATE 80. The Cantor dust is extraordinarily difficult to illustrate, because it is thin and spare to the point of being invisible. To help intuition by giving an idea of its form, thicken it into what may be called a Cantor bar. ◅ In technical terms, this is the Cartesian product of a Cantor dust of length 1, by an interval of length 0.03. ►

CURDLING. The construction of the Cantor bar results from the process I call *curdling*. It begins with a round bar (seen in projection as a rectangle in which width/length=0.03). It is best to think of it as having a very low density. Then matter "curdles" out of this bar's middle third into the end thirds, so that the positions of the latter remain unchanged. Next matter curdles out of the middle third of each end third into its end thirds, and so on ad infinitum until one is left with an infinitely large number of infinitely thin slugs of infinitely high density. These slugs are spaced along the line in the very specific fashion induced by the generating process. In this illustration, curdling (which eventually requires hammering!) stops when both the printer's press and our eye cease to follow; the last line is indistinguishable from the last but one: each of its ultimate parts is seen as a gray slug rather than two parallel black slugs.

CANTOR CAKE. When curdling starts with a pancake, much less thick than it is wide, and dough curdles into thinner pancakes (while exuding an appropriate filling), one ends up with an infinitely extrapolated Napoleon, which one might call *Cantor cake*.

SATURN'S RINGS. Saturn was originally believed to have a single ring around it. But eventually a break was discovered, then two, and now Voyager I has identified a very large number of breaks, mostly very thin ones. Voyager also established that the rings are diaphanous: they let sunlight through...as befits a set we called "thin and spare."

Thus, the rings' structure (see Stone & Minen 1981, especially the cover illustration) is suggestive of a collection of near circles, each with a radius corresponding to the distance from some origin to a point in Cantor dust. ◅ The technical term is Cartesian product of a Cantor dust by a circle. Actually, it may be that a closer picture is given by a circle's product with a dust with positive measure, like those examined in Chapter 15. ► Last minute insert: The same idea is stated independently to Avron & Simon 1981, which relates it to Hill's equation; their Note 6 includes many other relevant references.

SPECTRA. Harter 1979–1981 describes some spectra of organic molecules whose resemblance to a Cantor dust is stunning.

PLATE 81. Here, the Cantor dust's shape is clarified by being placed among generalized dusts with N=2 and variable r. The vertical coordinate is either r itself, ranging from 0 to ½ (bottom figure), or D ranging from 0 to 1 (top figure). Both theater curtains are topped by the full interval [0,1]. Every horizontal cut of either figure is some Cantor dust, with the arrows pointing out r=⅓ and D=0.6309.

A FAMOUS GREEK PARADOX. Greek philosophers believed that, in order to be indefinitely subdivisible, a body had to be continuous. They had not heard of Cantor dusts. ■

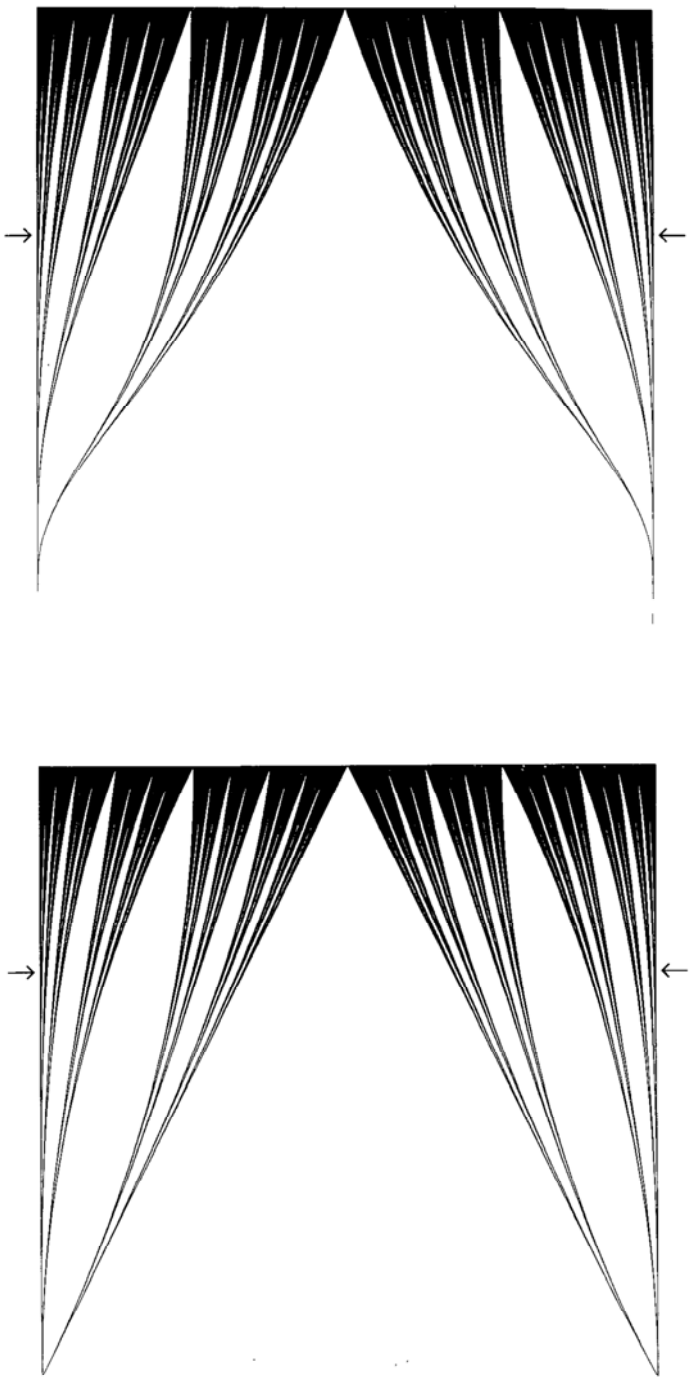


Plate 83 ¤ CANTOR FUNCTION, OR DEVIL'S STAIRCASE (DIMENSION D=1 THE RISERS' ABSCISSAS ARE OF DIMENSION D~0.6309). CANTOR MOTION

The Cantor function describes the distribution of mass along the Cantor bar of Plate 80. Many writers refer to its graph as the *Devil's Staircase*, because it is odd indeed. Set both the bar's length and mass as equal to 1, and for every value of the abscissa R define M(R) as the mass contained between 0 and R. Since there is no mass in the gaps, M(R) remains constant along intervals that add up to the whole length of the bar. However, since hammering does not affect the total mass in the bar, M(R) must manage to increase *somewhere* from the point of coordinates (0,0) to the point of coordinates (1,1). It increases over infinitely many, infinitely small, highly clustered jumps corresponding to the slugs. Hille & Tamarkin 1929 describes this function's odd properties in detail.

REGULARIZING MAPPINGS. The Devil's staircase accomplishes the feat on mapping the drastic nonuniformity of the Cantor bar into something uniform and homogeneous. Starting with two different intervals of the same length on the vertical scale, the inverse function of the Cantor staircase yields two collections of slugs that contain the same mass—even though they usually look very different from each other.

Since science thrives on uniformity, it often happens that such regularizing transformations make fractal irregularity accessible to analysis.

FRACTAL HOMOGENEITY. It is convenient to describe the distribution of mass in the Cantor bar as *fractally homogeneous*.

CANTOR MOTION. As in the case of the Koch curve reinterpreted as Koch motion, or of the Peano motion, it is useful to reinterpret the ordinate M(R) as a time. If so, the inverse function R(M) gives the position of a *Cantor motion* at time t. This motion is most discontinuous. Chapters 31 and 32 describe a randomized linear and spatial generalizations.

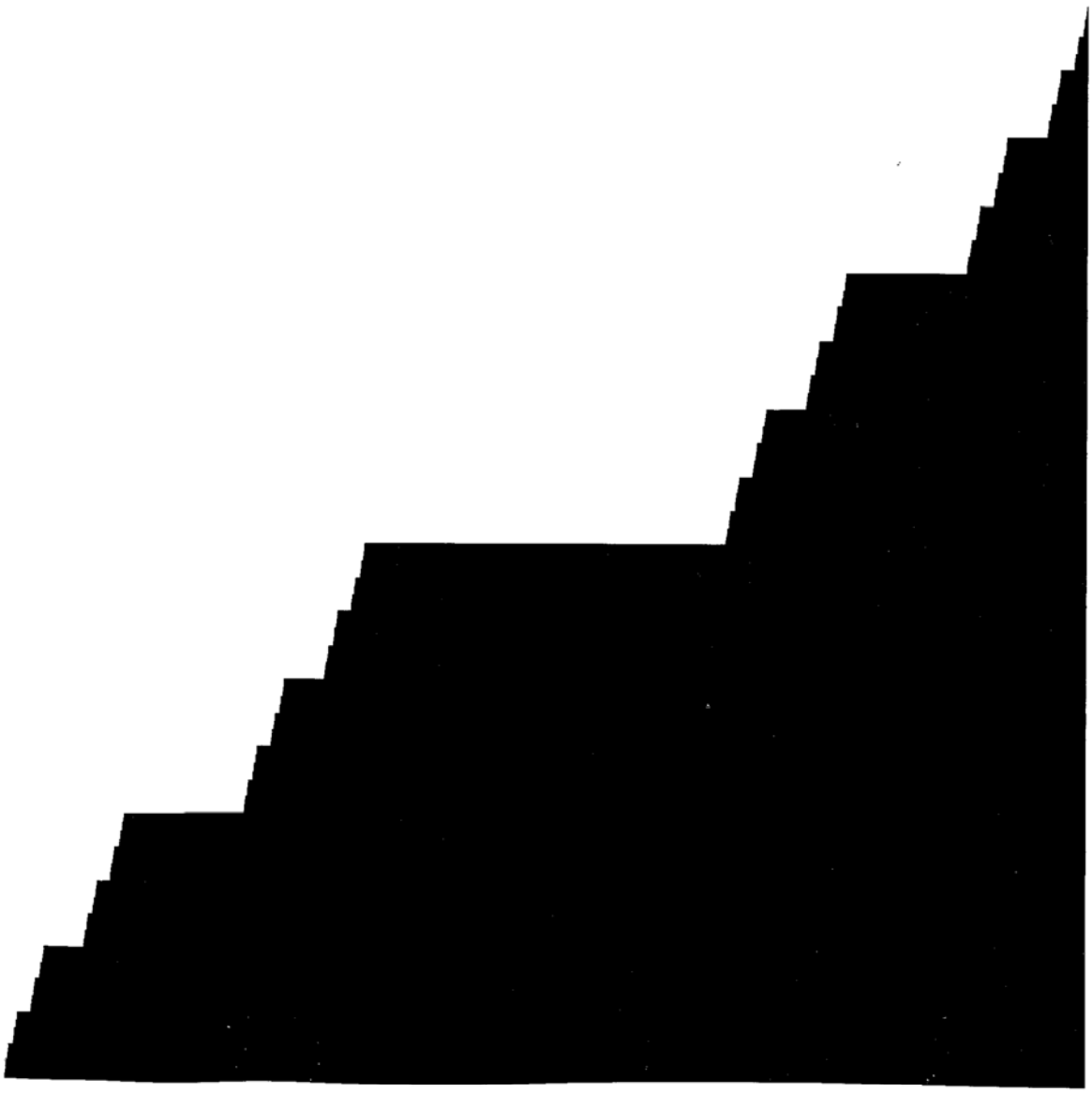
FRACTAL DIMENSION. The sums of the widths and of the heights of the steps both equal 1, and one finds in addition that this curve has a well-defined length equal to 2. A curve of finite length is called rectifiable and is of dimension D=1. This example demonstrates that the dimension D=1 is compatible with the presence of many irregularities, as long as they remain sufficiently scattered.

◅ One would love to call the present curve a fractal, but to achieve this goal we would have to define *fractals* less stringently, on the basis of notions other than D alone. ►

SINGULAR FUNCTIONS. The Cantor staircase is a nondecreasing and nonconstant function that is singular, in the sense that it is continuous but *non*differentiable. Its derivative vanishes almost everywhere, and its continuous variation manages to occur over a set whose length—i.e., linear measure—vanishes.

Any nondecreasing function can be written as the sum of a singular function, of a function made of discrete jumps, and of a differentiable function. The last two components are classical in mathematics and of wide use in physics. On the other hand, the singular component is widely regarded in physics as pathological and totally devoid of uses. A principal theme of this Essay is that this last opinion is totally devoid of merit.

DEVIL'S STAIRCASES IN STATISTICAL PHYSICS. The publication of this plate in my 1977 Essay brought the Devil's staircase to the physicists' attention, and stimulated an extensive literature. Diagrams analogous to the "curtains" of Plate 81, or the Fatou curtain of Plate 185, are encountered with growing frequency. See Aubry 1981. Important earlier work (Azbel 1964, Hofstadter 1976), which used to be isolated, merges with this new development. ■



9 ¤ Fractal View of Galaxy Clusters

In Chapters 6 and 7, the Koch and Peano fractals are introduced via geomorphology, but the most significant uses of fractals are rooted elsewhere. Inching toward the mainstream of science, this chapter and the next two tackle two issues of exceptional antiquity, importance and difficulty.

The distribution of the stars, the galaxies, the clusters of galaxies, and so on fascinates the amateur as well as the specialist, yet clustering remains peripheral to astronomy and to astrophysics as a whole. The basic reason is that no one has yet explained why the distribution of matter falls into an irregular hierarchy, at least within a certain range of scales. While there are allusions to clustering in most works on the subject, serious theoretical developments hasten to sweep it under the rug, claiming that on scales beyond some large but unspecified threshold galaxies are uniformly distributed.

Less fundamentally, the hesitation in dealing with the irregular arises from the absence of tools to describe it mathematically. Statistics is asked to decide between two assumptions, only one of which is thoroughly explored (asymptotic uniformity). Is it surprising that the results are inconclusive?

The questions, however, refuse to be set aside. In parallel with efforts to *explain,* I think it indispensable to *describe* clustering, and to mimic reality by purely geometric means. The fractal treatment of this subject, scattered over several chapters of this Essay, proposes to show by explicitly constructed models that the evidence is compatible with a degree of clustering that extends far beyond the limits suggested by existing models.

The present introductory chapter describes an influential theory of the formation of stars and galaxies, due to Hoyle, the principal descriptive model of their distribution, due to Fournier d'Albe (also known as the Charlier model), and, most important, sketches some

empirical data. It is shown that both theories and data can be interpreted in terms of a scaling fractal dust. I argue that the distribution of galaxies and of stars includes a zone of self-similarity in which the fractal dimension satisfies $0<D<3$. Theoretical reasons for expecting $D=1$ are sketched, raising the question of why the observed D is ~ 1.23.

PREVIEW. Chapter 22 uses fractal tools to improve our understanding of what the cosmological principle means, how it can and should be modified, and why the modification demands randomness. A discussion of improved model clusters is withheld until Chapters 22, 23, and 32 to 35.

IS THERE A GLOBAL DENSITY OF MATTER?

Let us begin with a close examination of the concept of global density of matter. As with the concept of the length of a coastline, things seem simple, but in fact go awry very quickly and most interestingly. To define and measure density, one starts with the mass M(R) in a sphere of radius R centered on Earth. The approximate density, defined as

$$M(R)/[(4/3)\pi R^3],$$

is evaluated. After that, the value of R is made to tend toward infinity, and the global density is defined as the limit toward which the approximate density converges.

But need the global density converge to a positive and finite limit? If so, the speed of convergence leaves a great deal to be desired. Furthermore, the estimates of the limit density had behaved very oddly in the past. As the depth of the world perceived by telescopes increased, the approximate density diminished in a surprisingly systematic manner. According to de Vaucouleurs 1970, it has remained $\propto R^{D-3}$. The observed exponent D is much smaller than 3, the best estimate, on the basis of indirect evidence, being $D=1.23$.

The thesis of de Vaucouleurs is that the behavior of the approximate density reflects reality, meaning that $M(R)\propto R^D$. This formula recalls the classical result that a ball of radius R in a Euclidean space of dimension E has a volume $\propto R^E$. In Chapter 6 we encounter the same formula for the Koch curve, with the major difference that the exponent is not the Euclidean dimension $E=2$ but a fraction-valued fractal dimension D. And Chapter 8 derives $M(R)\propto R^D$ for the Cantor dust on the time axis (for which $E=1$).

All these precedents suggest very strongly that the de Vaucouleurs exponent D is a fractal dimension.

ARE STARS IN THE SCALING RANGE?

Obviously, the scaling range in which D satisfies $0<D<3$ must end before one reaches objects with well-defined edges, such as planets. But does it, or does it not, include stars? According to data by Webbink reported in Faber & Gallagher 1980, the mass of the Milky

Way interior to radius R may very well be represented as $M(R) \propto R^D$, with the D extrapolated from galaxies. But we continue our discussion exclusively in terms of galaxies.

IS THERE AN UPPER CUTOFF TO THE SCALING RANGE?

The question of how far the range in which $0<D<3$ extends in the direction of very large scales is controversial and the subject of renewed activity. Many authors either state or imply that the scaling range admits of an outer cutoff corresponding to clusters of galaxies. Other authors disagree. De Vaucouleurs 1970 asserts that "clustering of galaxies, and presumably of all forms of matter, is the dominant characteristic of the structure of the universe on all observable scales with no indication of an approach to uniformity; the average density of matter decreases steadily as even larger volumes of space are considered, and there is no observational basis for the assumption that this trend does not continue out to much greater distances and lower densities."

The debate between these two schools of thought is interesting and important to cosmology—but not for the purposes of this Essay. Even if the range in which $0<D<3$ is cut off at both ends, its importance is sufficient in itself to warrant a careful study.

In either case, the Universe (just like the ball of thread discussed in Chapter 3) appears to involve a sequence of several different effective dimensions. Starting with scales of the order of Earth's radius, one first encounters the dimension 3 (due to solid bodies with sharp edges). Then the dimension jumps to 0 (matter being viewed as a collection of isolated points). Next is the range of interest, ruled by some nontrivial dimension satisfying $0<D<3$. If scaling clustering continues ad infinitum, so does the applicability of this last value of D. If, on the contrary, there is a finite outer cutoff, a fourth range is added on top, in which points lose their identity and one has a uniform fluid, meaning that the dimension again equals 3.

On the other hand, the most naive idea is to view the galaxies as distributed near uniformly throughout the Universe. Under this untenable assumption, one has the sequence $D=3$, then $D=0$, and again $D=3$.

◅ The general theory of relativity asserts that in the absence of matter, the local geometry of space tends to be flat and Euclidean, with the presence of matter making it locally Riemannian. Here we could speak of a globally flat Universe of dimension 3 with local $D<3$. This type of disturbance is considered in Selety 1924, an obscure reference which fails to refer to Koch but includes (p. 312) an example of the construction of Chapter 6. ►

THE FOURNIER UNIVERSE

It remains to construct a fractal that satisfies $M(r) \propto R^D$, and see how it agrees with accepted views concerning the Universe. The first fully described model of this kind is due to E. E.

Fournier d'Albe (Chapter 40). While Fournier 1907 is largely a work of fiction disguised as science, it also contains genuinely interesting considerations to which we come momentarily. It is best, however, to first describe the structure it proposes.

Its construction begins with the centered regular octahedron whose projection is represented near the center of Plate 95. The projection reduces to the four corners of a square whose diagonal is set to be of length 12 "units," and to this square's center. But the octahedron also includes two points above and below our plane, on the perpendicular drawn from the center of the square, and at the same distance of 6 units from this center.

Now, each point is replaced with a ball of radius 1, to be viewed as "stellar aggregate of order 0." And the smallest ball including the basic 7 balls is to be called a "stellar aggregate of order 1." An aggregate of order 2 is achieved by enlarging an aggregate of order 1 in the ratio $1/r=7$ and by replacing each of the resulting balls of radius 7 by a replica of the aggregate of order 1. In the same way, an aggregate of order 3 is achieved by enlarging an aggregate of order 2 in the ratio $1/r=7$ and by replacing each ball by a replica of the aggregate of order 2. And so on.

In sum, between two successive orders of aggregation, the number of points and the radius are enlarged in the ratio $1/r=7$. Consequently, whenever R is the radius of some aggregate, the function $M_0(R)$ expressing the number of points contained in a ball of radius R is $M_0(R)=R$. For intermediate values of R, $M_0(R)$ is smaller (reaching down to $R/7$), but the overall trend is $M_0(R)\propto R$.

Starting from aggregates of order 0, it is also possible to interpolate by successive stages to aggregates of orders -1, -2, and so on. The first stage replaces each aggregate of order 0 with an image of the aggregate of order 1, reduced in the ratio $1/7$, and so forth. If one does so, the validity of the relationship $M_0(R)\propto R$ is extended to ever smaller values of R. After infinite extra- and interpolation, we have a self-similar set with $D=\log 7/\log 7=1$.

We may also note that an object in 3-space for which $D=1$ need not be a straight line nor any other rectifiable curve. It need not even be connected. Each D is compatible with any lesser or equal value of the topological dimension. In particular, since the doubly infinite Fournier universe is a totally disconnected "dust," its topological dimension is 0.

DISTRIBUTION OF MASS; FRACTAL HOMOGENEITY

The step from geometry to the distribution of mass is obvious. If each stellar aggregate of order 0 is loaded with a unit mass, the mass $M(R)$ within a ball of radius $R>1$ is identical to $M_0(R)$, hence $\propto R$. Furthermore, to generate aggregates of order -1 from aggregates of order 0 amounts to breaking up a ball that had been viewed as uniform, and finding it to be made of seven smaller ones. This stage extends the rule $M(R)\propto R$ below $R=1$.

When viewed over the whole 3-space, the resulting mass distribution is grossly inhomogeneous, but over the Fournier fractal it is as homogeneous as can be. (Recall Plate 80.) In particular, any two geometrically identical portions of the Fournier universe carry identical masses. I propose that such a distribution of mass be called *fractally homogeneous.*

◅ The preceding definition is phrased in terms of scaling fractals, but the concept of fractal homogeneity is more general. It applies to any fractal for which the Hausdorff measure for the dimension D is positive and finite. Fractal homogeneity requires the mass carried by a set to be proportional to the set's Hausdorff measure. ►

FOURNIER UNIVERSE VIEWED AS CANTOR DUST. EXTENSION TO $D \neq 1$

I trust the reader was not distracted by the casual use of fractal terminology in the opening sections of this chapter. It is obvious that, without being aware of the fact, Fournier was traveling along a track parallel to that of Cantor, his contemporary. The main difference is that the Fournier construction is imbedded in space instead of the line. To further improve the resemblance, it suffices to change Fournier's aggregates from being balls to being bricks (filled-in cubes). Now, each aggregate of order 0 is a brick of side 1, and it includes 7 aggregates of side $1/7$: one of them has the same center as the initial cube, and the other six touch the central subsquares of the faces of the original cube.

Later we will examine how Fournier obtains the value $D=1$ from basic physical phenomena, and how Hoyle obtains this same value. Geometrically, however, $D=1$ is a special case, even if one preserves the overall octahedron and the value $N=7$. Since the balls do not overlap, $1/r$ can take any value between 3 and infinity, yielding $M(R) \propto R^D$, with $D=\log 7/\log(1/r)$ anywhere between 0 and $\log 7/\log 3=1.7712$.

Further, given any D satisfying $D<3$, it is easy by changing N to construct variants of Fournier's model having this dimension.

THE CHARLIER MODEL AND OTHER FRACTAL UNIVERSES

The above constructs share every one of the characteristic defects of first fractal models. Most conspicuously, just like the Koch curve model in Chapter 6 and the Cantor dust model in Chapter 8, the Fournier model is so regular as to be grotesque. As a corrective, Charlier 1908, 1922 suggests that one allow N and r to vary from one hierarchical level to another, taking on the values N_m and r_m.

The scientific eminence of Charlier was such that, despite the praise he lavished on Fournier, writing in the leading scientific languages of the day, even the simple model soon became credited to its famous expositor instead of its unknown author. It was much discussed in its time, in particular in Selety 1922, 1923a, 1923b, 1924. Furthermore, the

model attracted the attention of the very influential Emile Borel, whose comments in Borel 1922, while dry, were perceptive. But from then on, aside from fitful revivals, the model fell into neglect (for not very convincing reasons noted in North 1965, pp. 20–22 and 408–409). Nevertheless, it refuses to die. The basic idea was independently reinvented many times to this day, notably in Lévy 1930. (See the LÉVY entry in Chapter 40.) Most important, the fractal core notion of the Fournier universe is implicit in the considerations about turbulence and galaxies in von Weizsäcker 1950 (see Chapter 10), and in the model of the genesis of the galaxies due to Hoyle 1953, which will be discussed momentarily.

The basic fractal ingredient is also present in my models, Chapters 32 to 35.

In this light, the question arises of whether a model of galaxy distribution can *fail* to be a fractal with one or two cutoffs. I think not. If one agrees that the distribution must be scaling (for reasons to be elaborated in Chapter 11) and that the set on which matter concentrates is not a standard scaling set, it must be a fractal set.

Granted the importance of scaling, Charlier's nonscaling generalization of the Fournier model is ill-inspired. ◅ Incidentally, it allows $\log N_m / \log (1/r_m)$ to vary with m between two bounds; $D_{min} > 0$ and $D_{max} < 3$. We have here yet another theme: effective dimension need not have a single value, and may drift between an upper and a lower limit. This theme is picked up again in Chapter 15. ►

FOURNIER'S REASON TO EXPECT D=1

We now describe the impressive argument that leads Fournier 1907, p. 103, to conclude that D must be equal to 1. This argument is a strong reason for not forgetting its author.

Consider a galactic aggregate of arbitrary order with mass M and radius R. Using without misgivings a formula applicable to objects with spherical symmetry, assume that the gravitational potential on the surface is GM/R (G being the gravitational constant). A star falling on this universe impacts with the velocity V equal to $(2GM/R)^{1/2}$.

To paraphrase Fournier, an important conclusion may be drawn from the observation that no stellar velocity exceeds 1/300 of the velocity of light. It is that the mass comprised within a world ball increases as its radius, and not as its volume, or in other words, that the density within a world ball varies inversely as the surface of the ball... To make this clearer, the potential at the surface would be always the same, being proportional to the mass and inversely proportional to the distance. And as a consequence, stellar velocities approaching the velocity of light would not prevail in any part of the universe.

HOYLE CURDLING; THE JEANS CRITERION ALSO YIELDS D=1

A hierarchical distribution also arises in a theory advanced in Hoyle 1953, according to which galaxies and stars form by a cascade

process starting with a uniform gas.

Consider a gas cloud of temperature T and mass M_0, distributed with a uniform density throughout a ball of radius R. As shown by Jeans a "critical" situation prevails when $M_0/R_0=JkRT/G$. (Here, k is the Boltzmann constant and J a numerical coefficient.) In this critical case, the primordial gaseous cloud is unstable and must inevitably contract.

Hoyle postulates (a) that M_0/R_0 takes on this critical value at some initial stage, (b) that the resulting contraction stops when the volume of the gas cloud drops to 1/25-th, and (c) that each cloud then splits into five clouds of equal size, mass $M_1=M_0/5$, and equal radius $R_1=R_0/5$. Thus the process ends as it started: in an unstable situation followed by a second stage of contraction and subdivision, then a third, and so on. But curdling stops as clouds become so opaque that the heat due to gas collapse can no longer escape.

As in the diverse other fields where the same cascade process is encountered, I propose that the five clouds be called *curds*, and that the cascade process be called *curdling*. As said when I introduced this last term, I could not resist its juxtaposition with *galactic*.

Fournier injects N=7 to facilitate the graphical illustration, but Hoyle claims that N=5 has a physical basis. In another contrast with Fournier, whose geometrical illustration is detailed beyond what is reasonable or needed, Hoyle is vague about the curds' spatial scatter. An explicit implementation has to wait until we describe random curdling in Chapter 23. But these discrepancies do not matter: the main fact is that r=1/N, so that D=1 must be part of the design if curdling is to end as it began, in Jeans instability.

Further, if the duration of the first stage is taken to be 1, gas dynamics shows that the mth stage's duration is 5^{-m}. It follows that the same process could continue to infinity within a total time of 1.2500.

EQUIVALENCE OF THE FOURNIER AND HOYLE DERIVATIONS OF D=1

At the edge of an unstable gas cloud satisfying the Jeans criterion, the velocity and the temperature are linked by $V^2/2=JkT$, because GM/R is equal to $V^2/2$ (Fournier) and to JkT (Jeans). Now recall that in statistical thermodynamics the temperature of a gas is proportional to the mean square velocity of its molecules. Hence the combination of the Fournier and Jeans criteria suggests that at the edge of a cloud the velocity of the fall of a macroscopic object is proportional to the average velocity of its molecules. A careful analysis of the role of temperature in the Jeans criterion is bound to show the two criteria to be equivalent. ◅ Most likely, the analogy extends to the $M(R)\propto R$ relationship *within* galaxies, reported in Wallenquist 1957. ►

WHY D=1.23 AND NOT D=1?

The disagreement between the empirical D=1.23 and the Fournier and Hoyle theoreti-

cal D=1 raises an important issue. P. J. E. Peebles tackled it in 1974 by relativity theory. See Peebles 1980, a full treatment of the physics and of the statistics (but not of the geometry) of this topic.

THE SKY'S FRACTAL DIMENSION

The sky is a projection of a universe, in which every point is first described by its spherical coordinates ρ, θ, and ϕ and then replaced by the point of spherical coordinates 1, θ, and ϕ. When the universe is a fractal of dimension D, and the origin of the frame of references belongs to the universe (see Chapter 22), the structure of this projection is "typically" ruled by the following alternative: $D>2$ implies that the projection covers a nonzero proportion of the sky, while $D<2$ implies that the projection is itself of dimension D. ◅ As exemplified in Plates 95 and 96, *typical* allows for exceptions, due to the structure of the fractal and/or the choice of origin. It often means "true with probability 1." ►

ASIDE ON THE BLAZING SKY EFFECT (WRONGLY CALLED OLBERS PARADOX)

The rule in the preceding section bears directly upon the motivation that led diverse writers (including Fournier) to variants of a fractal Universe. They recognized that such universes "exorcise" geometrically the *Blazing Sky Effect*, often (but wrongly) called *Olbers paradox*. Under the assumption that the distribution of celestial bodies is uniform, meaning that D=3 for all scales, the sky is lit near uniformly, during the night and during the day, to the brighness of the solar disc.

This paradox is no longer of interest to physicists, having been eliminated by relativity theory and the theory of the expansion of the Universe, and other arguments. But its demise left a peculiar by-product: numerous commentators invoke their preferred explanation of the Blazing Sky Effect as an excuse for neglecting clustering, and even as an argument for denying its reality. This is a truly odd viewpoint: even if galaxies *need not* be clustered to avoid the Blazing Sky Effect, they *are* clustered, and this characteristic demands careful study. Furthermore, as seen in Chapter 32, the expansion of the Universe is compatible not only with standard homogeneity but also with fractal homogeneity.

The Blazing Sky argument is simplicity itself. When the light emitted by a star is proportional to its surface area, the amount of light reaching an observer at a distance of R is $\propto 1/R^2$, but the star's apparent surface is itself $\propto 1/R^2$. Thus, the apparent ratio of light to spherical angle is independent of R. Also, when the distribution of stars in the Universe is uniform, almost any direction in the sky sooner or later intersects some star. Therefore, the sky is uniformly bright, and seems ablaze. (The Moon's disc would form an exceptional *dark* domain, at least, in the absence of atmospheric diffusion.)

On the other hand, the assumption that the

universe is fractal with $D<2$ resolves the paradox. In that case, the universe's projection on the sky is a fractal with the same D, hence a set of zero area. Even if the stars are given a nonzero radius, a large proportion of directions go to infinity without encountering any star. Along these directions, the night sky is black. When the range in which $D<3$ is followed by a range in which $D=3$, the sky's background is not strictly black but illuminated extremely faintly.

The Blazing Sky Effect was noticed by Kepler shortly after Galileo's *Sidereal Message* had commented favorably on the notion that the Universe is unbounded. In his 1610 *Conversation with the Sidereal Messenger*, Kepler rejoined: "You do not hesitate to declare that there are visible over 10,000 stars... If this is true, and if [the stars have] the same nature as our sun, why do not these suns collectively outdistance our sun in brilliance?... But maybe the intervening ether obscures them? Not in the least... It is quite clear that...this world of ours does not belong to an undifferentiated swarm of countless others." (Rosen 1965, pp. 34–35.)

This conclusion remained controversial, but the argument was not forgotten, witness the comment by Edmund Halley, in 1720, that: "Another Argument I have heard urged, that if the number of Fixt Stars were more than finite, the whole superficies of their apparent Sphere would be luminous." Later, this conclusion was discussed by De Chéseaux and J. H. Lambert, but came to be credited to Gauss's great friend, Olbers. The term "Olbers paradox" that became attached to it is scandalous but symptomatic. Observations that had been rejected into the "unclassified residuum" (page 28) become all too often credited to the first Establishment figure who decorates them by a classifiable wrapping, however transient. Historical discussions are found in Gamow 1954, Munitz 1957, North 1965, Dickson 1968, Wilson 1965, Jaki 1969, Clayton 1975, and Harrison 1981.

ASIDE ON NEWTONIAN GRAVITATION

The Rev. Bentley kept pestering Newton with an observation closely related to the Blazing Sky Effect: if the stars' distribution is homogeneous, the force they exert on one among them is infinite. One may add that their gravitational potential is infinite. And that any distribution wherein $M(R) \propto R^D$ for large R yields an infinite potential unless $D<1$. The modern theory of potentials (Frostman theory) confirms that there is a privileged link between Newton's gravitation and the value $D=1$. The Fournier and Hoyle derivations of $D=1$ cannot fail to be related to this link. ◄ Fournier's theme of "the gravitational potential at the surface being always the same" is central to modern potential theory. ►

ASIDE ON RELATIVITY THEORY

◄ To paraphrase de Vaucouleurs 1970: "Relativity theory led us to believe that to be

optically observable, no stationary material ball can have a radius R less than the Schwarzchild limit $R_M=2GM/c^2$, where c is the velocity of light. In a plot of the mean density ρ and the characteristic radius R of various cosmical systems, $\rho_M = 3c^2/8\pi GR_M^2$ defines an upper limit. The ratio ρ/ρ_M may be called the Schwarzchild filling factor. For most common astronomical bodies (stars) or systems (galaxies), the filling factor is very small, on the order of 10^{-4} to 10^{-6}." The square of the velocity ratio postulated by Fournier is $(300)^{-2}\sim10^{-5}$, precisely in the range middle of the above. ►

AN AGGLUTINATED FRACTAL UNIVERSE?

Many authors think one may explain the genesis of stars and other celestial objects by an *ascending* cascade (i.e., the agglutination of greatly dispersed dust particles into increasingly bigger pieces) rather than by a *descending* cascade à la Hoyle (i.e., the fragmentation of very large and diffuse masses into smaller and smaller pieces).

An analogous alternative arises in connection with the cascades postulated in the study of turbulence, Chapter 10. Richardson's cascade descends toward ever smaller eddies, but ascending cascades may also be present; see Chapter 40, under RICHARDSON. Thus it may be hoped that the interrelations between descending and ascending cascades will be clarified soon.

FRACTAL TELESCOPE ARRAYS

To wind up this discussion, nothing can be more appropriate than a comment about the tools used to observe the galaxies. Dyson 1977 suggests that it may be advantageous to replace one piece telescopes by arrays of small telescopes. The diameter of each would be about 0.1 m, equal to the patch size of the smallest optically significant atmospheric disturbance, their centers would form a fractally hierarchical pattern, and they would be linked by Currie interferometers. A rough analysis leads to the conclusion that a suitable value for the dimension would be ⅔. Dyson's conclusion: "A 3-kilometer array of 1024 ten-centimeter telescopes connected by 1023 interferometers is not a practical proposition today. [It is offered] as a theoretical ideal, to show what can be done in principle."

SURVEY OF RANDOM FRACTAL MODELS OF GALAXY CLUSTERS

If one grants the claim that the distribution of galaxies is described usefully by unknowingly fractal models of limited subtlety and versatility, one should not be surprised that knowingly fractal random models provide even more useful descriptions. To begin with, our understanding of Hoyle curdling improves when it is set in its proper context: random fractals (Chapter 23). Of greater significance, I think, are the random models I developed and discuss in Chapters 32 to 35. One reason for

dwelling on several models is that improvement in the quality of description is "paid for" by increased complication. A second reason is that each model involves a fractal dust that deserves attention. Let me survey these models here, out of logical order.

Around 1965, my ambition was to implement the relationship $M(R) \propto R^D$ with $D<3$ with a model in which there is no "center of the universe." I first achieved this goal by the random walk model described in Chapter 32. Then, as an alternative, I developed a trema model, which consists in cutting out from space a collection of mutually independent *randomly placed* tremas of *random* radius, ranging up to an upper cutoff L that may be either finite or infinite.

Since both models had been selected solely on the basis of formal simplicity, it was delightfully surprising to discover they have predictive value. My theoretical correlation functions (Mandelbrot 1975u) agree with the curve-fitted ones reported in Peebles 1980 (see pp. 243-249). ◅ More precisely, my two approaches agree on the 2-point correlation, my random walk yields a good 3-point correlation and a bad 4-point correlation, and my spherical tremas model is very good for all known correlations. ►

Unfortunately, the appearance of samples generated by either model is quite unrealistic. Using a notion that I developed for this very purpose and describe in Chapter 35, they have unacceptable lacunarity properties. For the trema model this defect is corrected by introducing more elaborate trema shapes. For the random walk model, I use a less lacunar "subordinator."

Thus, the study of galaxy clusters has greatly stimulated the development of fractal geometry. And today the uses of fractal geometry in the study of galaxy clusters go well beyond the tasks of streamlining and housecleaning accomplished in the present chapter.

CUT DIAMONDS LOOK LIKE STARS

And the distribution of raw diamonds in the Earth's crust resembles the distribution of stars and galaxies in the sky. Consider a large world map on which each diamond mine or diamond rich site—past or present—is represented by a pin. Where examined from far away, these pins' density is extraordinarily uneven. A few are isolated here and there, but most concentrate in a few blessed (or accursed) areas. However, the Earth's surface in these areas is *not* uniformly paved with diamonds. When examined more closely, any of these areas turns out itself to be mostly blank, with scattered subareas of much greater diamond concentration. The process continues over several orders of magnitude.

Is it not irresistible to inject curdling in this context? Indeed, an unknowingly fractal model has been advanced by de Wijs, as seen under NONLACUNAR FRACTALS in Chapter 39. ■

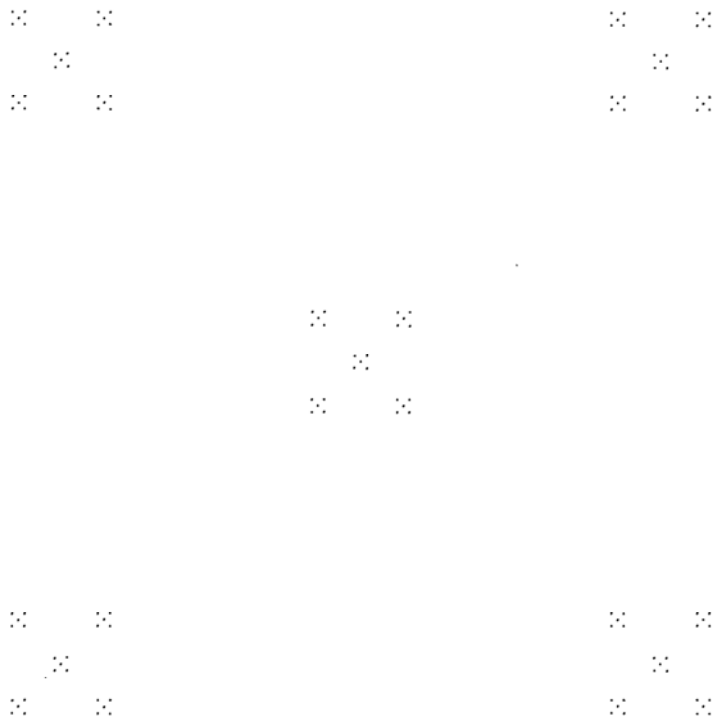


Plate 95 ¤ PROJECTION OF FOURNIER'S MULTIUNIVERSE (DIMENSION D~0.8270)

This plate represents to scale both the projection and the "equatorial" section of a Universe of dimension D=1 described in the text. See also Plate 96.

To paraphrase the caption in Fournier 1907: "A multiuniverse constructed upon a cruciform or octahedral principle is not the plan of the world but is useful in showing that an infinite series of similar successive universes may exist without producing a 'blazing sky.' *The matter in each world sphere is proportional to its radius.* This is the condition required for fulfilling the laws of gravitation and radiation. In some directions the sky would appear quite black, although there is an infinite succession of universes. The 'world ratio' in this case is N=7 instead of 10^{22}, as in reality."

In the sense described in Chapter 34, a universe with D=1 and N=10^{22} is of very low lacunarity, but extraordinarily stratified. ■

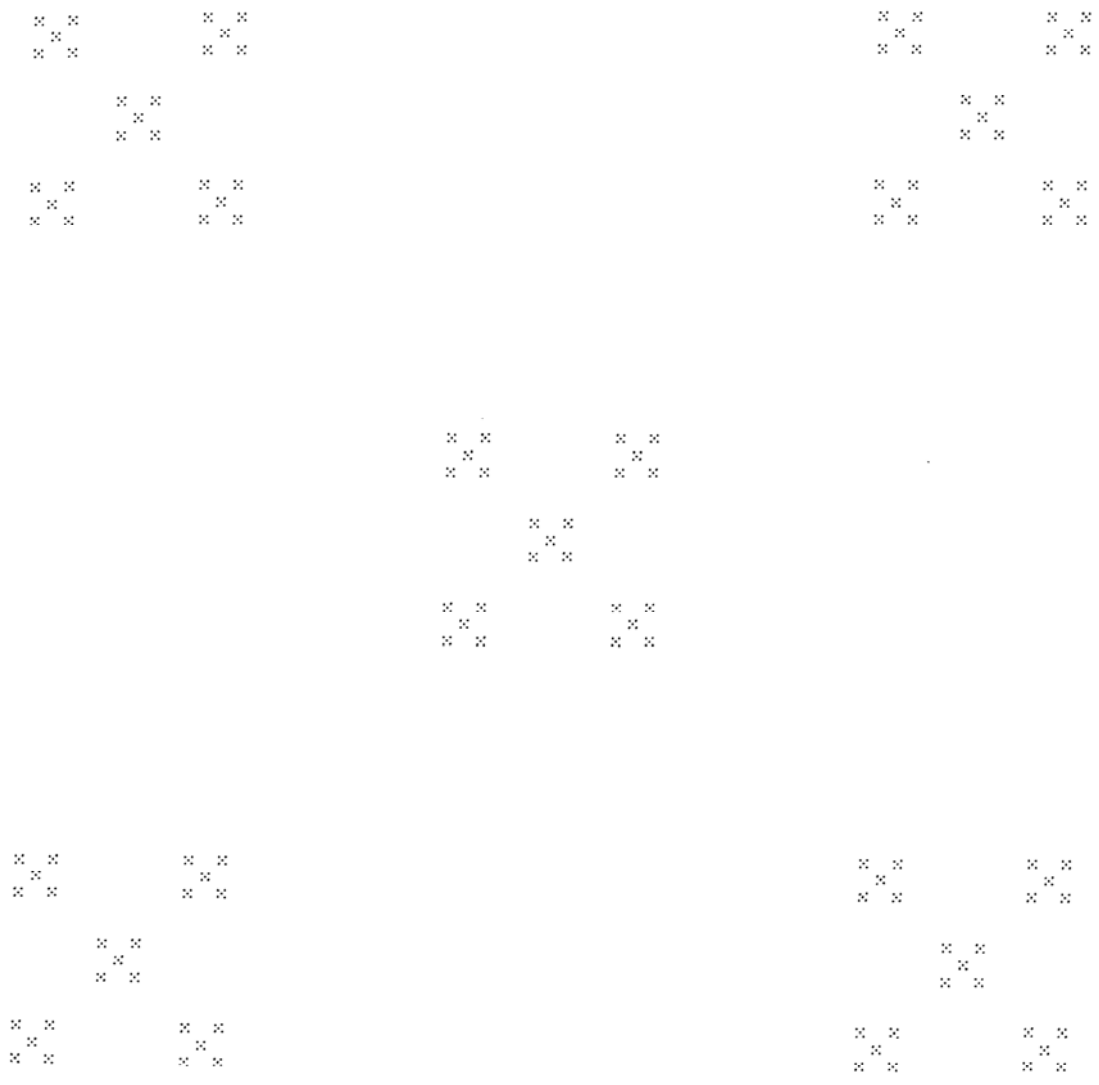


Plate 96 ¤ A FLAT FOURNIER UNIVERSE WITH D=1

Plate 95, being drawn to exact scale, is not only hard to print and to see, but potentially misleading. Indeed, it is *not* a universe of dimension D=1 but its planar projection, whose dimension is $D=\log 5/\log 7 \sim 0.8270 < 1$. Therefore, in order to avoid leaving the wrong impression, we hasten to exhibit a regular Fournier-like planar pattern of dimension D=1. The construction, which involves $1/r=5$ instead of $1/r=7$, is carried one step further than is possible in Plate 95. ■

10 ¤ Geometry of Turbulence; Intermittency

The study of turbulence is one of the oldest, hardest, and most frustrating chapters of physics. Common experience suffices to show that under certain circumstances the flow of a gas or a liquid is smooth, the technical term being "laminar," while under different circumstances it is not smooth at all. But where should we draw a line? Should the term "turbulence" denote all unsmooth flows, including much of meteorology and oceanography? Or is it better to reserve it for a narrow class, and, if so, for which one? Each scholar seems to answer these questions differently.

This disagreement does not matter here, because we focus on unquestionably turbulent flows, whose most conspicuous characteristic resides in the absence of a well-defined scale of length: they all involve coexistent "eddies" of all sizes. This feature can already be recognized in Leonardo's and Hokusai's drawings. It demonstrates that turbulence is necessarily foreign to the spirit of the "old" physics that focused upon phenomena having well-defined scales. But this same reason makes the study of turbulence of direct interest to us.

As some readers know, practically all investigations of turbulence concentrate upon the *analytic* study of fluid flow, and leave the *geometry* aside. I like to think that this lack of balance does not reflect a perceived lack of importance. In fact, many geometric shapes involved in turbulence are easily seen or made visible and cry out for a proper description. But they could not receive the attention they deserve until the development of fractal geometry. Indeed, as I immediately surmised, turbulence involves many fractal facets, of which I describe a few in this and later chapters.

Two disclaimers are necessary. First of all, we leave aside the problem of the onset of turbulence in a laminar fluid. There is strong reason to believe that this onset has fractal aspects of great importance, but they have not been clarified enough to be discussed here. Secondly, such periodic structures as Bénard cells and Kármán streets do not concern us here.

This chapter begins with pleas for a more geometric approach to turbulence and for the use of fractals. These pleas are numerous but each is brief, because they involve suggestions with few hard results as yet.

After that, we focus on the problem of intermittency, which I have investigated actively. My most important conclusion is that the region of dissipation, namely the spatial set on which turbulent dissipation is concentrated, can be modeled by a fractal. Measurements done for different purposes suggest that this region's D lies around 2.5 to 2.6, but probably below 2.66.

Unfortunately, the model cannot be pinpointed accurately, until we determine the topological properties of the region of dissipation. In particular, is it a dust, or a wiggly and branched curve (vortex tube), or a wiggly and layered surface (vortex sheet)? The first conjecture is unlikely, while the second and third suggest models akin to the ramified fractals of Chapter 14. But we are in no position to decide. Progress on the new fractal front does not help the old topological front at all. Our knowledge of the geometry of turbulence remains primitive indeed.

The bulk of this chapter requires no expertise. ◅ But the specialist will observe that part of fractal analysis of turbulence is the geometric counterpart of the analytic analysis of correlations and spectra. The relationship between turbulence and probability theory is an old story. Indeed, G. I. Taylor's earliest work was, after Perrin's Brownian motion, the second major influence on Norbert Wiener's creation of a mathematical theory of stochastic processes. Spectral analysis has long since "paid back" (with accrued interest) what it once borrowed from the study of turbulence and now it is time for the theory of turbulence to take advantage of the development of a sophisticated stochastic geometry. In particular, the Kolmogorov spectrum has a geometric counterpart examined in Chapter 30. ►

CLOUDS, WAKES, JETS, ETC.

A generic problem in the geometry of turbulence concerns the shape of the boundary of the region where some characteristic of the fluid is encountered. Striking examples are the billows upon billows which one finds in the ordinary (water) clouds, as well as in the clouds provoked by volcanic eruptions and in nuclear mushrooms. At this stage of this Essay, it is indeed difficult to escape the impression that, insofar as there is a range of scales wherein a cloud can be said to have a well-defined boundary, cloud boundaries *must* be fractal surfaces. The same remark applies to the patterns of rain squalls seen on radar screens. (For a first confirmation of this hunch, see Chapter 12.)

But I prefer to deal with simpler shapes. Turbulence may be restricted to a portion of an otherwise laminar fluid, say a wake or a jet. In the roughest approximation, each is a rod. If, however, the boundary is examined in detail, it reveals a hierarchy of indentations, whose depth increases with the value of the classic measure of hydrodynamic scale, called Reynolds number. This very visible and complex "local" structure does not evoke a rod as much as a rope with many loosely attached strings floating around. Its typical cross sec-

tion is not at all circular, but closer in shape to a Koch curve, and even closer to the most rugged among the coastlines with islands investigated in Chapters 5 and 28. In any event, a jet's boundary seems fractal. When vortex rings are present, their topology is of interest, but does not exhaust the structure.

The next comment requires the reader to have a mental picture of a wake, say, the lovely shape of a disabled tanker's oil spill. The "rod" that describes such a wake in the roughest approximation has a great deal of structure: it is not at all a cylinder, since its cross section broadens rapidly away from the ship, and its "axis" is not at all straight but shows meanders whose typical size again increases away from the ship.

Analogous features are found in the turbulence due to the shear between fluids masses rubbing past each other, as shown in Browand 1966 and Brown & Roshko 1974. The resulting coherent structures ("animals") attract wide attention, today. Fractals do not concern their overall form, but I think it is equally clear that the hierarchy of fine features that "ride" on the meanders is strikingly fractal in its structure.

Jupiter's celebrated red eye may also be an example of this sort.

Related but different problems arise when studying the Gulf Stream. It is not a single well-defined sea current but divides into multiple wiggly branches, and these branches themselves subdivide and ramify. An overall specification of its propensity to branch would be useful, and will doubtless involve fractals.

ISOTHERMS, DISPERSION ETC.

Similarly, it is interesting to study the shape of the surfaces of constant temperature or the isosurfaces of any other scalar characteristic of the flow. The isotherms may be delineated by the surface surrounding proliferating plankton that lives only in water at $T>45°$, and fills all the volume available to it. The boundary of such a blob is extremely convoluted; in the specific model in Chapter 30, it is demonstrably fractal.

A broad class of geometric problems occurs when a medium is completely filled by turbulence, but parts are marked by some "passive" or inert characteristic that does not affect the flow. The best example is when turbulence disperses a blob of color. Branches of all kinds shoot off in all directions, endlessly, but existing analyses and standard geometry are of little help in describing the resulting shapes. Plate 55 and Mandelbrot 1976c argue that these shapes must be fractals.

OTHER GEOMETRIC QUESTIONS

CLEAR-AIR TURBULENCE. Some scattered evidence I examined suggests that the set carrying this phenomenon is a fractal.

FLOW PAST A FRACTAL BOUNDARY. This is another typical case where fluid mechanics is bound to involve fractals (Plates 45 and 68).

VORTEX STRETCHING. Fluid motion forces vortices to stretch, and a stretching vortex must fold to accommodate an increasing

length in a fixed volume. To the extent that the flow is scaling, I conjecture the vortex tends toward a fractal.

THE TRAJECTORY OF A FLUID PARTICLE. In a crude approximation, inspired by the Ptolemaic model of planetary motion, let our particle be carried up vertically by an overall current of unit velocity, while it is perturbed by a hierarchy of eddies, each of which is a circular motion in a horizontal plane. The resulting functions $x(t)-x(0)$ and $y(t)-y(0)$ are sums of cosine and of sine functions. When the high frequency terms are very weak, the trajectory is continuous and differentiable, hence it is rectifiable and $D=1$. When, however, the high frequency terms are strong and continue down to 0, the trajectory is a fractal, with $D>1$. Assuming that eddies are self-similar, said trajectory happens to be identical to a famous counterexample of analysis: the Weierstrass function (Chapters 2, 39, and 41). This leads one to wonder whether or not the transition of all the fluid to being turbulent can be associated with the circumstances under which the trajectory is a fractal.

THE INTERMITTENCY OF TURBULENCE

Turbulence eventually ends in dissipation: due to the fluid's viscosity, the energy of visible motion transforms into heat. Early theories assume that the dissipation is uniform in space. But the hope that "homogeneous turbulence" would be a sensible model was dashed by Landau & Lifshitz 1953–1959, which notes that some regions are marked by very high dissipation, while other regions seem by contrast nearly free of dissipation. This means that the well-known property of wind, that it comes in gusts, is also reflected—in more consistent fashion—on smaller scales.

This phenomenon, *intermittency*, was first studied in Batchelor & Townsend 1949, p. 253. See also Batchelor 1953, Section 8.3, and Monin & Yaglom 1963, 1971, 1975. Intermittency is particularly clear-cut when the Reynolds number is very large, meaning that the outer cutoff of turbulence is large relative to its inner cutoff: in the stars, the ocean, and the atmosphere.

The regions in which dissipation concentrates are conveniently described as *carrying* or *supporting* it.

The fact that this Essay brings together the intermittency of turbulence and the distribution of galaxies is natural and not new. A while ago, physicists (von Weizäcker 1950) attempted to explain the genesis of the galaxies by turbulence. Recognizing that homogeneous turbulence cannot account for stellar intermittency, von Weizäcker sketched some amendments that are in the spirit of the Fournier ("Charlier") model (Chapter 9), hence of the theory presented here. If von Weizsäcker's unifying efforts are taken up again, they may establish a physical link between two kinds of intermittency and the corresponding self-similar fractals.

One goal of such a unifying effect would be to relate the dimension of the distribution

of galaxies, which we know to be D~1.23, with the dimensions involved in turbulence, which we noted lies around 2.5 to 2.7.

A DEFINITION OF TURBULENCE

We noted that, odd as it may seem, the same term, *turbulence*, is applied to several different phenomena. This continuing lack of a definition becomes easy to understand if, as I claim and propose to demonstrate, a proper definition requires fractals.

The customary mental image of turbulence is nearly "frozen" in the terms first isolated by Reynolds, about one hundred years ago, for fluid flow in a pipe: when the upstream pressure is weak, the motion is regular and "laminar"; when the pressure is increased sufficiently, everything suddenly becomes irregular. In this prototype case, the support of turbulent dissipation is either "empty," nonexistent, or is the entire tube. In either case there is not only no geometry to study, but also no imperative reason to define turbulence.

In wakes, things become more complicated. There is a boundary between the turbulent zone and the surrounding sea, and one ought to study its geometry. However, this boundary is again so clear that an "objective" criterion to define turbulence is not really necessary.

In fully developed turbulence in a wind tunnel, matters are again simple, the whole appearing turbulent like the Reynolds pipe. Nevertheless, the procedures used to achieve this goal are sometimes curious, if we are to believe certain stubbornly held stories. It is rumored that wind tunnels when first "blown" are unfit for the study of turbulence. Far from filling up the volume offered to it, turbulence itself seems "turbulent," presenting itself in irregular gusts. Only gradual efforts manage to stabilize the whole thing, after the fashion of the Reynolds pipe. Because of this fact, I am among those who wonder to what extent the nonintermittent "laboratory turbulence" in wind tunnels can be regarded as the same physical phenomenon as the intermittent "natural turbulence" in the atmosphere. Hence we must define the terms.

We approach this task indirectly, starting from an ill-defined concept of what is turbulent and examining the one-dimensional records of the velocity at a point. The motions of the center of gravity of a large airplane illustrate a rough analysis of such records. Every so often, the airplane is shaken about, which shows that certain regions of the atmosphere are strongly dissipative. A smaller airplane acts as a more sensitive probe: it "feels" turbulent gusts that leave the large airplane undisturbed, and it experiences each shock received by the large airplane as a burst of weaker shocks. Thus, when a strongly dissipative piece of the cross section is examined in detail, laminar inserts become apparent. And further smaller inserts are seen when the analysis is refined further.

Each stage demands a redefinition of what is turbulent. The notion of a *turbulent minute of record* becomes meaningful if interpreted as "minute of record that is not completely

free of turbulence." On the other hand, the more demanding notion of a *solidly turbulent minute of record* seems devoid of observable significance. Proceeding to successive stages of analysis, turbulence becomes increasingly sharp over an increasingly small fraction of the total record length. The volume of the support of dissipation seems to decrease. Our next task is to model this support.

ROLE OF SELF-SIMILAR FRACTALS

As already said, it is not surprising, in my view, that very few geometric aspects of turbulence have actually been investigated, because the only available techniques have been Euclidean. To escape their limitations, many pre-Euclidean terms are used. For example, papers on intermittency make an uncommonly heavy use of terms such as *spotty* and *lumpy*, and Batchelor & Townsend 1949 envisions "only four possible categories of shapes: blobs, rods, slabs, and ribbons." Some lecturers (but few writers) also use the terms *beans, spaghetti,* and *lettuce,* an imaginative terminology that does not attempt to hide the poverty of the underlying geometry.

By contrast, the investigations I carried out since 1964, and first presented at the 1966 Kyoto Symposium (Mandelbrot 1967k), augment the classical geometric toolbox by the addition of self-similar fractals.

To advocate the use of fractals is a radical new step, but to restrict the fractals of turbulence to be self-similar is orthodox, because the very notion of self-similarity was first conceived to describe turbulence. The pioneer was the Lewis Fry Richardson whom we first encounter in Chapter 5. Richardson 1926 introduced the concept of a hierarchy of eddies linked by a cascade. (See Chapter 40.)

It is also in the context of turbulence that the theory of cascades and of self-similarity achieved its triumphs of prediction between 1941 and 1948. The main contributors were Kolmogorov, Obukhov, Onsager, and von Weizsäcker, but tradition denotes the developments of the period by Kolmogorov's name. However, a subtle change occurred between Richardson and Kolmogorov.

While self-similarity is suggested by the consideration of visually perceived eddies, the Kolmogorov theory is purely *analytic.* Fractals, on the other hand, make it possible to apply the technique of self-similarity to the *geometry* of turbulence.

The fractal approach should be contrasted with the peculiar fact that the blobs, rods, slabs, and ribbons involved in yesterday's four-way choice fail to be self-similar. This may be why Kuo & Corrsin 1972 admit that this choice is "primitive" and that one needs *in-between* patterns.

A number of possible ad hoc changes in the standard patterns come to mind. For example, one might split rods into ropes surrounded with loose strands (remember the analogous situation with wakes or jets) and slice slabs into sheets surrounded with loose layers. Somehow those strands and layers might be made self-similar.

However, an ad hoc injection of self-similarity has never been implemented, and I find it both unpromising and unpalatable. I prefer to follow an entirely different tack, allowing the overall shapes and the details of strand and layer to be generated by the same process. Since the basic self-similar fractals are devoid of privileged direction, our study leaves aside (for now) all the interesting geometric questions that combine turbulence with strong overall motion.

◅ Obukhov 1962 and Kolmogorov 1962 are the first analytic studies of intermittency. In immediate influence, they nearly matched the 1941 papers of the same authors, but they are seriously flawed, and their long run influence promises to be small. See Mandelbrot 1972j, 1974f, 1976o; Kraichnan 1974. ►

INNER AND OUTER CUTOFFS

Due to viscosity, the inner cutoff of turbulence is positive. And wakes, jets, and analogous flows clearly show a finite outer cutoff Ω. But the widespread current belief in the finiteness of of Ω should be subjected to criticism. Richardson 1926 claims that "observation shows that the numerical values [presumed to converge for samples of size about Ω] would depend entirely upon how long a volume was included in the mean. Defant's researches show that no limit is attained within the atmosphere." The meteorologists have discounted, then forgotten, this assertion, far too hastily to my mind. New data in Chapter 11 and the study of lacunarity in Chapter 34 add to my conviction that the matter is not yet closed.

CURDLING AND FRACTALLY HOMOGENEOUS TURBULENCE

In a rough preliminary stage, we may represent the support of turbulence by one of the self-similar fractals which the preceding chapters obtain through curdling. This curdling is a crude "de-randomized" form of the Novikov & Stewart model of Chapter 23. After a finite number m of stages of a curdling cascade, dissipation is distributed uniformly over $N=r^{-mD}$ out of r^{-3m} mth-order nonoverlapping subeddies, whose positions are specified by a generator. After a cascade has continued without end, the limit distribution of dissipation spreads uniformly over a fractal of dimension $D<3$. I propose that the limit be called *fractally* homogeneous turbulence.

G. I. Taylor's homogeneous turbulence is obtained for $D\to 3$. The salient fact is that curdling does not exclude $D=3$, but it allows the novel possibility $D<3$.

DIRECT EXPERIMENTAL EVIDENCE THAT INTERMITTENCY SATISFIES $D>2$

From the viewpoint of linear sections, wide classes of unbounded fractals behave very simply: the section is almost surely empty when $D<2$ and is nonempty with positive

probability when $D>2$. (Chapter 23 proves this result for a simple class of fractals.)

Had the set that supports turbulent dissipation satisfied $D<2$, the preceding statement should imply that nearly all experimental probes would slip between turbulent regions. The fact that such is not the case suggests that in reality $D>2$. This inference is extraordinarily strong, because it relies upon an experiment that is repeated constantly, and for which the possible outcomes are reduced to an alternative between "never" and "often."

A tentative topological counterpart $D_T>2$, Mandelbrot 1976o, is tempting, but too special to be recounted here.

GALAXIES & TURBULENCE COMPARED

The inequality $D>2$ for the set that supports turbulent dissipation, and the opposite inequality $D<2$ for the distribution of mass in the cosmos, Chapter 9, spring from the closely related effects of the sign of $D-2$ on the typical section of a fractal and on its typical projection on a plane or the sky. For the phenomenon studied in the present chapter, the section has to be nonempty. In Chapter 9, on the contrary, the Blazing Sky Effect is "exorcised" if the majority of straight lines drawn from the Earth *never* meet a star. This requires the stars' projection on the sky to be of vanishing area.

The contrast between the signs of $D-2$ in these two problems must have a vital bearing on a constrast between their structures.

(IN)EQUALITIES BETWEEN EXPONENTS (MANDELBROT 1967k, 1976o)

Many useful characteristics of fractally homogeneous turbulence depend solely upon D. This topic is studied in Mandelbrot 1976o, where intermittent turbulence is characterized by a series of conceptually distinct exponents linked by (in)equalities. ◅ The situation is reminiscent of critical point phenomena. ►

SPECTRUM (IN)EQUALITIES. The (in)equality first stated in Mandelbrot 1967k (which uses the notation $\theta=D-2$), is ordinarily expressed in terms of the spectrum of the turbulent velocity, but is here stated in terms of variance. In fractally homogeneous turbulence, the velocity v at point x satisfies

$$\langle[v(x)-v(x+r)]^2\rangle=|r|^{2/3+B},$$

where $B=(3-D)/3$.

In Taylor homogeneous turbulence, $D=3$, and B vanishes, leaving the classic Kolmogorov exponent $\frac{2}{3}$, which we meet again in Chapter 30.

Mandelbrot 1976o also shows that the more general model of weighted curdling, as described in Mandelbrot 1974f, involves the inequality $B\leq(3-D)/3$.

THE β MODEL. Frisch, Nelkin & Sulem 1978 grafts a pseudo dynamic vocabulary upon the geometry of fractally homogeneous turbulence, as described in Mandelbrot 1976o. The interpretation has proven helpful, but the mathematical arguments and the conclusions are identical to mine. The term "β-model"

given to their interpretation has gained some currency, and is often identified with fractal homogeneity.

THE TOPOLOGY OF TURBULENCE REMAINS AN OPEN ISSUE

The preceding chapters make it abundantly clear that the same value of D can be encountered in sets that differ in terms of topological connectedness. The topological dimension D_T yields a lower bound to the fractal dimension D, but this bound is frequently exceeded by such a wide margin as to be of no use. A shape with a fractal dimension D between 2 and 3 may be either "sheetlike," "linelike," or "dustlike," and can achieve configurations in such variety as to make it hard to coin or find names for them all. For example, even in fractal shapes that are most nearly ropelike, the "strands" can be so heavy that the result is really "more" than ropelike. Similarly, fractal near sheets are "more" than sheetlike. Also, it is possible to mix sheetlike and ropelike features at will. Intuitively, one might have hoped that some closer relationship should exist between fractal dimension and degree of connectedness, but this is a hope mathematicians lost between 1875 and 1925. We turn to a special problem of this kind in Chapter 23, but it may be said that the actual loose relationship between these structures is essentially unexplored territory.

The question of ramification, raised in Chapter 14, is also vital, but its impact on the study of turbulence is as yet unexplored.

KURTOSIS INEQUALITIES. Using a measure of intermittency called kurtosis, the issue of connectedness is tackled in Corrsin 1962, Tennekes 1968, and Saffman 1968. Ostensibly, those models deal with shapes that share the topological dimension of the plane (sheets) or the straight line (rods). However, they test the topology indirectly, through the exponent of a predicted power law relationship between the kurtosis and a Reynolds number. Unfortunately, this attempt fails because the kurtosis exponent is in fact dominated by diverse additional assumptions, and ultimately depends solely on the fractal dimension D of the shape generated by the model. Corrsin 1962 predicts a value of D equal to the topological dimension it postulates, $D_T=2$. The prediction is incorrect, expressing the fact that the data involve fractals, but this model does not. On the other hand, Tennekes 1968 postulates $D_T=1$ but yields the fractional value $D=2.6$, hence does involve an approximate fractal. Nevertheless, the attempted inference from the kurtosis to a combination of intuitive "shape" and topological dimension is unwarranted. ■

11 ¤ Fractal Singularities of Differential Equations

The present chapter concerns a first connection between the fractal geometry of Nature and the mainstream of mathematical physics. The topic is so vital that it deserves a separate chapter. Readers whose interests lie elsewhere should forge ahead.

A SPLIT IN TURBULENCE THEORY

A major defect of the current theoretical study of turbulence is that it separates into at least two disconnected parts. One part includes the successful phenomenology put forth in Kolmogorov 1941 (examined in greater detail in Chapter 30). And the other part includes the differential equations of hydrodynamics, due to Euler for nonviscous fluids, and to Navier (and Stokes) for viscous fluids. These two parts remain unrelated: If "explained" and "understood" mean "reduced to basic equations," the Kolmogorov theory is not yet explained or understood. And Kolmogorov has not helped solve the equations of fluid motion.

My assertion in Chapter 10, that turbulent dissipation is not homogeneous over the whole space, only over a fractal subset, may seem at first sight to make the gap even greater. *But I contended that the opposite is the case.* And there is increasing evidence in my favor.

THE IMPORTANCE OF SINGULARITIES

Let us review the procedure that allows an equation of mathematical physics to be solved successfully. Typically, one draws up a list that combines solutions obtained by solving the equation under special conditions, and solutions guessed on the basis of physical observation. Next, neglecting details of the solutions, one draws a list of elementary "singularities" characteristic of the problem. From then on, more complex instances of the equation can often be solved in the first approximation by identifying the appropriate singularities and stringing them together as required. This is how the student of calculus draws the graph of a rational function. Of

course, the standard singularities are standard Euclidean sets: points, curves, and surfaces.

CONJECTURE: THE SINGULARITIES OF FLUID MOTION ARE FRACTAL SETS (MANDELBROT 1976c)

In this perspective, I interpret the difficulties experienced in deriving turbulence from the Euler and Navier-Stokes solutions as implying that *no* standard singularity accounts for what we perceive intuitively to be the characteristic features of turbulence.

I contend instead (Mandelbrot 1976c) that the turbulent solutions of the basic equations involve singularities or "near singularities" of an entirely new kind. The singularities are locally scaling fractal sets, and the near singularities are approximations thereto.

An unspecific motivation for this contention is that, standard sets having proven inadequate, one may as well try the next best known sets. But more specific motivation is available.

NONVISCOUS (EULER) FLUIDS

FIRST SPECIFIC CONJECTURE. Part of my contention is that the singularities of the solutions of the Euler equations are fractal sets.

MOTIVATION. This belief relies on the very old notion that the symmetries and other invariances present in an equation "ought" to be reflected in the equation's solution. (For a self-standing, careful and eloquent description, see Chapter IV of Birkhoff 1960.) Of course, preservation of symmetries is by no means a general principle of Nature, hence one cannot exclude the possibility of "broken symmetry" here. I propose, however, that one try the consequences of symmetry preservation. Since the Euler equations are scale-free, the equations' typical solutions should also be scale-free, and the same should hold of any singularities they may possess. If the failure of past efforts is taken as evidence that the singularities are not standard points or lines or surfaces, they must be fractals.

It may of course happen that a scale is imposed by the boundary's shape and the initial velocities. It is, however, likely that the solutions' local behavior is ruled by a "principle of not feeling the boundary." Hence the solutions should be locally scaleless.

ALEXANDRE CHORIN'S WORK. Chorin 1981 provides strong support for my contention, by applying a vortex method to the analysis of the inertial range in fully developed turbulence. The finding is that the highly stretched vorticity collects itself into a body of decreasing volume, and of dimension $D \sim 2.5$ compatible with the conclusions in Chapter 10. The correction to the Kolmogorov exponents, $B = .17 \pm 0.03$, is compatible with experimental data. The calculations suggest that the solutions of Euler's equations in three dimensions blow up in a finite time.

Unpublished work of Chorin comes even closer to experiment: $2.5 < D < 2.6$.

VISCOUS (NAVIER-STOKES) FLUIDS

SECOND SPECIFIC CONJECTURE. Furthermore, I contended that the singularities of the solutions of the Navier-Stokes equations can only be fractals.

DIMENSION INEQUALITIES. Furthermore, we have the intuitive feeling that the solutions of the Navier-Stokes equations are necessarily smoother, hence less singular, than those of the Euler equations. Hence the further conjecture that the dimension is larger in the Euler than in the Navier-Stokes case. The passage to zero viscosity is doubtless singular.

NEAR SINGULARITIES. A final conjecture in the implementation of my overall contention concerns the peaks of dissipation involved in the notion of intermittency: they are Euler singularities smoothed out by viscosity.

V. SCHEFFER'S WORK. The examination of my conjectures for the viscous case was pioneered by V. Scheffer, recently joined by others in studying in this light a finite or infinite fluid subject to the Navier-Stokes equations with a finite kinetic energy at t=0.

Assuming that singularities are indeed present, Scheffer 1976 shows that they necessarily satisfy the following theorems. First, their projection over the time axis has at most the fractal dimension ½. Second, their projection on the space coordinates is at most a fractal of dimension equal to 1.

It turns out, after the fact, that the first of the above results is a corollary of a remark in an old and famous paper Leray 1934 ends abruptly after a formal inequality of which Scheffer's first theorem is a corollary, in fact merely a restatement. But is it fair to say "merely"? Restating a result in more elegant terminology is (for sound reasons) rarely viewed as a scientific advance, but I think that the present instance is different. The inequality in Leray's theorem was nearly useless until the Mandelbrot-Scheffer corollary placed it in proper perspective.

The almost routine uses of Hausdorff Besicovitch dimension in recent studies of the Navier-Stokes equations can all be traced back to my conjecture.

SINGULARITIES OF OTHER NONLINEAR EQUATIONS OF PHYSICS

The other phenomena which this Essay claims involve scaling fractals have nothing to do with either Euler or Navier and Stokes. For example, the distribution of galaxies is ruled by the equations of gravitation. But the symmetry preservation argument applies to all scaling equations. As a matter of fact, an obscure remark by Laplace (see the entry SCALING IN LEIBNIZ AND LAPLACE, Chapter 41) can now be construed (with 20/20 hindsight!) as pointing toward the theme of Chapter 9.

More generally, the singularities' fractal character is likely to be traceable to generic features shared by many different equations of mathematical physics. Can it be some very broad kind of nonlinearity? The issue is joined again, in different terms, in Chapter 20. ■

12 ¤ Length-Area-Volume Relations

Chapters 12 and 13 extend the properties of fractal dimension through numerous mini case studies of varying importance and increasing difficulty, and Chapter 14 shows that fractal geometry necessarily involves concepts beyond the fractal dimension.

The present chapter describes, and applies to diverse concrete cases, the fractal counterparts I developed for certain standard results of Euclidean geometry. They can be viewed as parallel to the fractal relations of the form $M(R) \propto R^D$ obtained in Chapters 6, 8, and 9.

STANDARD DIMENSIONAL ANALYSIS

From the facts that the circumferential length of a circle of radius R is equal to $2\pi R$, and the area of the disc bounded by the circle is πR^2, it follows that

$$(\text{length}) = 2\pi^{1/2}(\text{area})^{1/2}.$$

Among squares, the corresponding relation is

$$(\text{length}) = 4(\text{area})^{1/2}.$$

More generally, within each family of standard planar shapes that are geometrically similar and have different linear extents, the ratio $(\text{length})/(\text{area})^{1/2}$ is a number entirely determined by the common shape.

In space (E=3), length, $(\text{area})^{1/2}$, and $(\text{volume})^{1/3}$ provide alternative evaluations of the linear extent of the shape, and the ratio of any two of them is a shape parameter independent of the units of measurement.

The equivalence of different linear extents is very useful in many applications. And its extension when time and mass are added lead to a powerful technique, known to physicists as "dimensional analysis." (Birkhoff 1960 is a recommended exposition of its basic features.)

PARADOXICAL DIMENSIONAL FINDINGS

However, in increasingly numerous instances, the equivalence between alternative linear extents proves distressingly elusive. For example, mammalian brains satisfy

$$(\text{volume})^{1/3} \propto (\text{area})^{1/D},$$

with $D \sim 3$, *far above* the anticipated value of 2. In river drainage basins, Hack 1957 measures length along the main river, and finds

$$(\text{area})^{1/2} \propto (\text{length})^{1/D},$$

with D *definitely above* the anticipated value of 1. Early writers interpret this last result as implying that river basins fail to be self-similar, large ones being elongated and small ones being chubby. Unfortunately, this interpretation conflicts with the evidence.

The present chapter describes how I explain these and related findings in more convincing fashion. My tool is a new, fractal, length-area-volume relation.

FRACTAL LENGTH-AREA RELATION

To pinpoint the argument, consider a collection of geometrically similar islands with fractal coastlines of dimension $D > 1$. The standard ratio $(\text{length})/(\text{area})^{1/2}$ is infinite in this context, but I propose to show it has a useful fractal counterpart. We denote as G-length the coast length measured with a yardstick length of G, and as G-area the island area measured in units of G^2. Knowing the dependence of G-length upon G to be nonstandard, while the dependence of G-area is standard, we form the generalized ratio

$$(\text{G-length})^{1/D}/(\text{G-area})^{1/2}.$$

I claim that this ratio takes the same value for our geometrically similar islands.

As a result, there are two different ways of evaluating the linear extent of each island in units of G: the standard expression $(\text{G-area})^{1/2}$ but also the nonstandard $(\text{G-length})^{1/D}$.

The novel feature is that if G is replaced by a different yardstick length G' the ratio of the alternative linear extents is replaced by

$$(\text{G'-length})^{1/D}/(\text{G'-area})^{1/2},$$

which differs from the original one by a factor of $(G'/G)^{1/D-1}$.

As for the ratio of linear extents, it varies between one family of mutually similar bounded shapes and another, whether they are fractal or standard. Hence it quantifies one facet of the shapes' form.

Note that the length-area relation may be used to estimate the dimension of a fractal curve that bounds a standard domain.

PROOF OF THE RELATION. The first step is to measure each coastline length with the intrinsic area-dependent yardstick

$$G^* = (\text{G-area})^{1/2}/1000.$$

When we approximate each of our island coastlines by a polygon of side G^*, these polygons are also mutually similar, and their lengths are proportional to the standard linear extents $(\text{G-area})^{1/2}$.

Next replace G^* by the prescribed yardstick G. We know from Chapter 6 that the measured length changes in the ratio $(G/G^*)^{1-D}$. Hence,

$$\begin{aligned}(\text{G-length}) &\propto (\text{G-area})^{1/2}(G/G^*)^{1-D}\\ &= (\text{G-area})^{1/2-1/2(1-D)}G^{1-D}1000^{D-1}\\ &= (\text{G-area})^{1/2D}G^{1-D}1000^{D-1}.\end{aligned}$$

Finally, by raising each side to the power $1/D$, we obtain the relation I claimed.

HOW WINDING IS THE MISSOURI RIVER?

The preceding arguments also throw light on the measured river lengths. To define a length for the leading river of a drainage basin, we approximate the river's course by a wiggly self-similar line of dimension $D>1$ going from a point called source to a point called mouth. If all rivers as well as their basins are mutually similar, the fractal length-area argument predicts that

$$(\text{river's G-length})^{1/D} \text{ is proportional to } (\text{basin's G-area})^{1/2}.$$

Moreover, standard reasons predict that

$$(\text{basin's G-area})^{1/2} \text{ is proportional to (straight distance from source to mouth).}$$

Combining the two results, we conclude that

$$(\text{river's G-length})^{1/D} \text{ is proportional to (straight distance from source to mouth).}$$

Most remarkably, as already mentioned, Hack 1957 finds empirically that the ratio

$$(\text{river's G-length})/(\text{basin's G-area})^{.6}$$

is indeed common to all rivers. This indirect estimate of $D/2=.6$ yields $D=1.2$, reminiscent of the values inferred from coastline lengths. If one measures the degree of irregularity by D, the degrees of irregularity of local wiggles of the banks and of enormously global bends turn out to be identical!

However, for basins of area $>10^4$ km^2 and correspondingly long rivers, J. E. Mueller observes that the value of D goes down to 1. The two different values of D suggest that if one maps all basins on sheets of paper of the same size, maps of short rivers look about the same as maps of long rivers, but maps of *extremely long* rivers are more nearly straight. It may be that nonstandard self-similarity breaks around an outer cutoff Ω whose value is of the order of 100 km.

CUMULATIVE LENGTH OF A RIVER TREE. The preceding argument also predicts that the cumulative length of all the rivers in a drainage basin should be proportional to that basin's area. I am told this prediction is correct, but I

have no reference.

BACK TO GEOMETRY. For the rivers and watersheds relative to the "snowflake sweep" curve of Plates 68 and 69, D~1.2618, somewhat above the observed value. The corresponding dimensions in Plates 70 and 71 are D~1.1291, on the low side.

The Peano curves of Plates 63 and 64 are well off the mark, since D=1.

Note that the identity between the dimensions of the rivers and of the watersheds is not a logical necessity, only a feature of certain specific recursive models. By way of contrast, a river network linked with the arrowhead curve (Plate 141) and described in Mandelbrot 1975m involves rivers of dimension D=1, which is too small, and watersheds of dimension D~1.5849, which is too large.

GEOMETRY OF RAIN AND OF CLOUDS

Pages 1, 10, 11, and 94 mention the possible use of fractals to model clouds. This hunch has now been confirmed by Lovejoy 1982, via the fractal area-perimeter graph in Plate 115. Very few graphs in meteorology involve all the available data over an enormous range of sizes, and are nearly as straight as this one.

The data combine radar observations from tropical Atlantic rain areas (with rainrate above .2 mm/hr), with geostationary satellite infrared observations of cloud areas over the Indian Ocean (= areas where the top of the cloud temperature is below −10°C). The areas range from 1 to over 1,000,000 km^2. The dimension of the perimeter, fitted over at least six orders of magnitude, is 4/3. The pleasure of providing a physical explanation is left to Dr. Lovejoy.

The largest cloud extended from central Africa to South India, a distance well above the thickness of the atmosphere, to which the outer cutoff L of atmospheric turbulence is all too often assimilated. Richardson's quote on p. 103 may prove prophetic.

THE AREA-VOLUME RELATION. CONDENSATION BY MICRO-DROPLETS

The derivation of the length area relationship generalizes easily to spatial domains bounded by fractal surfaces, and leads to the relation

$$(\text{G-area})^{1/D} \propto (\text{G-volume})^{1/3}.$$

To illustrate this relation, consider the condensation of vapor into liquid. This is a very familiar physical phenomenon, yet its theory is a recent development. To paraphrase Fisher 1967, the following geometric picture was put forward apparently quite independently by J. Frenkel, W. Band, and A. Bijl in the late 1930's. A gas consists of isolated molecules well separated from one another, except for occasional clusters which are bound together more-or-less tightly by the attractive forces. Clusters of different sizes are in mutual statistical equilibrium, associating and dissassociating, but even fairly large clusters resembling "droplets" of liquid have a small

chance of occurring. For a large enough cluster (which is not too "drawn out," like a piece of seaweed for example!), the surface area is fairly well defined. The surface of a cluster gives it stability. If the temperature now is lowered, it becomes advantageous for clusters to combine to form droplets and for droplets to amalgamate, thereby reducing the total surface area and hence lowering the total energy. If conditions are favorable, the droplets grow rapidly. A macroscopic droplet's presence indicates that condensation has taken place!

Building on this picture, M. E. Fisher proposes that a condensing droplet's area and volume are related by a formula equivalent to $\text{area}^{1/D} = \text{volume}^{1/3}$. Fisher evaluates D analytically without concern for its geometric meaning, but it is unavoidable that one should now conjecture that the underlying droplet surfaces are fractals of dimension D.

MAMMALIAN BRAIN FOLDS

To illustrate the area-volume relation in the important limit case D=3, and at the same time to buttress the exorcism of Peano shapes presented in Chapter 7, let us interpret a famous problem of comparative anatomy in terms of near-space-filling surfaces.

Mammalian brain volumes vary from 0.3 to 3000 ml, small animals' cortex being relatively or completely smooth, while large animals' cortex tends to be visibly convoluted, irrespective of the animals' positions on the scale of evolution. Zoologists argue that the proportion of white matter (formed by the neuron axons) to gray matter (where neurons terminate) is approximately the same for all mammals, and that in order to maintain this ratio a large brain's cortex must necessarily become folded. Knowing that the extent of folding is of purely geometric origin relieves Man from feeling threatened by Dolphin or Whale: they are bigger than us but need not be more highly evolved.

A quantitative study of such folding is beyond standard geometry but fits beautifully in fractal geometry. The gray matter's volume is roughly equal to its thickness multiplied by the area of the brain's surface membrane, called "pia." If the thickness ϵ were the same in all species, the pia area would be proportional not only to the gray matter volume but also to the white matter volume, hence to the total volume V. Therefore, the area-volume relationship would yield D=3, and the pia would be a surface that comes within ϵ of filling the space.

The empirical area-volume relation is better fitted by $A \propto V^{D/3}$ with $D/3 \sim 0.91$ to 0.93 (Jerison, private communication, based on the data of Elias & Schwartz, Brodman, and others). The most immediate interpretation is that the pia is only partly space filling, with D in the range between 2.79 and 2.73. A more sophisticated argument is sketched when we resume this topic in Chapter 17.

ALVEOLAR AND CELL MEMBRANES

Will a biologist kindly stand up and proclaim that the preceding section brings no hard result and no unexpected notion? I delight at hearing this objection because it buttresses further the argument with which Chapter 7 begins. Despite the fact that a biologist would run a mile from a Peano surface as adorned by mathematicians, I claim that the basic idea is indeed quite familiar to the good theoretical minds in this field.

Thus, the main novelty of the preceding sections lies with surfaces of $D<3$, which (as we saw) are required for a good fit. Let us pursue their novel application to biology by sketching how they help unscramble the detailed structure of several living membranes.

First, a paragraph to summarize Weibel 1979, section 4.3.7. Estimates of the human lung's alveolar area are conflicting: light microscopy yields 80 m^2, while electron microscopy claims 140 m^2. Does this discrepancy matter? The fine details to which it is due play no role with respect to gas exchange, being smoothed by a fluid lining layer (resulting in an even smaller functional area), but they are important with respect to solute exchanges. Measurements (triggered by my *Coast of Britain* paper) indicate in the first approximation that over a wide range of scales the membrane dimension is $D=2.17$.

Paumgartner & Weibel 1979 examine *sub*cellular membranes in liver cells. Again, the sharp past disagreement between different estimates of area per volume disappear by postulating that $D=2.09$ for the outer mitochondrial membrane (which wraps the cell, and departs only slightly from the smoothness characteristic of membranes with minimal area/volume ratio). On the other hand, $D=2.53$ for *inner* mitochondrial membranes, and $D=1.72$ for the endoplasmic reticulum.

Also let it be noted that many animals' nasal bone structure is of extraordinary complication, allowing the "skin" that covers this bone to have a very large area in a small volume. In Deer and Arctic Fox, this membrane *may* serve the sense of smell, but (Schmidt-Nielsen 1981) the goal of an analogous shape in Camel is to husband scarce water.

MODULAR COMPUTER GEOMETRY

To illustrate the area-volume relationship further, let us tackle a facet of computers. Computers are not natural systems, but this should not stop us. This and a few other case studies help demonstrate that, in the final analysis, fractal methods can serve to analyze any "system," whether natural or artificial, that decomposes into "parts" articulated in a self-similar fashion, and such that the properties of the parts are less important than the rules of articulation.

Complex computer circuits are always subdivided into numerous modules. Each contains a large number C of components and is connected with its environment by a large number T of terminals. Within an error of a few percent, one finds that $T^{1/D} \propto C^{1/E}$. The way

the exponent is written will be justified in a moment. Within IBM, the above rule is credited to E. Rent; see Landman & Russo 1971.

The earliest raw data suggested $D/E=\frac{2}{3}$, a value that Keyes 1981 extrapolates to huge "circuits" in the nervous system (optic nerve and *corpus callosum*). However, the ratio D/E increases with the circuit's performance. Performance, in turn, reflects the degree of parallelism that is present in the design. In particular, the designs with extreme characteristics lead to extreme values of D. In a shift register, the modules form a chain and $T=2$, independently of C, hence $D=0$. With integral parallelism, each component requiring its own terminal, $T=C$, hence $D=E$.

To account for $D/E=\frac{2}{3}$, R.W. Keyes noted that components are typically arranged within the volume of the modules, while the connections go through their surfaces. To show that this observation demands Rent's rule, it suffices to assume that all the components have roughly the same volume v and surface σ. Since C is the total volume of the module divided by v, $C^{1/3}$ is roughly proportional to the radius of the module. On the other hand, T is the total surface of the module divided by σ, thus $T^{1/2}$ is also roughly proportional to the radius of the module. Rent's rule simply expresses the equivalence of two different measures of the radius in a standard spatial shape. $E=3$ is the Euclidean dimension of the circuit and $D=2$ is the dimension of a standard surface.

Note that the concept of the module is ambiguous and almost indefinite, but Rent's rule is quite compatible with this characteristic, insofar as any module's submodules are interconnected by their surfaces.

It is just as easy to interpret the extreme cases mentioned above. In a standard linear structure, $E=1$ and the boundary reduces to two points, hence $D=0$. In a standard planar structure, $E=2$ and $D=1$.

However, when the ratio E/D is neither $3/2$, nor $2/1$, nor $1/0$, standard Euclidean geometry does not make it possible to interpret C as an expression of volume and T as an expression of surface. Yet such interpretations are very useful, and in fractal geometry they are easy. In a spatial circuit in contact with the outside by its whole surface, $E=3$, and D is anywhere between 2 and 3. In a plane circuit in contact with the outside by its whole bounding curve, $E=2$ and D is anywhere between 1 and 2. The case of integral parallelism, $D=E$, corresponds to Peano boundaries. Furthermore, if the boundary is utilized incompletely, the "effective boundary" may be any surface with D between 0 and E. ■

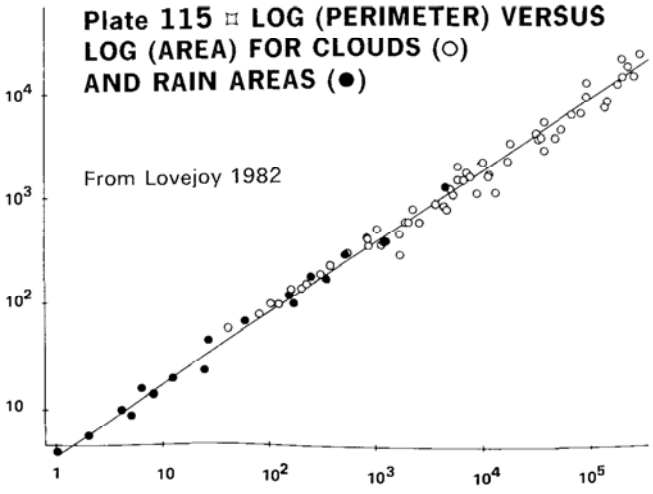


Plate 115 ¤ LOG (PERIMETER) VERSUS LOG (AREA) FOR CLOUDS (○) AND RAIN AREAS (●)

13 ¤ Islands, Clusters, and Percolation; Diameter-Number Relations

This chapter is devoted to fractal σ-curves, that is, to fractals that decompose into an infinity of disjoint fragments, each of them a connected curve. The concrete cases range from the coastlines of islands in an archipelago to an important problem of physics: percolation. The material in the first few sections was new to the 1977 *Fractals*, and the bulk of the chapter's remainder is new.

To begin, let us echo "How Long Is the Coast of Britain" and ask how many islands surround Britain's coast? Surely, their number is both very large and very ill-determined. As increasingly small rock piles become listed as islands, the overall list lengthens, and the total number of islands is practically infinite.

Since earth's relief is finely "corrugated," there is no doubt that, just like a coastline's length, an island's total area is geographically infinite. But the domains surrounded by coastlines have well defined "map areas." And the way in which a total map area is shared among the different islands is an important geographic characteristic. One might even argue that this "area-number relation" contributes more to geographic form than do the shapes of the individual coastlines. For example, it is difficult to think of the Aegean Sea's shores without also including those of the Greek islands. The issue clearly deserves a quantitative study, and this chapter provides one, by generalizing the Koch curves.

Next, this chapter examines diverse other fragmented shapes obtained by generalizing the familiar fractal-generating processes: either the Koch procedure or curdling. The resulting shapes are called *contact clusters* here, and the diameter-number distribution is shown to be the same for them as for islands.

Special interest attaches to the plane-filling contact clusters, in particular those clusters generated by certain Peano curves, whose teragons do not self-intersect but have carefully controlled points of self-contact. The saga of the taming of Peano monsters is thereby enriched by a new scene!

Last but not least, this chapter includes the first part of a case study of the geometry

of percolation, a very important physical phenomenon also studied in Chapter 14.

KORČAK EMPIRICAL LAW, GENERALIZED

List all the islands of a region by decreasing size. The total number of islands of size above a is to be written as $Nr(A>a)$ ◅ patterned after the notation $Pr(A>a)$ of probability theory. ► Here, a is a possible value for an island map's area, and A denotes the area when it is of unknown value.

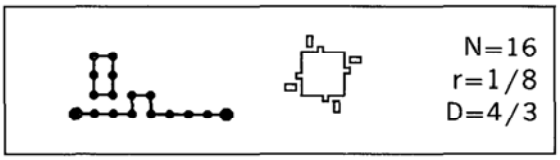
B and F' being two positive constants, to be called exponent and prefactor, one finds the following striking area-number relation:

$$Nr(A>a)=F'a^{-B}$$

Korčak 1938 (the name is pronounced Kor'chak) comes close to deserving credit for this rule, except that it claims that $B = ½$, which I found incredible, and which the data showed is unfounded. In fact, B varies between regions and is always $>½$. Let me now show that the above generalized law is the counterpart of the distribution Chapter 8 obtains for the gap lengths in a Cantor dust.

KOCH CONTINENT AND ISLANDS, AND THEIR DIVERSE DIMENSIONS

To create a Koch counterpart to the Cantor gaps, I let the generator split into disconnected portions. To insure that the limit fractal remains interpretable in terms of coastlines, the generator includes a connected broken line of $N_c<N$ links, joining the end points of the interval [0,1]. This portion will be called the *coastline generator*, because it determines how an initially straight coastline becomes transformed into a fractal coastline. The remaining $N–N_c$ links form a closed loop that "seeds" new islands and will be called *island generator*. Here is an example:



In later stages, the sub-island always stays to the left of the coastline generator (going from 0 to 1), and of the island generator (going clockwise).

A first novelty is that the limit fractal now involves two distinct dimensions. Lumping all the islands' coastlines together, $D = \log N/\log(1/r)$, but for the coastline of each individual island $D_c = \log N_c/\log(1/r)$, with the inequalities

$$1 \le D_c < D.$$

The cumulative coastline, not being connected, is not itself a curve but an infinite sum (Σ, sigma) of loops. I propose for it the term *sigma-loop*, shortened into *σ-loop*.

Note that modeling of the observed relation between D and D_c in actual islands requires additional assumptions, unless it can be

derived from a theory, as in Chapter 29.

THE DIAMETER-NUMBER RELATION

The proof that the Korčak law holds for last section's islands is simplest when the generator involves a single island, and teragons are self-avoiding. (Recall that the *teragons* are the approximating broken lines). Then the first stage of construction creates 1 island; let its "diameter," defined by $\sqrt{a}$, be λ_0. The second stage creates N islands of diameter $r\lambda_0$, and the mth stage creates N^m islands of diameter $\lambda=r^m\lambda_0$. Altogether, as λ is multiplied byr, $Nr(\Lambda>\lambda)$ is multiplied by N. Hence the distribution of Λ (for all values of λ of the form $r^m\lambda_0$) takes the form

$$Nr(\Lambda>\lambda)=F\lambda^{-D},$$

in which the crucial exponent is the coastline's fractal dimension! As a corollary

$$Nr(A>a) = F'a^{-B}, \text{ with } B=\tfrac{1}{2}D,$$

we have thus derived the Korčak law. For other values of λ or a, one has the staircase curve familiar from the distribution of Cantor gaps' lengths, Chapter 8.

This result is independent of N_c and D_c. It extends to the case when the generator involves two or more islands. We note that the empirical B regarding the whole Earth is of the order of 0.6, very close to one half of D measured from the coastline lengths.

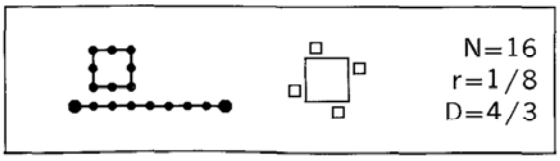
GENERALIZATION TO E>2

In the same construction extended to space, it continues to be true that the E dimensional diameter, defined as $volume^{1/E}$, is ruled by a hyperbolic expression of the form $Nr(volume^{1/E}>\lambda) = F\lambda^{-D}$, wherein the crucial exponent is D.

The exponent D also rules the special case of Cantor dusts for E=1, but there is a major difference. The length outside the Cantor gaps vanishes, while the area outside the "Koch" islands can be, and in general is, positive. We return to this topic in Chapter 15.

FRACTAL DIMENSION MAY BE SOLELY A MEASURE OF FRAGMENTATION

The preceding construction also allows the following generator



The overall D is unchanged, but the coastline D_c takes the smallest allowable value, $D_c=1$. In the present model, island coastlines are allowed to be rectifiable! When such is the case, the overall D is not a measure of irregularity, but solely of fragmentation. Instead of the wiggliness of individual curves, D measures the number-area relationship for an infinite family of rectangular islands.

It is still true that, when the length is measured with a yardstick of ϵ, the result tends to infinity as $\epsilon \to 0$, but there is a new reason for this. A yardstick of length ϵ can only measure islands with a diameter of at least ϵ. However, the number of such islands increases as $\epsilon \to 0$, and the measured length behaves like ϵ^{1-D}, exactly as in the absence of islands.

In the general case where $D_C > 1$, the value of D_C measures irregularity alone, while the value of D measures irregularity and fragmentation in combination.

A FRAGMENTED FRACTAL CURVE MAY HAVE TANGENTS EVERYWHERE. By rounding off the islands' corners, one may make every coastline have a tangent at every point, while the areas, hence the overall D, are unaffected. Thus, being a fractal σ-curve and being without tangent are *not* identical properties.

THE INFINITY OF ISLANDS

AN INNOCUOUS DIVERGENCE. As $a \to 0$, $Nr(A>a) = Fa^{-B}$ tends to infinity. Hence, the Korčak law agrees with our initial observation that islands are practically infinite in numbers.

LARGEST ISLAND'S RELATIVE AREA. This last fact is mathematically acceptable because the cumulative *area* of the very small islands is finite and negligible. ◅ All islands of area below ϵ have a total area that behaves like the integral of $a(Ba^{-B-1}) = Ba^{-B}$ from 0 to ϵ. Since $B<1$, this integral converges, and its value $B(1-B)^{-1}\epsilon^{1-B}$ tends to 0 with ϵ. ►

Consequently, the largest island's relative contribution to all the islands' cumulative area tends to a positive limit as the islands increase in numbers. It is *not* asymptotically negligible.

LONGEST COASTLINE'S RELATIVE LENGTH. On the other hand, assuming $D_C = 1$, the coastline lengths have a hyperbolic distribution with the exponent $D>1$. Hence the cumulative coastline length of small islands is infinite. And, as the construction advances and the number of islands increases, the coastline length of the largest island becomes relatively negligible.

RELATIVELY NEGLIGIBLE SETS. More generally, the inequality $D_C < D$ expresses that the curve drawn using the coastline generator alone is negligible in comparison to the whole coastline. In the same way, a straight line ($D=1$) is negligible in comparison to a plane ($D=2$). Just as a point chosen at random in the plane almost never falls on the x-axis, a point chosen at random on the coastline of a "core" island surrounded with sub-islands almost never falls on the core island's coastline.

SEARCH FOR THE INFINITE CONTINENT

In a scaling universe, the distinction between the islands and the continent cannot be based on tradition or "relative size." The only sensible approach is to define the continent as a special island with an *infinite* diameter. Let

me now show that the constructions at the beginning of this chapter *practically never* generate a continent. ◅ For those who know probability: the probability of a continent being generated is zero. ►

In a sensible search for a continent, we must no longer choose the initiator and the generator separately. From now on, the same generator must be made to serve both for interpolation and for extrapolation. The process runs by successive stages, each subdivided into steps. It strongly resembles the extrapolation of the Cantor set in Chapter 8, but deserves to be described even more thoroughly.

The first step upsizes our chosen generator in the ratio of $1/r$. The second step puts a "mark" on *one* of the links of the upsized generator. The third step displaces the upsized generator, to make its marked link coincide with [0,1]. The fourth and last step interpolates the upsized generator's remaining links.

The same process is repeated ad infinitum, its progress and outcome being determined by the sequence of positions of the "marked" links. This sequence can take diverse forms.

The first form requires the coastline generator to include a positive number N_c-2 of "nonextreme" links, defined as belonging to the coastline generator but not ending on either 0 or 1. If the mark is consistently put on a nonextreme link, each stage of extrapolation expands the original bit of coastline, and eventually causes it to be incorporated into a fractal coastline of infinite extent in both directions. This proves that it is indeed *possible* to obtain a continental coastline in this setup.

Secondly, always mark an extreme link of the coastline generator, each possibility being chosen an infinite number of times. Then our bit of coastline again expands without end. If we always choose the same link, the coastline expands in only one direction.

Thirdly, always mark a link that belongs to the island generator. Then the biggest island before extrapolation is made to lie off a bigger island's shore, then off-off a still bigger island's, and so on ad infinitum. No continent is *ever* actually reached.

The next comment involves a bit of "probabilistic common sense," which must be familiar to every reader. We suppose that the marks fall according to the throws of an N-sided die. In order for the extrapolation to generate a continent, it is obviously necessary that all the marks beyond a finite (kth) stage be placed upon one of N_c-2 nonextreme links of the coastline generator. Call them "winning" links. To know one has reached a continent after k stages, one must know that thereafter *every* throw of our die, with *not one* exception, will win. Such luck is not impossible, but it is of vanishing probability.

ISLAND, LAKE AND TREE COMBINATION

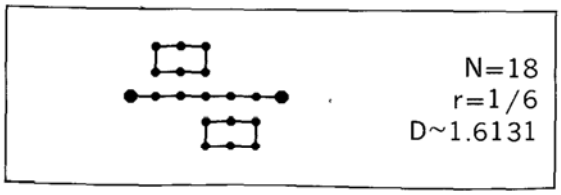
The Koch islands being mutually similar, their diameter Λ can be redefined as the distance between any two specified points, best chosen on the coastline. Next, we observe that the derivation of the diameter-number relation makes specific use of the assumption that

the generator includes a coastline generator. But the assumption that the generator's remaining links form islands, or are self-avoiding, is never actually used. Thus, the relation

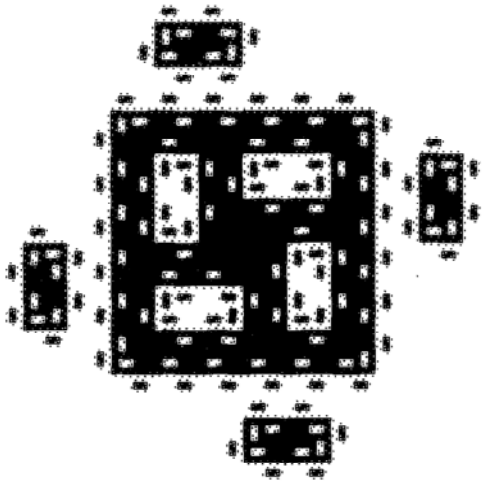
$$Nr(\Lambda>\lambda)=F\lambda^{-D}$$

is of very wide validity. ◅ One can even release the condition the teragons initiated by two intervals must not intersect. ► Let us now show by examples how the configuration of the original $N-N_C$ links can affect the resulting fractal's topology.

COMBINATION OF ISLANDS AND LAKES. Relieve the generator from the requirement of being placed to the left, going clockwise. When it is placed to the right, it forms lakes instead of islands. Alternatively, one may include *both* lakes and islands in the same generator. Either way, the final fractal is a σ-loop whose component loops are nested in each other. For example, consider the generator



When initiated by a square, this generator yields the following advanced teragon



THE ELUSIVE CONTINENT. In the above diagram, the length of the initiator's side injects a nonintrinsic outer cutoff. A more consistent approach is to extrapolate it as we did for islands without lakes. Again, it is almost sure that no continent is ever reached, and that the nesting of islands within lakes within islands continues without bound.

AREA-NUMBER RELATION. In order to define the area of an island (or lake), one may at will take either the total area, or the area of land (or water), within its coastline. The two differ by a fixed numerical factor, hence affect $Nr(A>a)$ through its prefactor F', not its exponent $\frac{1}{2}D$.

COMBINATION OF INTERVALS AND TREES. Now assume that the $N-N_C$ links form either a broken line with two free ends, or a tree. In either case, the fractal splits into an infinite

number of disconnected pieces, each of them a curve. This σ-curve is no longer a σ-loop; it is either a σ-tree or a σ-interval.

THE NOTION OF CONTACT CLUSTER

The generator may also combine loops, branches and diverse other topological configurations. If so, the limit fractals' connected portions recall the *clusters* of percolation theory (as seen later in this chapter) and of many other areas of physics. To us, this usage is terribly unfortunate, due to the alternative meaning of *cluster* in the study of dusts (Chapter 9). We need therefore a more specific and cumbersome term. I settled on "contact cluster." Luckily, the term σ-cluster is not ambiguous.

(It may be observed that *contact cluster* has a unique and natural mathematical definition, while the notion of clustering in a dust is diffuse and intuitive, and is at best defined via arguable statistical rules.)

PLANE-FILLING CONTACT CLUSTERS. As D reaches its maximum D=2, the arguments in the preceding section remain valid, but additional comments become necessary. Each individual cluster tends to a limit, which may be a straight line, but in most cases is a fractal curve. On the other hand, all the clusters together form a σ-curve, whose strands fill the plane increasingly tightly. The limit of this σ-curve behaves as in Chapter 7: it is no longer a σ-curve, but a domain of the plane.

THE ELUSIVE INFINITE CLUSTER. *No actually infinite* cluster is involved in the present approach. It is easy to arrange the generator's topology so that any given bounded domain is almost surely surrounded by a contact cluster. This cluster is in turn almost surely surrounded by a larger cluster, etc. There is no upper bound to cluster size. More generally, when a cluster seems infinite because it spans a very large area, the consideration of an even larger area will almost surely show it to be finite.

MASS-NUMBER AND WEIGHTED DIAMETER-NUMBER RELATIONS. THE EXPONENTS $D-D_c$ AND D/D_c.

Now let us reformulate the function $Nr(\Lambda>\lambda)$ in two ways: first by replacing a cluster's diameter λ by its mass μ, then by giving increased weight to large contact clusters.

Here, a cluster's mass is simply the number of links of length b^{-k} in the clusters itself (do *not* count the links *within* a looping cluster!). In effect, Chapters 6 and 12, we create a modified Minkowski sausage (Plate 33), by centering a square of side b^{-k} on each vertex, and adding half a square at each end-point.

The mass of a cluster of diameter Λ being the area of its modified sausage, $M \propto (\Lambda/b^k)^{D_c}(b^k)^2 = \Lambda^{D_c}/(b^k)^{D_c-2}$. Since $D_c<2$, $M \to 0$ as $k \to \infty$. The mass of all the contact clusters taken together is $\propto (b^k)^{D-2}$; if $D<2$, it too $\to 0$. And the relative mass of any individual contact cluster is $\propto (b^k)^{D_c-D}$; it tends to 0 at a rate that increases with $D-D_c$.

MASS-NUMBER RELATION. Clearly,

$$Nr(M>\mu) \propto (b^k)^{-D+2D/D_c}\mu^{-D/D_c}.$$

DISTRIBUTION OF DIAMETER WEIGHTED BY MASS. Observe that $Nr(\Lambda>\lambda)$ counts the number of lines above line λ in a list that starts with the largest contact cluster, continues with the next largest, etc. But we shall momentarily have to attribute to each contact cluster a number of lines equal to its mass. The resulting relation is easily seen to be

$$Wnr(\Lambda>\lambda) \propto \lambda^{-D+D_c}.$$

THE MASS EXPONENT $Q=2D_c-D$

Denote by $\mathcal{T}$ a fractal of dimension D, constructed recursively with $[0,\Lambda]$ as initiator, and take its total mass to be Λ^D. When $\mathcal{T}$ is a Cantor dust, Chapter 8 shows that the mass in a disc of radius $R<\Lambda$ centered at 0 is $M(R) \propto R^D$. ◅ The quantity $\log[M(R)R^{-D}]$ is a periodic function of $\log_b(\Lambda/R)$, but we shall not dwell on these complications because they vanish when the fractal is modified so that all $r>0$ are admissible self-similarity ratios. ►

We know that $M(R) \propto R^D$ also applies to the Koch curve of Chapter 6. Furthermore, this formula extends to the recursive islands and clusters of this chapter, with D replaced by D_c. In all cases, the mass in a disc of radius R centered at 0 takes the form

$$M(R,\Lambda) = R^{D_c}\phi(R/\Lambda),$$

with ϕ a function deducible from the shape of $\mathcal{T}$. In particular,

$$M(R,\Lambda)\propto R^{D_c} \text{ when } R<<\Lambda,$$
$$\text{and } M(R,\Lambda)\propto\Lambda^{D_c} \text{ when } R>>\Lambda.$$

Now consider the weighted average of M(R), to be denoted by $\langle M(R)\rangle$, corresponding to the case when Λ is variable with the widely spread-out hyperbolic distribution $Wnr(\Lambda>\lambda) \propto \lambda^{-D+D_c}$. We know that $1\le D_c<D\le 2$. Excluding the combination of $D=2$ and $D_c=1$, $0<D-D_c<D_c$. It follows that

$$\langle M(R)\rangle \propto R^Q \text{ with } Q=2D_c-D>0.$$

When the disc's center is a point of $\mathcal{T}$ other than 0, the factor of proportionality changes, but its exponent is unchanged. It also remains unchanged by averaging over all positions of the center in $\mathcal{T}$, and by the replacement of [0,1] by a different initiator. ◅ Usually, an arc of random size Λ is also of random shape. But the above formulas for $M(R,\Lambda)$ apply to $\langle M(R,\Lambda)\rangle$ averaged over all shapes. The final result is unchanged. ►

REMARK. The preceding derivation does not refer to the clusters' topology: they can be loops, intervals, trees, or anything else.

CONCLUSION. The formula $\langle M(R)\rangle\propto R^Q$ shows that, when Λ is hyperbolically distributed, hence of very wide scatter, one of the essential roles of dimension is taken up by an exponent *other than* D. The most natural exponent is $2D_c-D$, but different weighting function give different Q's.

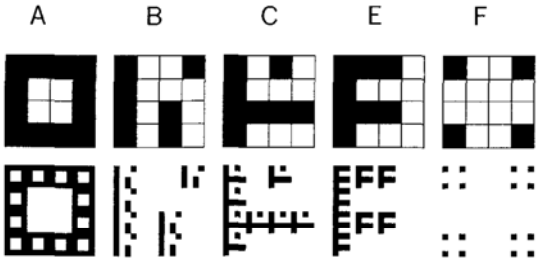
WARNING: NOT EVERY MASS EXPONENT IS A DIMENSION. The combined quantity Q is important. And, since it is a mass exponent, it is tempting to call it a dimension, but this temptation has no merit. Mixing many clusters with identical D_C but varying Λ leaves D_C unchanged, because dimension is *not* a property of a mixed population of sets, but a property of an individual set. Both D and D_C are fractal dimensions, but Q is not.

More generally, many areas of physics involve relations of the form $\langle M(R) \rangle \propto R^Q$, but such a formula does *not by itself* guarantee that Q is a fractal dimension. And calling Q an *effective dimension*, as some authors propose, is an empty gesture because Q does not possess any of the other properties that characterize D (for example, sums or products of D's have a meaning with no counterpart in the case of Q). Moreover, this empty gesture has proven a source of potential confusion.

NONLUMPED CURDLING CLUSTERS

We now proceed to describe two additional methods for generating contact clusters. One is based on curdling and applies for $D<2$, while the other is based on Peano curves and applies for $D=2$. The reader interested in percolation may skip this section and the next.

First, let us replace the Koch construction by the natural generalization of Cantor curdling to the plane. As illustration, consider the following five generators, with the next construction stage drawn underneath



In all these cases, the limit fractal is of zero area and contains no interior point. Its topology can take diverse forms, determined by the generator.

With generator A, the precurd of every stage k is connected, and the limit fractal is a curve, an example of the very important Sierpiński carpet examined in Chapter 14.

With generator F, the precurd splits into disconnected portions, whose maximum linear scale steadily decreases as $k \to \infty$. And the limit fractal is a dust, akin to the Fournier model of Chapter 9.

The generators B, C and E are more interesting: in their case, the precurd splits into pieces to be called *preclusters*. Each stage can be said to transform every "old" precluster by making it thinner and wigglier, and to give birth to "new" preclusters. Nevertheless, by deliberate choice of generators, each newborn precluster is entirely contained in a single smallest cell in the lattice prevailing before its birth. By contrast with the "cross lumped clusters" of the next section, the present ones are to be called "nonlumped." It follows that the limit contact clusters have a dimension of the form $\log N_C / \log b$, where N_C is an integer

at most equal to the number of cells in the generator's largest component. This maximum is attained for generators B and C, for which the contact clusters are, respectively, intervals with $D_c=1$ and fractal trees with $D_c=\log 7/\log 4$. But the fractal based on the generator E does not attain this maximum: in its case, the F-shaped preclusters keep splitting into parts, and the limit, again, is made of straight intervals with $D_c=1$.

Replacing the pseudo-Minkowski sausage by the collection of cells of side b^{-k} intersected by a contact cluster, the diameter-number relation and the other results of the preceding sections extend unchanged.

CROSS LUMPED CURDLING CLUSTERS

Next, let the generator of plane curdling takes either of the following shapes, with the next construction stages drawn to the side

Both cases exhibit massive "cross lumping," meaning that each newborn precluster combines contributions coming from several smallest lattice cells prevailing before its birth.

In the Koch context, an analogous situation prevails when the teragons are allowed to self-contact, resulting in the merger of small cluster teragons. In either case, the analysis is cumbersome, and we cannot dwell on it. But $Nr(\Lambda>\lambda)\propto\lambda^{-D}$ remains a valid relation for small λ.

◅ However, if one attempts to estimate D from this relation, without excluding the large λ's, the estimate is systematically biased and smaller than the true value. ►

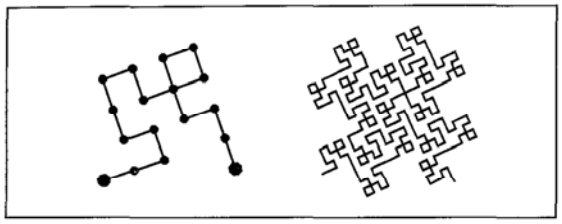
Novel features arise concerning the quantity b^{D_c}: it need not be an integer deducible from the generator by simple inspection, but it may be a fraction. The reason is that every contact cluster combines: (a) an integer number of versions of itself, downsized in the ratio $1/b$, and (b) many downsized versions due to lumping, which involve smaller ratios of the form $r_m=b^{-k(m)}$. The dimension-determining equation $\Sigma r_m{}^D=1$ of page 56, when rewritten in terms of $x=b^{-D}$, takes the form $\Sigma a_m x^m=1$. Cases where $1/x$ is an integer can only occur as exceptions.

KNOTTED PEANO MONSTERS, TAMED

A plane-filling collection of clusters (D=2) cannot be created by curdling, but I found an alternative approach, using Peano curves beyond those we saw being tamed in Chapter 7. As the reader must recall, Peano curves with self-avoiding teragons create river and watershed trees. But some other Peano curve teragons (for example the teragons in Plate 63, assuming that the corners are not rounded off) are simply chunks of lattice. As the construction proceeds, the open lattice cells sepa-

rated by such curves "converge" to an everywhere dense dust, e.g., to the points for which neither x nor y is a multiple of b^{-k}.

Between these extremes stands a new interesting class of Peano curves. Their generators are exemplified by the following, shown together with the next step



This class of Peano curves is now ready to be tamed. We observe that each point of self-contact "knots off" an open precluster, which may acquire branches and self-contacts, sees chunks of itself "knotted away," and eventually thins down to a highly ramified curve that defines a contact cluster. A cluster's diameter Λ, defined as in previous sections of this chapter, is fixed from the moment of birth: roughly equal to the side of the square that "seeded" this cluster. Its distribution is ruled by the familiar relation $\mathrm{Nr}(\Lambda>\lambda)\propto\lambda^{-2}$.

Observe in passing that, while Koch contact clusters are limits of recursively constructed curves, the present clusters are limits (in a peculiar sense) of the open components of the *complement* of a curve.

BERNOULLI PERCOLATION CLUSTERS

Whichever method is used to generate fractal contact clusters with $D=E$ and $D_C<D$, they provide a geometric model that had been lacking in a very important problem of physics: Bernoulli percolation through lattices. J. M. Hammersley, who posed and first investigated this problem, did not inject Bernoulli's name in this context, but the fractal percolation we encounter in Chapter 23 makes the full term unavoidable here. (It is independently adopted by Smythe & Wiermann 1975.)

LITERATURE. Bernoulli percolation is surveyed in Shante & Kirkpatrick 1971, Domb & Green 1972-, especially a chapter by J. W. Essam, Kirkpatrick 1973, deGennes 1976, Stauffer 1979, and Essam 1980.

DEFINITIONS. Percolation involves probabilistic notions, hence would not be discussed at this stage if we were entirely consistent. But an occasional lack of consistency has its rewards. The simplest percolation problem for $E=2$ is bond percolation on a square lattice. To illustrate it in homely fashion, imagine we construct a large square lattice with sticks made either of insulating vinyl or of conducting copper. A *Bernoulli lattice* obtains if each stick is selected at random, independently of the other sticks, the probability of choosing a conducting stick being p. Maximal collections of connected copper or vinyl sticks are called copper or vinyl *clusters*. When the lattice includes at least one uninterrupted string of copper sticks, the current can *flow through* from one side of the lattice to the other, and

the lattice is said to *percolate*. (In Latin, *per* = through, and *colare* = to flow.) The sticks in uninterrupted electric contact with the top and bottom sides of the lattice form a "percolating cluster," and the sticks actually active in conducting form the percolating cluster's "backbone."

The generalization to other lattices, and to $E>2$, is immediate.

CRITICAL PROBABILITY. Hammersley's most remarkable finding concerns the special role played by a certain threshold probability: the *critical probability* p_{crit}. This quantity enters in when the Bernoulli lattice's size (measured in numbers of sticks) tends to infinity. One finds that, when $p>p_{crit}$, the probability that there exists a percolating cluster increases with lattice size, and tends to 1. When $p<p_{crit}$, to the contrary, the probability of percolation tends to 0.

Bond percolation on square lattices being such that either copper or vinyl must percolate, $p_{crit} = \frac{1}{2}$.

ANALYTICAL SCALING PROPERTY. The study of percolation long devoted itself to the search for analytic expressions to relate the standard quantities of physics. All these quantities were found to be *scaling*, in the sense that the relations between them are given by power laws. For $p \neq p_{crit}$, scaling extends up to an outer cutoff dependent on $p-p_{crit}$ and denoted by ξ. As $p \to p_{crit}$, the cutoff satisfies $\xi \to \infty$. Physicists postulate (see Stauffer 1979, p. 21) that $\langle M(R,\Lambda) \rangle$ follows the rule obtained on p. 123.

THE CLUSTERS' FRACTAL GEOMETRY

THE CLUSTERS' SHAPE. Let $p=p_{crit}$, and let individual sticks decrease in size while the total lattice size remains constant. The clusters become increasingly thin ("all skin and no flesh"), increasingly convoluted, and increasingly rich in branches and detours ("ramified and stringy"). Specifically, Leath 1976, the number of sticks situated outside of the cluster, but next to a stick within the cluster, is roughly proportional to the number of sticks within the cluster.

HYPOTHESIS THAT CLUSTERS ARE FRACTALS. It is natural to conjecture that the property of scaling extends from analytic properties to the clusters' geometry. But this idea could not be implemented in standard geometry, because the clusters are *not* straight lines, Fractal geometry is of course designed to eliminate such difficulties: thus, I conjectured that clusters are representable by fractal σ-curves satisfying $D=2$ and $1<D_c<D$. This claim has been accepted, and found to be fruitful. It is elaborated upon in Chapter 36.

◅ To be precise, scaling fractals are taken to represent the clusters that are *not* truncated by the boundary of the original lattice. This excludes the percolating cluster itself. (The term *cluster* has a gift for generating confusion!) To explain the difficulty, start with an extremely large lattice, pick a cluster on it, and a smaller square that is spanned by this cluster. By definition, the intersection of this cluster and the smaller square includes a smaller percolating cluster, but in addition it

includes a "residue" that connects with the smaller percolating cluster through links *outside* the square. Note that neglect of this residue creates a downward bias in the estimation of D_C. ►

VERY ROUGH BUT SPECIFIC NONRANDOM FRACTAL MODELS. To be valid, the claim that any given natural phenomenon is fractal must be accompanied by the description of a specific fractal set, to serve as first approximation model, or at least as mental picture. My Koch curve model of coastlines, and the Fournier model of galaxy clusters, demonstrate that rough nonrandom picture may be very useful. Similarly, I expect recursively constructed contact clusters (like those introduced in this chapter) to provide useful fractal models of the ill-known natural phenomena that are customarily modeled by Bernoulli clusters.

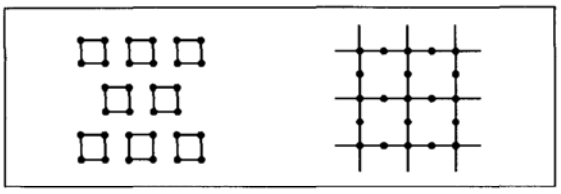
However, the Bernoulli clusters themselves are fully known (at least in principle), hence modeling them via explicit recursive fractals is a different task. The Koch contact clusters I studied are not suitable, due to dissymetry between vinyl and copper, even when there are equal numbers of sticks of both kinds. Next examine the knotted Peano curve clusters. Take an advanced teragon, and cover the cells to the left of the curve with copper, and the other cells with vinyl. The result involves a form of percolation applied to lattice cells (or to their centers, called sites). The problem is symmetric. But it differs from the Bernoulli problem, because the configuration of copper or vinyl cells are *not* the same as in the case of independence: for example, 9 cells forming a supersquare can all be of copper or vinyl in the Bernoulli case, but not in the knotted Peano curve case. (On the other hand, both models allow groups of 4 cells forming a supersquare to take any of the possible configurations.) This difference has far-reaching consequences: for example, neither copper nor vinyl percolate in the Bernoulli site problem with $p=\frac{1}{2}$, while both percolate in knotted Peano clusters, implying that $\frac{1}{2}$ is a critical probability.

The list of variants of Bernoulli bond percolation is already long, and can easily be lengthened further. And I have already examined many variants of recursively constructed fractal contact clusters. The detailed comparison of these lists is unfortunately complicated, and I shall not dwell on it here.

Let me therefore be satisfied with stating the loose conclusion that significant fractal essentials of the Bernoulli percolation problem seem to be illustrated by nonrandom space-filling σ-clusters defined earlier in this chapter. This model's principal weakness is that it is completely indeterminate beyond what has been said. It can accommodate any observed degree of irregularity and fragmentation. On the matter of topology, see Chapter 14.

MODEL OF CRITICAL CLUSTERS. Specifically, consider the critical clusters, defined as the clusters for $p=p_{crit}$. To represent them, a recursive σ-cluster is extrapolated as indicated in earlier sections of this chapter. Then it is truncated by stopping the interpolation so that the positive inner cutoff is the cell size in the original lattice.

MODELS OF NONCRITICAL CLUSTERS. To extend this geometric picture to noncritical clusters, that is, to clusters for $p \neq p_{crit}$, we seek fractals with a positive inner cutoff and a finite outer cutoff. Analysis calls for the largest copper cluster's extent to be of the order of ξ when $p<p_{crit}$, and to be infinite when $p>p_{crit}$. Either outcome is readily implemented. For example, one can start with the same generator as in the preceding subsection. But, instead of extrapolating it naturally, one initiates it with either of the following shapes



SUBCRITICAL CLUSTERS. The initiator to the left, which is geared towards $p<p_{crit}$, is made of squares of side $\frac{1}{2}\xi$. Now let the chosen generator be positioned *in* through each initiator's left side, and *out* through the other sides. The initiator square will transform into an atypical cluster of length ξ, surrounded by many typical clusters of length $<\xi$.

SUPERCRITICAL CLUSTERS. The initiator to the right, which is geared towards $p>p_{crit}$, is made of those lines of the initial square lattice, whose x or y coordinates are even integers. Four links radiate from each node whose coordinates are even integers; the chosen generator is always positioned to the left. In the special case when the coastline generator involves no loops nor dangling links, the resulting picture is a de-randomized and systematized variant of a crude model of clusters based solely on "nodes and links."

Observe that the fractal geometric picture deduces the noncritical clusters from the critical ones, while physicists prefer to consider the critical clusters as limits of the noncritical clusters for $\xi \to \infty$.

CRITICAL BERNOULLI CLUSTERS' D_C

The value of D_C is immediately inferred from either the exponent $D/D_C=E/D_C$ in the formula for $Nr(M>\mu)$, or the exponent $Q=2D_C-D=2D_C-E$ in the formula for $\langle M(R)\rangle$. Using the Greek letters τ, δ and η with the meanings customary in this context, we find that $E/D_C=\tau-1$ and $2D_C-E = 2-\eta$. Hence,

$$D_C = E/(\tau-1) = E/(1+\delta^{-1}),$$
$$\text{and } D_C=1+(E-\eta)/2.$$

Due to relations that physicists established between τ, δ and η, the above formulas for D_C are equivalent. Conversely, their equivalence does not reside in physics alone, because it follows from geometry.

Independently of each other, Harrison, Bishop & Quinn 1978, Kirkpatrick 1978, and Stauffer 1979 obtain the same D_C. They start from the properties of clusters for $p>p_{crit}$, hence express their result in terms of different critical exponents (β, γ, ν and σ). These derivations do not involve a specific underlying fractal picture. The dangers inherent in this

approach, against which we warned earlier in this chapter, are exemplified by the fact that it misled Stanley 1977 into advancing Q and D_C are equally legitimate dimensions.

For $E=2$, the numerical value is $D_C=1.89$. It is compatible with the empirical evidence, as obtained by a procedure familiar in other guises. Pick r, which need *not* be of the form $1/b$ (b an integer). Then take a big eddy, which is simply a square or cubic lattice of side set to 1. Pave it with subeddies of side r, count the number N of the squares or cubes that intersect the cluster, and evaluate $\log N/\log(1/r)$. Then repeat the process with each nonempty subeddy of side r by forming subsubeddies of side r^2. Continue as far as feasible. The most meaningful results obtain when r is close to 1. Some early simulations gave the biased estimate $D^+\sim 1.77$ (Mandelbrot 1978h, Halley & Mai 1979), but large simulations (Stauffer 1980) confirm D.

◅ The biased experimental D^+ is very close to Q, hence briefly seemed to confirm the theoretical arguments in Stanley, Birgenau, Reynolds & Nicoll 1976 and Mandelbrot 1978h, which were both in error in claiming that the dimension is Q. The error was brought to my attention by S. Kirkpatrick. A different and even earlier incorrect estimate of D is found in Leath 1976. ►

THE CYPRESS TREES OF OKEFENOKEE

When a forest that is not "managed" systematically is observed from an airplane, its boundary is reminiscent of an island's coastline. Individual tree patches' outlines are extremely ragged or scalloped, and each large patch is trailed by satellite patches of varying area. My hunch that these shapes may follow the Richardson and/or Korčak laws, is indeed confirmed by an unpublished study of the Okefenokee swamp (Kelly 1951) by H. M. Hastings, R. Monticciolo & D. VunKannon. The patchiness of cypress is great, with $D\sim 1.6$; the patchiness of broadleaf and mixed broadleaf trees is much less pronounced, with D near 1. My informants comment on the presence of an impressive variety of scales both on personal inspection and on examination of vegetation maps. There is an inner cutoff of about 40 acres, probably a consequence of aerial photography. ■

14 ¤ Ramification and Fractal Lattices

Chapter 6 investigates planar Koch curves that satisfy $D<2$ and are devoid of double points, hence can be called self-avoiding or nonramified. And Chapter 7 investigates Peano curves, for which everywhere dense double points are unavoidable in the limit. The present chapter takes the next step, and investigates examples of deliberately ramified self-similar shapes: planar curves with $1<D<2$, spatial curves with $1<D<3$, and surfaces with $2<D<3$. In a ramified self-similar curve, the number of double points is infinite.

This chapter's mathematics is old (though known to very few specialists), but my applications to the description of Nature are new.

THE SIERPIŃSKI GASKET AS MONSTER

Sierpiński gasket is the term I propose to denote the shape in Plate 141. An extension to space is shown in Plate 143. The constructions are described in the captions.

Hahn 1956 comments that "A point on a curve is called a branch point if the boundary of any arbitrarily small neighborhood has more than two points in common with the curve... Intuition seems to indicate that it is impossible for a curve to be made up of nothing but... branch points. This intuitive conviction had been refuted [by the] Sierpiński... curve, *all of whose points are branch points.*"

THE EIFFEL TOWER: STRONG AND AIRY

Again, Hahn's view is totally without merit, and his uncharacteristic "seems to indicate" is a wise choice of words. My first counterargument is borrowed from engineering. (As argued before we tackled computers at the end of Chapter 12, there is nothing illogical about including articulated engineering systems in this work concerned with Nature.)

My claim is that (well before Koch, Peano, and Sierpiński) the tower that Gustave Eiffel built in Paris deliberately incorporates the idea of a fractal curve full of branch points.

In a first approximation, the Eiffel Tower is made of four A-shaped structures. Legend has it that Eiffel chose A to express *Amour* for his work. All four A's share the same apex

and any two neighbors share an ascender. Also, a straight tower stands on top.

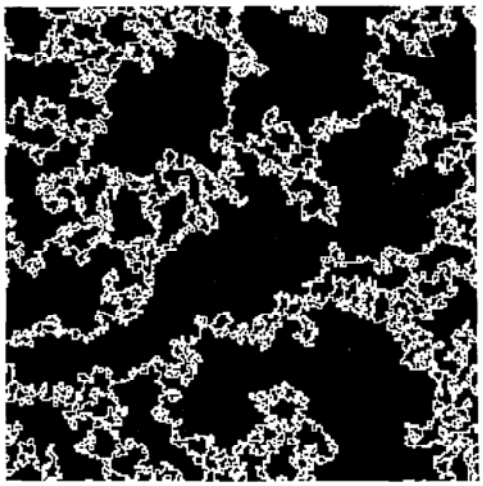
However, the A's and the tower are not made up of solid beams, but of colossal trusses. A truss is a rigid assemblage of interconnected submembers, which one cannot deform without deforming at least one submember. Trusses can be made enormously lighter than cylindrical beams of identical strength. And Eiffel knew that trusses whose "members" are themselves subtrusses are even lighter.

The fact that the key to strength lies in branch points, popularized by Buckminster Fuller, was already known to the sophisticated designers of Gothic cathedrals. The farther we go in applying this principle, the closer we get to a Sierpiński ideal! An infinite extrapolation of the Eiffel Tower design is described in Dyson 1966, p. 646, wherein a former student of Besicovitch seeks strong interplanetary structures of low weight.

CRITICAL PERCOLATION CLUSTERS

Let us now return to nature, or more precisely to an image of nature provided by statistical physics. I think the kin of the Sierpiński gasket is *demanded* by the study of percolation through lattices. Chapter 13, which began our case study of this topic, claims that percolation clusters are fractals. Now I add the further claim that the Sierpiński gasket's branching structure is a promising model of the structure of cluster backbones.

The physicists will mostly judge this model on the fact that it rapidly fulfilled its promise: Gefen, Aharony, Mandelbrot & Kirkpatrick 1981 shows the model allows usual calculations to be carried out *exactly*. But the details are much too technical to be included in this Essay, and the original reasons for my claim remain of interest. It arose from a resemblance I perceived between the gasket and the cluster backbones, as shown in this diagram:



The most conspicuous feature resides in the tremas left vacant by the elimination of dangling bonds (when a cluster was reduced to its backbone), and of clusters contained entirely within the cluster of interest. Second, the fact that the branching is self-similar in a Sierpiński gasket is shown in Chapter 13 to be an eminently desirable property in a geometric model of the percolation cluster. Finally, the dimensions fit to a degree that can hardly be coincidental! S. Kirkpatrick estimates that

in the plane D~1.6, astonishingly close to the D of the Sierpiński gasket! And in space, D~2.00, astonishingly close to the D of the fractal skewed web in Plate 143. Furthermore, Gefen, Aharony, Mandelbrot & Kirkpatrick 1981 observes that the identity between the D of the backbone and that of the generalized gasket persists in $\mathbb{R}^4$. An additional argument in favor of the gasket model is mentioned later, as a last application of ramification.

THE TRIADIC SIERPIŃSKI CARPET

Let us now switch from triangular to orthogonal lattices. They allow great versatility in design, yielding *curves* in the plane or in space, or *surfaces* in space. And the curves they yield, despite a superficial resemblance to the Sierpiński gasket, are very different from the fundamental viewpoint of ramification, to which we turn after defining them.

The literal planar extension of Cantor's method of deleting mid-thirds initiates with a square, and iś described in the caption on page 142. The fractal obtained by continuing ad infinitum is widely known by the homely term *triadic Sierpiński carpet*. Its dimension is $D=\log 8/\log 3=1.8927$.

NONTRIADIC FRACTAL CARPETS

Given an integer $b>3$, and writing $r=1/b$ as usual, a "large centered medallion" carpet is obtained by taking as initiator a square, as trema a square of side $1-2r$, with the same center, and as generator a thin ring of $4(b-1)$ squares of side r. The dimensions are $D=\log[4(b-1)]/\log b$. Given an odd integer $b>3$, a "small centered medallion" carpet is obtained by taking as trema a single subsquare of side r, with the same center as the initiator, and as generator a thick ring of b^3-1 small squares. The dimensions are $D=\log(b^3-1)/\log b$. Thus, any D between 1 and 2 can be approximated arbitrarily closely in a centered carpet.

Noncentered carpets can be defined for $b\geq 2$. For example, when $b=2$ and $N=3$, a trema made of one subsquare can be positioned in the subsquare on the top right. The corresponding limit set turns out to be the Sierpiński *gasket* built with the triangle forming the bottom left half of the square.

TRIADIC FRACTAL FOAM

The literal spatial extension of the triadic carpet consists in removing a cube's mid 27-th subcube as trema, leaving a shell of 26 subcubes. The resulting fractal is to be called *triadic fractal foam*. Its dimension is $D=\log 26/\log 3=2.9656$.

Here, every trema is entirely enclosed by an uninterrupted boundary split into infinitely many, infinitely thin layers of infinite density. In order to join two points situated in different tremas, it is necessary to cross an infinite number of layers. One is reminded, but this is a topic I do not master thoroughly enough to

attempt to account for it here, of the "space-time foam" which characterizes the finest structure of matter according to J. A. Wheeler and G. W. Hawking.

MENGER'S TRIADIC FRACTAL SPONGE

Karl Menger selects a different trema, shaped like a cross with spikes front and back, consisting of $N=20$ subcubes of side $\frac{1}{3}$, connected to one another. Among them, 12 form "rods" or ropes, and the remaining 8 are knots, connectors, or ties. The limit (Plate 145) satisfies $D=\log 20/\log 3=2.7268$. I call it *a sponge*, because both the curd and the whey are connected sets. One can conceive of water flowing between any two points in the whey.

To obtain a mixture of ropes and sheets, let the trema be a triadic cross continued by a single spike in front. By changing the direction of the spike every so often, one may end up with punctured sheets. It may be worth mentioning that I thought of all these shapes before reading Menger, while looking for models of turbulent intermittency.

NONTRIADIC SPONGES AND FOAMS

Given a nontriadic base $b>3$, generalized Menger sponges are obtained when the trema is the union of three square based cylinders: the axis of each coincides with an axis of the unit cube, its length is 1, and its base has sides parallel to the other axes. The sponge is called "light" when the bases' sides are as large as possible. For $E=3$, they are of length $1-2/b$, leaving as generator a collection of $12b-16$ cubes of side $r=1/b$. Hence the dimension is $D=\log(12b-16)/\log b$. Similarly, a "heavy sponge" is obtained, but only in case b is odd, when the cylinder bases' sides are of length $1/b$. For $E=3$, they leave as generator a collection of b^3-3b+2 cubes of side $r=1/b$. Now $D=\log(b^3-3b+2)/\log b$.

Fractal foams generalize in analogous fashion. For $E=3$, "thick wall" foams yield $D=\log(b^3-1)/\log b$, and "thin wall" foams yield $D=\log(6b^2-12b+8)/\log b$. With big holes and D near 2, the foam resembles an overly airy Emmenthaler. With small holes and D near 3, it resembles a different cheese delicacy, Appenzeller.

GAPS' SIZE DISTRIBUTIONS

The sponges' tremas merge together but carpets' and foams' tremas remain as gaps analogous to those of the Cantor dust (Chapter 8). The distribution of their linear scale Λ satisfies

$$N_r(\Lambda>\lambda)\propto F\lambda^{-D},$$

where F is a constant. We know this rule well from the gaps of a Cantor dust, and the islands and clusters of Chapter 13.

THE NOTION OF FRACTAL NET, LATTICE

The *lattices* of standard geometry are formed by parallel lines bounding equal squares or triangles, and analogous regular designs. The same term seems applicable to regular fractals in which any two points can be linked by at least two paths that do not otherwise overlap. When the graph is *not* regular, for example is random, I replace *lattice* by *net*.

However, a closer comparison of standard and fractal lattices reveals considerable differences. The first difference is that the standard lattices are invariant by translation but not by scaling, while for the fractal lattices the contrary is true. A second difference is that any standard lattice, if downsized, converges to the whole plane. Also, several standard lattices in the plane can be interpolated by adding lines halfway between existing parallel lines, and repeating ad infinitum. Again, the result converges towards the whole plane. Similarly, when a standard spatial lattice can be interpolated, its limit is the whole space. Thus, the limit is *not* a lattice. In the fractal context, to the contrary, the limit of an approximate fractal lattice is a fractal lattice.

The term, *ramified fractal lattices* can also be applied to the fractal foams.

THE SECTIONS' FRACTAL DIMENSIONS

A BASIC RULE. In many studies of fractals, it is important to know the dimensions of the linear and planar sections. The basic fact (used in Chapter 10 to show that $D>2$ for turbulence) concerns the section of a planar fractal shape by an interval "independent of the fractal." One finds that if the section is nonempty, it is "almost sure" that its dimension is $D-1$.

The corresponding value in space is $D-2$.

EXCEPTIONS. Unfortunately, this result is hard to illustrate in the case of nonrandom fractals that have axes of symmetry. The intervals that impose themselves upon our consideration are parallel to these axes, hence atypical, and nearly *every* simple section by an interval belongs to the exceptional set wherein the general rule fails to apply.

For example, take the Sierpiński carpet, the triadic Menger sponge and the triadic foam. $D-1$, which is the almost sure dimension of sections by intervals, is, respectively

$$\log(8/3)/\log 3,$$
$$\log(20/9)/\log 3, \text{ and } \log(26/9)/\log 3,$$

On the other hand, let x be the abscissa of an interval parallel to the y-axis of the Sierpiński carpet. When x, written in counting base 3, ends up by an uninterrupted infinite string of 0's or 2's, the sections are themselves intervals, hence $D=1$, larger than expected. When x ends up by an uninterrupted infinite string of 1's, to the contrary, the sections are Cantor dusts, hence $D=\log 2/\log 3$ is too small. And when x terminates by a periodic pattern of period M, including pM times 1 and $(1-p)M$ times 0 or 2, the sections are of dimension $p(\log 2/\log 3)+(1-p)$. The expected D prevails for $p\sim.29$. ◅ The same

holds if the digits of x are random. ► Thus, three dimensions are involved here: the largest, the smallest and the average.

Closely analogous results apply in space.

As to the Sierpiński gasket, the almost sure D is $\log(3/2)/\log 2$, but the D's relative to "natural" cuts range from 1 to 0. For example, a short interval through the midpoint of one of the gasket's sides, if close enough to the perpendicular, intersects the gasket on a *single* point, with $D=0$.

In part, the variability of these special sections is traceable to the regularity of the original shapes. But in another part, it is inevitable: the most economical section (not necessarily by a straight line) is the basis of the notions of topological dimension and of order of ramification, to which we proceed now.

THE RAMIFIED FRACTALS VIEWED AS CURVES OR SURFACES

As often stated, *curve* is used in this Essay as a synonym of "connected shape of topological dimension $D_T=1$." Actually, this phrase is not fully satisfactory to the mathematicians, and the precise restatements are delicate. Luckily, Chapter 6 could be content with a simple reason why any Koch curve with [0,1] as initiator deserves to be called a curve: like [0,1] itself, it is connected, but becomes disconnected if any point other than 0 or 1 is removed. And a snowflake boundary is like a circle: it is connected, but becomes disconnected if any two points are removed.

Restated more pedantically, as is now necessary, the topological dimension is defined recursively. For the empty set, $D_T=-1$. For any other set $\mathcal{S}$, the value of D_T is 1 higher than the smallest D_T relative to a "cutset" that disconnects $\mathcal{S}$. Finite sets and Cantor dusts satisfy $D_T = 1-1 = 0$, because nothing (the empty set) need be removed to disconnect them. And the following connected sets are all disconnected by the removal of a cutset that satisfies $D_T=0$: circle, [0,1], snowflake boundary, Sierpiński gasket, Sierpiński carpets, Menger sponges. (In the last three cases, it suffices to avoid the special intersections that include intervals.) Hence, all these sets are of dimension $D_T=1$.

By the same token, a fractal foam is a surface, with $D_T=2$.

Here is an alternative proof that $D_T=1$ for the gasket, all carpets, and all sponges with $D<2$. Since D_T is an integer $\leq D$, the fact that $D<2$ means that D_T is either 0 or 1. But the sets in question are connected, hence D_T is *no less* than 1. The only solution is $D_T=1$.

A CURVE'S ORDER OF RAMIFICATION

Topological dimension, and the corresponding notions of dust, curve, and surface, yield only a first level classification. Indeed, two finite sets containing M' and M'' points, respectively, have the same $D_T=0$, but they differ topologically. And Cantor dust differs from *all* finite dusts.

Let us now see how a parallel distinction

based on the number of points in a set ◅ its "cardinality" ► carries on to curves, leading to the topological notion of *order of ramification*, defined by Paul Urysohn and Karl Menger in the early 1920's. This notion is mentioned in few mathematics books other than the pioneers', but is becoming indispensable in physics, hence becoming better known after being tamed than in the wild. It shows that the reasons for discussing first a gasket, then a carpet, go beyond esthetics and the search for completeness.

The order of ramification involves the cutset containing the *smallest* number of points, that must be removed in order to disconnect the set $\mathcal{S}$. And it involves separately the neighborhood of every point P in $\mathcal{S}$.

THE CIRCLE. As background from standard geometry, begin by taking for $\mathcal{S}$ a circle of radius 1. A circle $\mathcal{B}$ centered on P cuts $\mathcal{S}$ in R=2 points, except if $\mathcal{B}$ has a radius exceeding 2, in which case R=0. The disc bounded by $\mathcal{B}$ is called a neighborhood of P. Thus, any point P lies in arbitrarily small neighborhoods whose boundaries intersect $\mathcal{S}$ at R=2 points. This is the best one can do: when $\mathcal{B}$ is the boundary of a general neighborhood of P, not necessarily circular but "not too large," R is at least 2. The terms "not too large" in the preceding sentence are a complication, but are unfortunately unavoidable. R=2 is called the order of ramification of the circle. We note that it is the same at all points of the circle.

THE GASKET. Next, let $\mathcal{S}$ be a Sierpiński gasket, constructed via tremas. Here R is no longer the same for every P. Let me show after Sierpiński that, excluding the initiator's vertices, R can be either $3=R_{min}$ or $4=R_{max}$.

The value R=4 applies to the vertices of any finite approximation of $\mathcal{S}$ by triangles. A vertex in an approximation of order $h \geq k$ is the common vertex P of two triangles of side 2^{-k}. Again, circles of center P and radius 2^{-k}, with $h>k$, intersect $\mathcal{S}$ in 4 points, and bound arbitrarily small neighborhoods of P. And if $\mathcal{B}$ bounds a "sufficiently small" neighborhood of P (in the new sense that the initiator's vertices lie outside $\mathcal{B}$), one can show that $\mathcal{B}$ intersects $\mathcal{S}$ in at least 4 points.

The value R=3 applies for every point of $\mathcal{S}$ that is the limit of an infinite sequence of triangles, each contained in its predecessor and having vertices distinct from its predecessor's. Circles circumscribed to these triangles intersect $\mathcal{S}$ in 3 points, and bound arbitrarily small neighborhoods of P. Also if $\mathcal{B}$ bounds a sufficiently small neighborhood of P (again, the initiator's vertices must lie outside), one can show that $\mathcal{B}$ intersects $\mathcal{S}$ at 3 points at least.

THE CARPETS. When $\mathcal{S}$ is a Sierpiński carpet, the result is *radically* different. Any neighborhood's boundary, if sufficiently small, intersects $\mathcal{S}$ in a nondenumerably infinite cutset, regardless of the parameters N, r, or D.

COMMENT. In this finite versus infinite dichotomy, the gasket does not differ from the standard curves, while the carpets do not differ from the whole plane.

HOMOGENEITY. UNICITY. Denoting by R_{min} and R_{max} the smallest and the largest R attained on a point of $\mathcal{S}$, Urysohn proves that

$R_{max} \geq 2R_{min}-2$. The ramification is called *homogeneous* when the equality $R_{max}=R_{min}$ holds; this is the case when $R \equiv 2$, as in simple closed curves, and when $R \equiv \infty$.

For other lattices with $R_{max} = 2R_{min}-2$, I propose the term *quasi-homogeneous*. One simple and famous example, the Sierpiński gasket, is self-similar. The other nonrandom examples are part of a collection set up by Urysohn 1927, and are *not* self-similar. Thus, the conditions, of being quasi-homogeneous and self-similar, have only one known solution, the Sierpiński gasket. Could this seeming unicity be confirmed rigorously?

STANDARD LATTICES. Here the order of ramification ranges from a minimum of 2 for all points off the lattice sites, to a variable finite maximum attained on the lattice sites: 4 (squares), 6 (triangles or cubes) or 3 (hexagons). However, as a standard lattice of any kind is downsized, it transforms from a curve into a plane domain, and its ramification becomes $R=\infty$.

This last fact is made more obvious by exchanging the infinitely small and the infinitely large, holding to a lattice of fixed cell size, and observing that in order to isolate an increasingly *large* portion of lattice, one must cut points whose number has no finite bound.

FORMAL DEFINITION. ◅ See Menger 1932 and p. 442 of Blumenthal & Menger 1970. ►

APPLICATIONS OF RAMIFICATION

Let us now face a familiar question. Whatever interest the Sierpiński and Menger shapes, and their kin, may have for the mathematician, is it not obvious that the order of ramification can be of no interest to the student of Nature? The response is as familiar—to us!—as the question. The order of ramification is already meaningful in the "real world" of the finite approximations which obtain when the interpolation leading to a fractal is stopped at some positive inner cutoff, ϵ.

Indeed, given an approximate Sierpiński gasket made of filled triangles of side ϵ, a domain whose linear scale is above ϵ can be disconnected by removing 3 or 4 points, each of which belongs to 2 neighboring gaps' boundaries. This number (3 or 4) does not change as this approximation is refined. Hence, from the viewpoint of ramification, all approximate gaskets are curve-like.

To the contrary, all carpets have the property that the boundaries of any two gaps fail to overlap. To disconnect a finite approximation of such a shape, in which the gaps of diameter $<\epsilon$ are disregarded, it is necessary to remove whole intervals. And these intervals' number increases as $\epsilon \to 0$. Whyburn 1958 shows that all the fractal curves that possess this property are topologically identical ◅ homeomorphic ►, and are characterized by the fact they contain no part that can be disconnected by the removal of a single point.

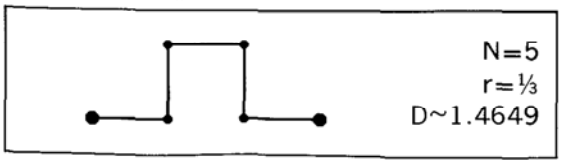
Due to the preceding comments, it is not surprising that the finiteness of ramification acquires clearcut implications when fractal geometry is called to determine in detail how much a plane fractal curve partakes of its two

standard limits: the straight line and the whole plane. In general, knowing the fractal dimension does *not* suffice. For example, Gefen, Mandelbrot & Aharony 1980 examines critical phenomena for Ising models on a fractal lattice, and finds that the most important issue ◅ whether the critical temperature is 0 or positive ► depends on the finiteness of R.

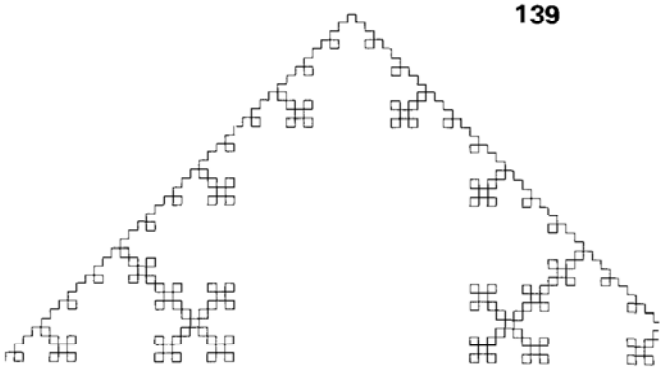
We are now in a position to give an explanation we had postponed. The reason why a cluster backbone in critical Bernoulli percolation seems better modeled by a gasket than by a carpet lies in this finding reported in Kirkpatrick 197?. Even on extremely large lattices, a critical backbone can be cut by removing an essentially unvarying small number of bonds, of the order of 2. Even allowing for certain biases I could think of, this points out very strongly toward $R<\infty$.

ALTERNATIVE FORM OF RAMIFICATION

Two variants of the Koch snowflake achieve ramification through branches without loops. The first is a plane curve obtained when the initiator is a square and the generator is



The resulting shape is totally different from the snowflake, as shown overleaf.



The next example is a surface of zero volume, infinite area, and a dimension equal to $\log 6/\log 2=2.58497$. The initiator is a regular tetrahedron. On the mid-quarter of each face (= the triangle having as vertices the sides' midpoints), one attaches a tetrahedron reduced in the ratio ½. One repeats the procedure with each face of the resulting regular (skew and nonconvex) 24-hedron, and so on ad infinitum. From the second stage on, the added tetrahedrons self-contact along lines, without self-intersecting. And eventually they swarm all over the initiator. Let each fourth of this shape, growing on a face of the initiator, be called a Koch pyramid.

SECRETS OF THE KOCH PYRAMID

A Koch pyramid is a wondrous shape—plain when seen from above, but with a wealth of hidden chambers to defy the imagination.

Seen from above, it is a tetrahedron whose base is a equilateral triangle, but whose three other faces are right isosceles triangles joined at their 90° vertices. Three Koch pyramids, if put together on the sides of a regular tetrahedron, add to a plain cubic box.

Now lift such a pyramid from the floor of the desert. From a distance, we see its base

subdivides into four equal regular triangles. But in place of the middle triangle there is a hole opening up on a "chamber of order 1," shaped like a regular tetrahedron whose fourth vertex coincides with the pyramid's top vertex. Next, as we approach and perceive finer detail, we find that the regular triangles that form the peripheral fourths of the base and the top faces of the chamber of order 1 are not smooth either. Each is broken by a tetrahedral chamber of order 2. Similarly, as we explore the chambers of order 2, each of their triangular walls reveals a chamber of order 3 in its middle portion. And increasingly tiny chambers appear without end.

All the chambers together add up precisely to the Koch pyramid's volume. On the other hand, if the chambers are viewed as including their bases but not their three other faces, they do *not* overlap. Were our pyramid to be dug from a mound, the chamber diggers would have to scoop out all its volume, leaving a mere shell. The curve along which this surface rests on the base's plane, and the chamber "walls," are Sierpiński gaskets.

SPHERICAL TREMAS AND LATTICES

Lieb & Lebowitz 1972 makes an unwitting contribution to fractal geometry, by packing $\mathbb{R}^E$ with balls whose radii are of the form $\rho_k=\rho_0 r^k$, with $r<1$; the per-unit-volume number of balls of radius ρ_k is of the form $n_k=n_0\nu^k$, where ν is an integer and is of form $\nu=(1-r)r^{-E}$, which strongly restricts r. Thus, the exponent of the distribution of gap sizes is

$$D=\log \nu/\log (1/r)=E-\log (1-r)/\log r.$$

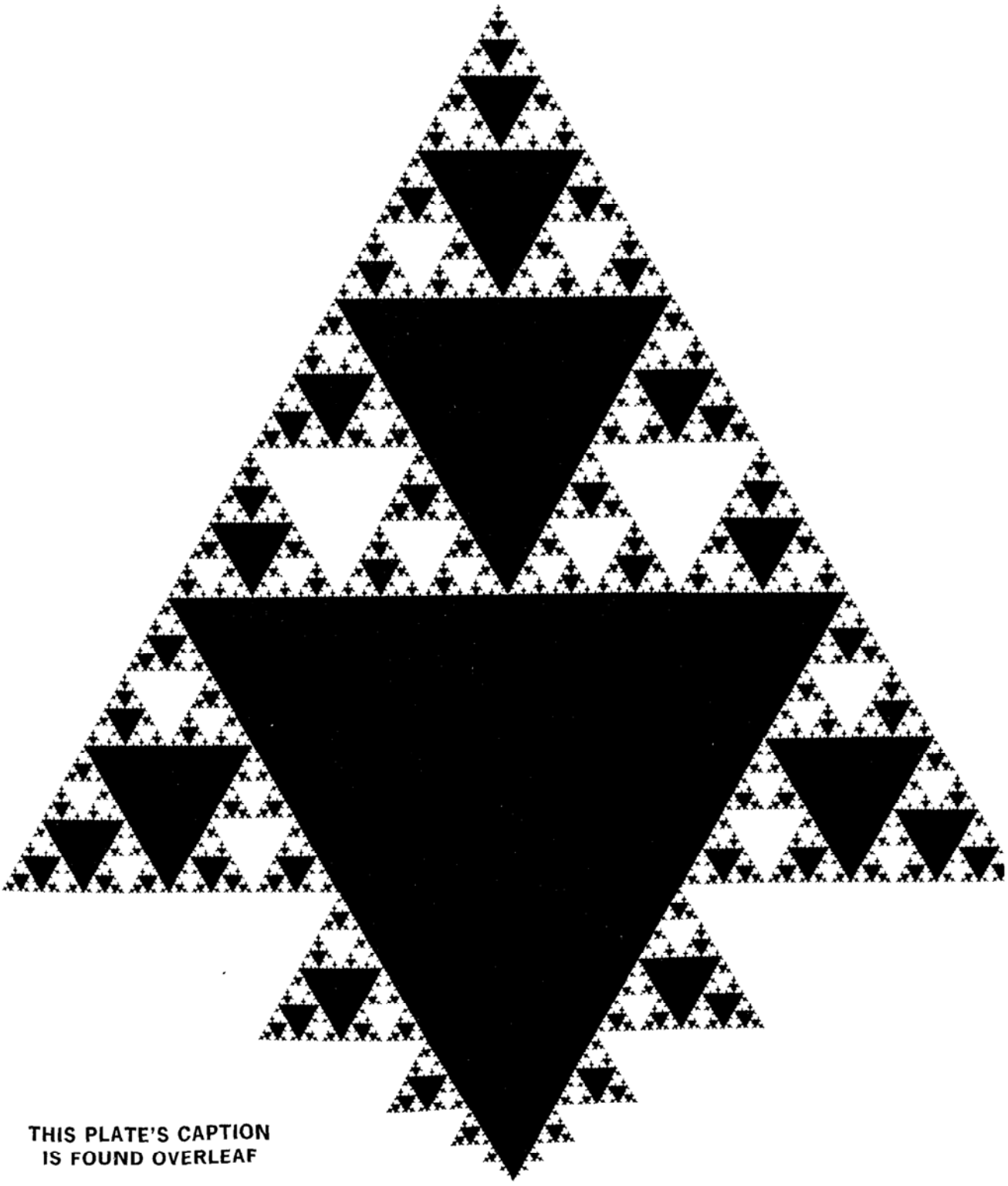
First, one centers big spheres of radius ρ_1 on a lattice of side $2\rho_1$. The vertices of a lattice of side $2\rho_2$ that lie outside of the big spheres are numerous enough to serve as centers for the next smaller spheres, and so on. The construction involves these upper bounds on r:

$$\text{for } E=1,\ r\le 1/3; \quad \text{for } E=2,\ r\le 1/10;$$
$$\text{for } E=3,\ r\le 1/27; \quad \text{as } E\to\infty,\ r\to 0.$$

Packing of $\mathbb{R}^3$ by nonoverlapping balls can proceed more rapidly. For example, on the line, the maximum r is ⅓, corresponding to the triadic dust of Cantor! The existence of Cantor dusts with $r>⅓$ demonstrates that one-dimensional packing can leave a remainder of arbitrarily low dimension. However, a tighter packing involves richer structure.

PREVIEW OF LACUNARITY

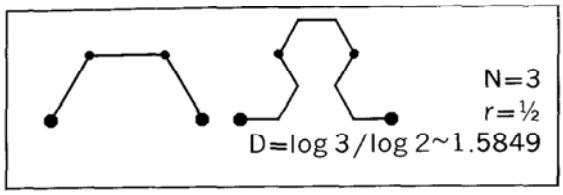
Even after the order of ramification R is added to the dimensions D_T and D, a fractal remains incompletely specified for many purposes. Of special importance is the additional notion of lacunarity that I developed. A very lacunar fractal's gaps are very large, and conversely. The basic definitions could have been described here, but it is more expedient to wait until Chapter 34. ■



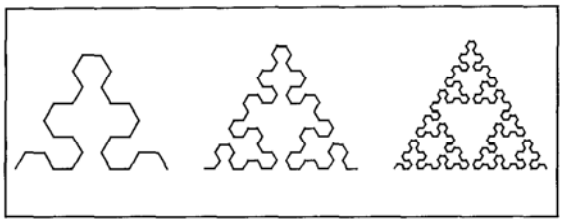
THIS PLATE'S CAPTION
IS FOUND OVERLEAF

Plate 141, OVERLEAF ¤ SIERPIŃSKI ARROWHEAD (BOUNDARY DIMENSION D~1.5849)

In Sierpiński 1915, the initiator is [0,1], and the generator and second teragon are



This construction's next two stages are

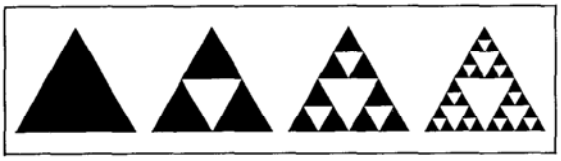


And an advanced stage is shown as the "coastline" of the upper portion of Plate 141 (above the largest solid black triangle).

SELF-CONTACTS. Finite construction stages are free of points of self-contact, as in Chapter 6, but the limit curve *does* self-contact infinitely often.

TILING ARROWHEADS. The arrowhead in Plate 141 (turned sideways, it becomes a tropical fish) is defined as a piece of the Sierpiński curve contained between two successive returns to a point of self-contact, namely the midpoint of [0,1]. Arrowheads tile the plane, with neighboring tiles being linked together by a nightmarish extrapolation of Velcro. (To mix metaphors, one fish's fins fit exactly those of two other fish). Furthermore, by fusing together four appropriately chosen neighboring tiles, one gets a tile increased in the ratio of 2.

THE SIERPIŃSKI GASKET'S TREMAS. I call Sierpiński's curve a *gasket*, because of an alternative construction that relies upon cutting out "tremas," a method used extensively in Chapters 8 and 31 to 35. The Sierpiński gasket is obtained if the initiator, the generator, and next two stages are these closed sets:



This trema generator includes the above stick generator as a proper subset.

WATERSHED. I first encountered the arrowhead curve without being aware of Sierpiński, while studying a certain watershed in Mandelbrot 1975m. ■

Plate 143 ¤ A FRACTAL SKEWED WEB (DIMENSION D=2)

This web obtains recursively, with N=4 and r=½, using a closed tetrahedron as initiator and a collection of tetrahedrons as generator.

Its dimension is D=2. Let us project it along a direction joining the midpoints of either couple of opposite sides. The initiator tetrahedron projects on a square, to be called initial. Each second-generation tetrahedron projects on a subsquare, namely (¼)th of the initial square, etc. Thus, the web projects on the initial square. The subsquares' boundaries overlap. ■

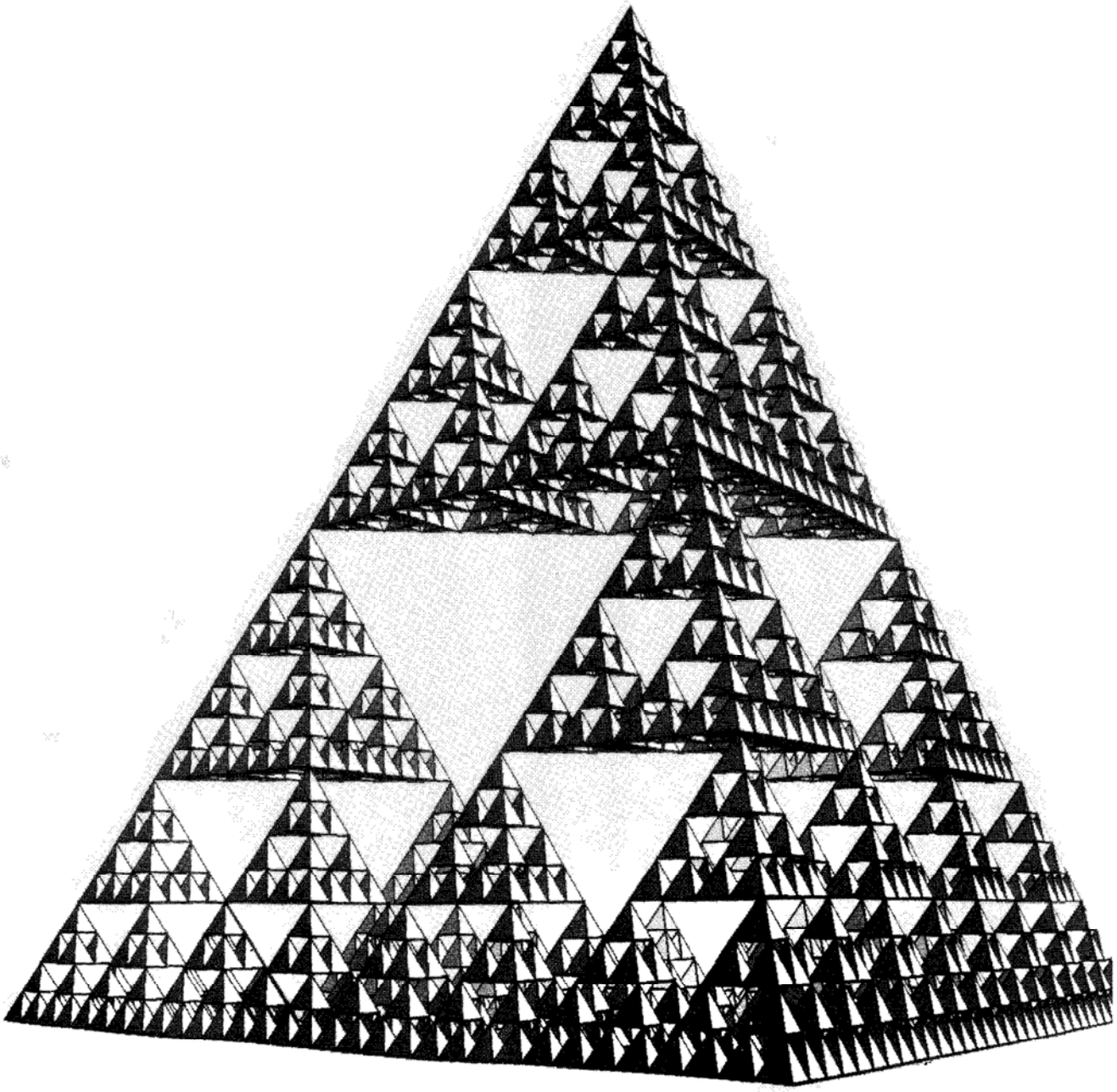
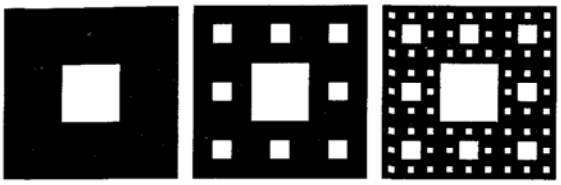


Plate 145 ¤ THE SIERPIŃSKI CARPET (DIMENSION D~1.8928), AND THE MENGER SPONGE (DIMENSION D~2.7268)

SIERPIŃSKI CARPET. In Sierpiński 1916, the initiator is a filled square, while the generator and the next two steps are



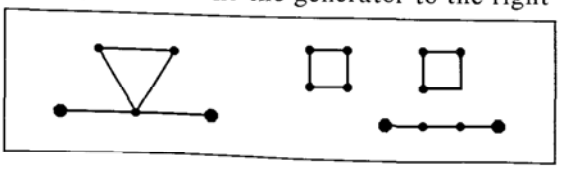
N=8, r=⅓, D~1.8928.

This carpet's area vanishes, while the total perimeter of its holes is infinite.

PLATE 145. THE MENGER SPONGE. The principle of the construction is evident. Continued without end, it leaves a remainder to be called a Menger sponge. I regret having credited it wrongfully in earlier Essays, to Sierpiński. (Reproduced from *Studies in Geometry*, by Leonard M. Blumenthal and Karl Menger, by permission of the publishers, W. H. Freeman and Company, copyright 1970.) The intersections of the sponge with medians or diagonals of the initial cube are triadic Cantor sets.

FUSED ISLANDS. The carpet, as well as the gasket in Plate 143, may also be obtained by yet another generalization of the Koch recursion, wherein self-overlap is allowed, but overlapping portions count only once.

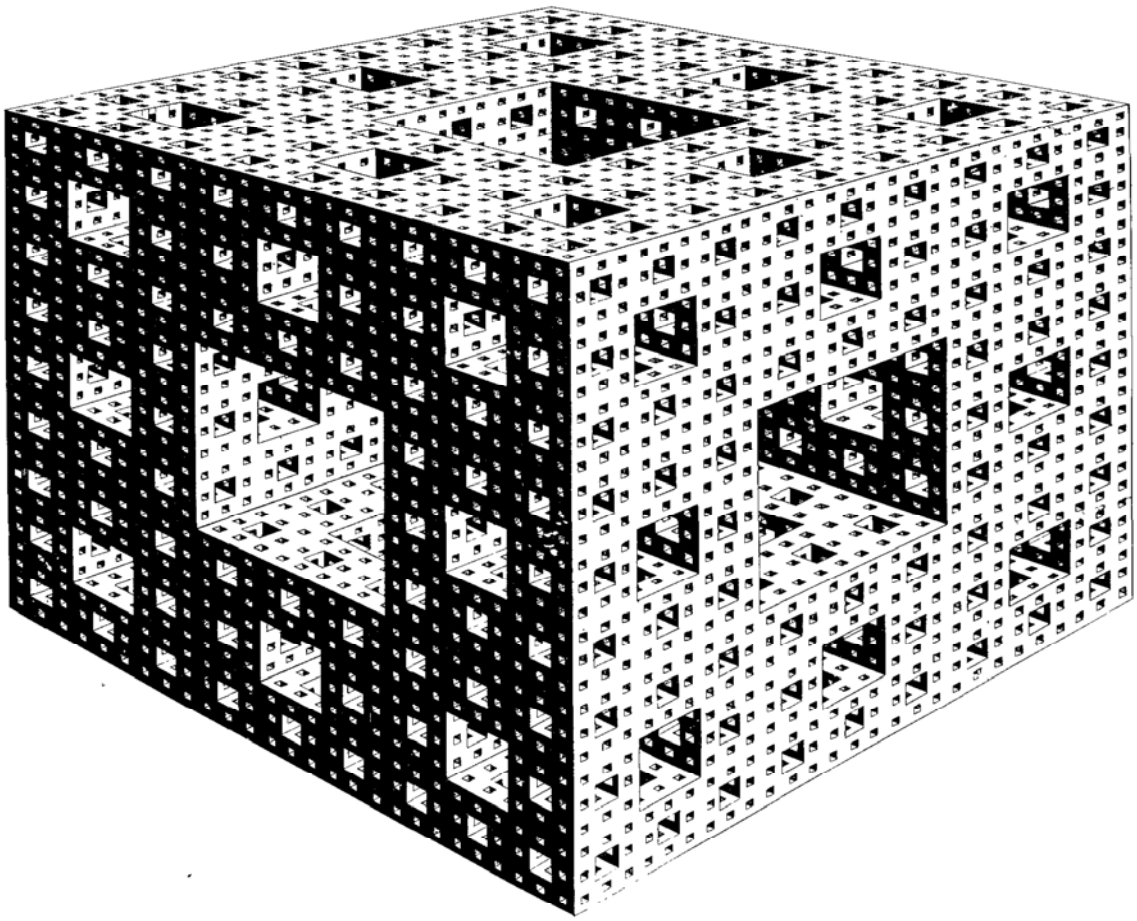
To obtain a gasket, the initiator is a regular triangle, and we take the generator to the left. To obtain a carpet, the initiator is a square, and we take the generator to the right



Two phenomena familiar from Chapter 13 are encountered again: each island's coastline is rectifiable and therefore of dimension 1, and the dimension of the gasket or the carpet expresses the degree of fragmentation of land into islands rather than the degree of irregularity of the islands' coastlines.

Otherwise, the result is unfamiliar: in Chapter 13 the sea is connected, which seems to be a proper topological interpretation of nautical openness. It is also open in the set topological sense of not including its boundary. The novelty brought in by the present construction is that it is possible for the Koch islands to "fuse" *asymptotically* into a solid superisland; there is no continent, and the coastlines combine into a lattice.

◅ Topologically, every Sierpiński carpet is a plane universal curve, and the Menger sponge is a spatial universal curve. That is, see Blumenthal & Menger 1970, pp. 433 and 501, these shapes are respectively the most complicated curve in the plane, and the most complicated curve in any higher dimensional space. ► ■



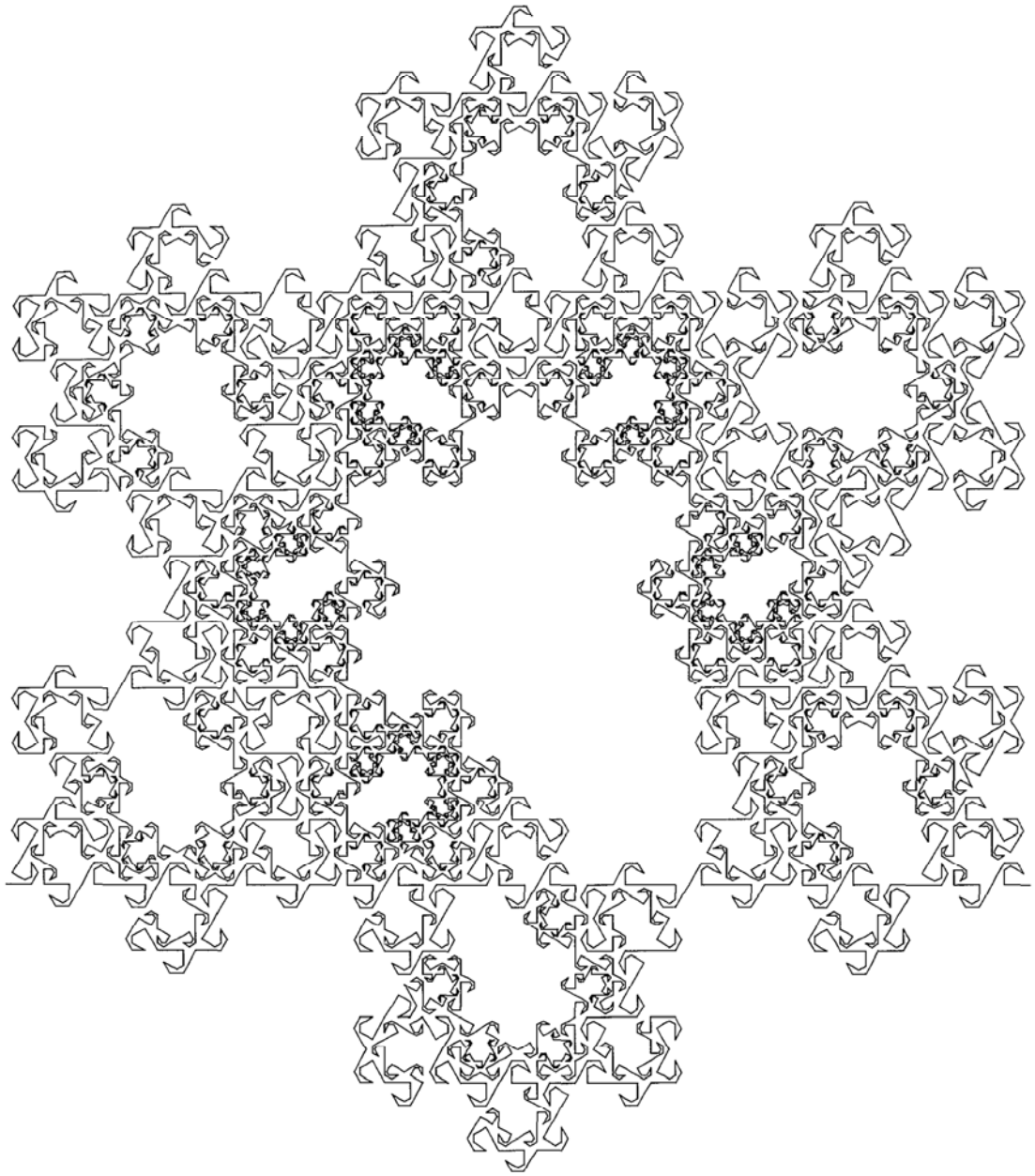


Plate 146 ¤ SPLIT SNOWFLAKE HALLS (DIMENSION D~1.8687)

Long ago and far away, the Great Ruler and his retinue had sat their power in the splendid Snowflake Halls. A schism occurs, a war follows, ending in stalemate, and finally Wise Elders draw a line to divide the Halls between the contenders from the North and the South.

RIDDLES OF THE MAZE. Who controls the Great Hall, and how is it reached from outside? Why do some Halls fail to be oriented toward either of the cardinal points? For hints, see the Monkeys Tree on Plate 31. ■

15 ¤ Surfaces with Positive Volume, and Flesh

The fractal curves, surfaces, and dusts which the present Part describes and tames for the purposes of science, are only scaling in an asymptotic or otherwise limited sense.

This first chapter centers on surfaces with a positive (nonvanishing!) volume. What a mad combination of contradictory features! Have we not finally come to mathematical monsters without conceivable utility to the natural philosopher? Again, the answer is emphatically to the negative. While believing they were fleeing Nature, two famous pure mathematicians unknowingly prepared the precise tool I need to grasp (among others) the geometry of...flesh.

CANTOR DUSTS OF POSITIVE MEASURE

A preliminary step is to review Cantor's construction of the triadic set $\mathcal{C}$. Its being of zero length (more pedantically, of zero linear measure) follows from the fact that the lengths of the mid third tremas add to

$$1/3+2/3^2... +2^k/3^{k+1}... = 1.$$

But the fact that $\mathcal{C}$ is totally disconnected, hence of topological dimension $D_T=0$, is independent of the trema lengths. It comes from the basic fact that each construction stage bisects every interval created in the preceding stage, by removing a trema centered on the "host" interval's midpoint. Denoting the ratio of the trema and host lengths by λ_k, the cumulative length of the intervals that remain after K stages is $\Pi_0^K(1-\lambda_k)$. It decreases as $K\rightarrow\infty$ to a limit denoted by P. In Cantor's original construction, $\lambda_k\equiv 2/3$, hence $P=0$. But $P>0$ whenever $\Sigma_0^\infty\lambda_k<\infty$. In that case, the remainder set $\mathcal{C}_*$ has the positive length $1-P$. This set is not self-similar, hence has no similarity dimension, but the Hausdorff Besicovitch definition, Chapter 5, concludes that

$D=1$. It follows from $D>D_T$ that $\mathcal{C}_*$ is a fractal set. Since D and D_T are both independent of the trema lengths λ_k, their values describe $\mathcal{C}_*$ very superficially.

The construction is even more perspicuous in the plane. Cut out from the unit square a cross of area λ_1, leaving four square tiles. Next cut out from each a cross of relative area λ_2. This cascade generates a dust, $D_T=0$, having the area $\Pi_0^{\infty}(1-\lambda_k)$. When this area does not vanish, $D=2$.

In E-dimensional space, one can similarly achieve *a dust with positive volume,* satisfying $D_T=0$ and $D=E$.

SLOWLY DRIFTING $\log N/\log(1/r)$

◅ Although the Cantor dusts with positive length, area or volume have no similarity dimension, it is useful to set $r_k=(1-\lambda_k)/2$, and to investigate the formal dimensions defined as $D_k=\log N/\log(1/r_k)$.

◅ When D_k drifts slowly, it embodies the idea of effective dimension discussed in Chapter 3 when describing a ball of thread. On the line, the dimension $D=1$ of the limit set $\mathcal{C}_*$ is the limit of $\log 2/\log(1/r_k)$. Furthermore, the conclusion $D=1$ does not require $\Sigma\lambda_k<\infty$, only the weaker condition $\lambda_k\to 0$. Consequently there are three classes of linear Cantor dusts: (a) $0<D<1$ and length$=0$, (b) $D=1$ and length$=0$, and (c) $D=1$ and length>0.

◅ The counterpart of the above category (c) can occur for Koch curves. It suffices to change the generator at each construction stage and to let its D tend to 2. For example, take $r_k=\frac{1}{2}k$ and adopt for N_k, hence for D_k, the maximal value discussed in the caption of Plate 53. The limit has a remarkable combination of properties: its fractal dimension $D=2$ is nonstandard for a curve; but its topological dimension is standard: it is $D_T=1$, and its area is standard: it vanishes.

◅ The same properties coexist in Brownian motion, Chapter 25, but here they are achieved while avoiding double points.

◅ The formal dimension may also drift *away* from $D=2$. For example, k stages of a plane filling tree construction may be finished off by stages with $D<2$. The result may be of use in modeling certain river trees that seem plane filling on scales above the inner cutoff η but crisscross finer scale domains less thoroughly. This η would be very big in deserts, and very small in soaked jungles, possibly equal to 0. Such rivers' effective dimension would be $D=2$ for scales above η, and $D<2$ for scales below η. ►

CURVES WITH POSITIVE AREA

Our intuition of dusts being imperfect, it is not bothered by dusts of positive length or volume. But *curves* of positive area are truly hard to swallow. Thus, after Lebesgue 1903 and Osgood 1903 showed that swallow them we must, they came to supersede the Peano curve as supreme monsters. After describing an example, I show that the thought is worse than the reality: in the most textual sense,