Hausdorff Dimension, Its Properties, and Its Surprises

Dierk Schleicher

1. INTRODUCTION. The concept of *dimension* has many aspects and meanings within mathematics, and there are a number of very different definitions of what the dimension of a set should be. The simplest case is that of $\mathbb{R}^d$: in order to distinguish points in $\mathbb{R}^d$, we need d different (real) coordinates, so $\mathbb{R}^d$ has dimension d as a (real) vector space. Similarly, a d-dimensional manifold is a space that locally looks like a piece of $\mathbb{R}^d$.

Another interesting concept is the *topological dimension* of a topological space: every discrete set has topological dimension 0 (e.g., any finite sets of points in $\mathbb{R}^d$), an injective curve has topological dimension 1, a disk has dimension 2 and so on. The idea is that a set of dimension d can be disconnected in a neighborhood of *every* point by a set of dimension $d-1$: curves and circles can be disconnected by removing isolated points, disks can be disconnected by removing curves and circles, etc. A formal definition is recursive, starting conveniently with the empty set: *the empty set has topological dimension* -1*, and a set has topological dimension at most* d *if each point has a basis of open neighborhoods whose boundaries have topological dimension at most* $d-1$.

All these dimensions, if finite, are integers (we will ignore infinite-dimensional spaces). An interesting discussion of various concepts of dimension, different in spirit from ours, can be found in the recent article of Manin [**17**].

2000 *Mathematics Subject Classification.* 28A78, 28A80, 30D05, 37-01, 37C45, 37F10, 37F20, 37F35.

Key words and phrases. Hausdorff dimension, dimension paradox, dynamic ray, Julia set, entire function, iteration, complex dynamics, fractal.

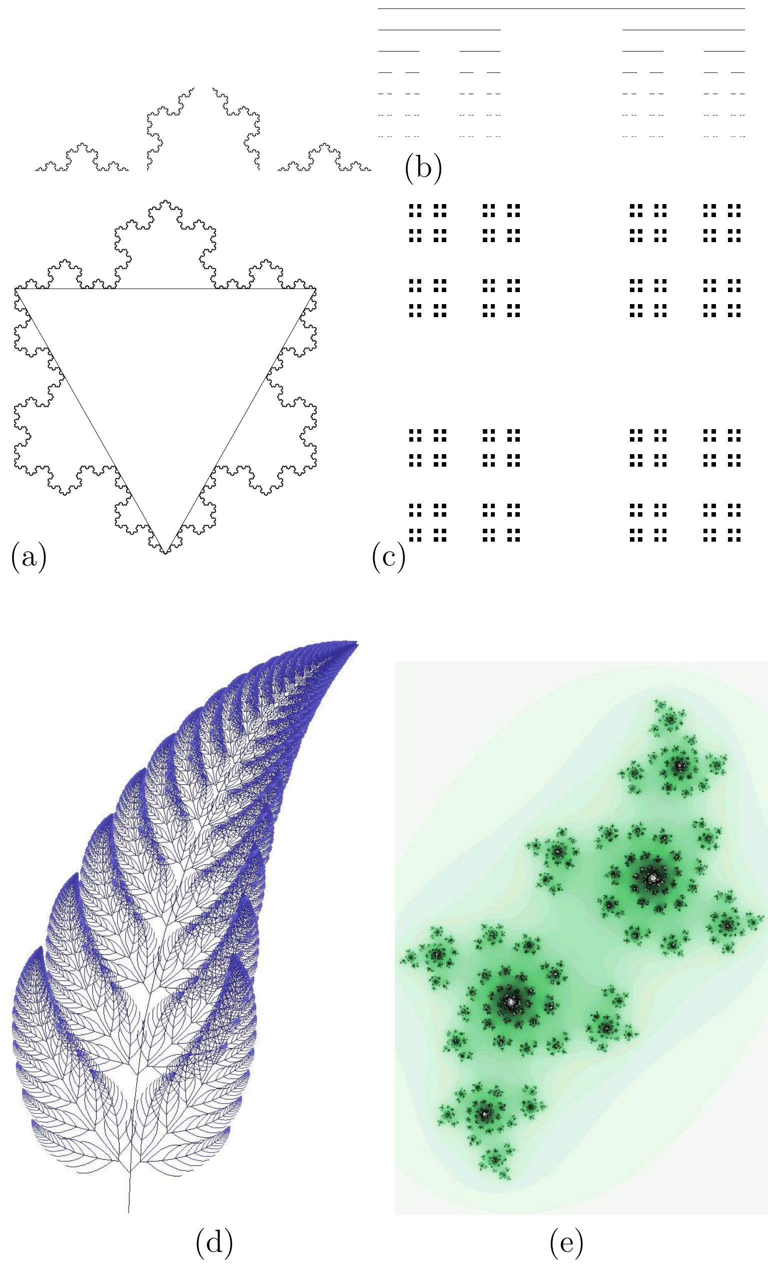

FIGURE 1. Several "fractal" subsets of $\mathbb{R}^2$: (a) the "snowflake" (von Koch) curve: each of its three (fractal) sides can be disassembled into four pieces, each of which is a copy of the entire side, shrunk by a factor 1/3; (b) the Cantor middle-third set, consisting of two copies of itself, shrunk by 1/3; (c) a "fractal" square in the plane, consisting of four shrunk copies of itself with a factor 1/3 (so it has the same dimension as the snowflake!); (d) a fern; (e) the Julia set of a quadratic polynomial.

We will be concerned with a different aspect of dimension, having to do with self-similarity of "fractal" sets such as those shown in Figure 1. As Mandelbrot points out [**16**, p. 1], "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," so many objects occurring in nature are not manifolds. For instance, the fern in Figure 1 is constructed by a simple affine self-similarity process, and people have tried to describe the hairy systems of roots of trees or plants in terms of "fractals", rather than as smooth manifolds. Similar remarks apply to the human lung or to the borders of most states and countries.

The concept of Hausdorff dimension is almost a century old, but it has received particularly prominent attention since the advent of computer graphics and the computer power to simulate and visualize beautiful objects with importance in a number of sciences. Earlier, such sets were often constructed by ad hoc methods as counterexamples to intuitive conjectures. In the first part of this paper, we try to convince interested readers that Hausdorff dimension is the "right" concept to describe interesting properties of a metric set X: for each number d in $\mathbb{R}_0^+$ we define the d-dimensional Hausdorff measure $\mu_d(X)$; if d is a positive integer and $X = \mathbb{R}^d$, then this measure coincides with Lebesgue measure (up to a normalization factor). There is a threshold value for d, called $\dim_H(X)$, such that $\mu_d(X) = 0$ if $d > \dim_H(X)$ and $\mu_d(X) = \infty$ if $d < \dim_H(X)$. This value $\dim_H(X)$ is the *Hausdorff dimension* of X.

We first help to develop intuition for this natural concept, and then we challenge it by describing a number of relatively newly discovered sets with very remarkable and surprising (possibly counterintuitive!) properties of Hausdorff dimension. To describe such sets, imagine a curve $\gamma\colon (0, \infty) \to \mathbb{C}$ that connects the point 0 to ∞ (we identify a curve $\gamma\colon I \to \mathbb{C}$ with its image set $\{\gamma(t)\colon t \in I\}$ in $\mathbb{C}$). Curves have dimension at least 1, possibly more, but the two endpoints certainly have dimension 0. Now take a collection of disjoint curves γ_h, each connecting a different point z_h to ∞. For example, let $z_h = ih$ for h in $[0, 1]$ and $\gamma_h(t) = ih + \gamma(t)$ (provided γ is such that all γ_h are disjoint). Then the endpoints are an interval with dimension 1, while the union of all curves γ_h covers an open set of $\mathbb{C}$ and should certainly have dimension 2. This is true and intuitive: the union of the endpoints has smaller dimension than the union of the curves. In this paper, we describe the following situation [**24**]:

Theorem 1 (A Hausdorff Dimension Paradox). *There are subsets E and R of $\mathbb{C}$ with the following properties:*

(1) *E and R are disjoint;*
(2) *each path component of R is an injective curve (a "ray") $\gamma\colon(0,\infty)\to\mathbb{C}$ connecting some point e of E to ∞ (i.e., $\lim_{t\to 0}\gamma(t)=e$ and $\lim_{t\to\infty}\gamma(t)=\infty$);*
(3) *each point e of E is the endpoint of one or several curves in R;*
(4) *the set $R=\bigcup\gamma((0,1))$ of rays has Hausdorff dimension 1;*
(5) *the set E of endpoints has Hausdorff dimension 2 and even full 2-dimensional Lebesgue measure (i.e., the set $\mathbb{C}\setminus E$ has measure zero);*
(6) *stronger yet, we have $E\cup R=\mathbb{C}$: the set of endpoints E is the complement of the 1-dimensional set R, yet each point in E is connected to ∞ by one or several curves in R!*

Mathematics is full of surprising phenomena, and often very artful methods are used to construct sets that exhibit these phenomena. This result is another illustration that many of these phenomena arise quite naturally in dynamical systems, especially complex dynamics. It comes at the end of a series of successively stronger results. The story started with a surprising result by Karpińska [**13**]: she established the existence of natural sets E and R arising in the dynamics of complex exponential maps $z\mapsto\lambda e^z$ for certain values of λ, where E and R enjoy properties (1)–(4), as well as (5) in the form that E has Hausdorff dimension 2. In [**25**], this result was extended to exponential maps with λ in $\mathbb{C}^*=\mathbb{C}\setminus\{0\}$ arbitrary. In [**23**], this was carried over to maps of the form $z\mapsto ae^z+be^{-z}$; in this case, E always has positive 2-dimensional Lebesgue measure. Finally, condition (6) was established for maps like $z\mapsto\pi\sinh z$ [**24**].

We start this paper with a discussion of several concepts of dimension (section 2). In section 3, we give the definition of Hausdorff dimension together with a number of its fundamental properties. In section 4, we describe a beautiful example constructed by Bogusława Karpińska in which E has positive 2-dimensional Lebesgue measure. In the remainder of the paper, we show that sets E and R satisfying all the assertions of Theorem 1, including $\mathbb{C}=E\dot\cup R$, appear naturally in complex dynamics, when iterating maps such as $z\mapsto\pi\sin z$.

The basic features of iterated complex sin and sinh maps are described in section 5, and a fundamental lemma for estimating Hausdorff dimension is given in section 6. In section 7, we then describe the dynamics of the map $z\mapsto\pi\sin z$ in detail and finish the proof of Theorem 1. Finally, we discuss some related known results about planar

Lebesgue measure, including a theorem of McMullen and a conjecture of Milnor.

The purpose of this paper is to highlight interesting phenomena that are observed at the interface between dimension theory and transcendental dynamics. It cannot serve as an exhaustive survey on the exciting work that has been done on these two areas, and we can mention only a few of the most interesting references. A good survey of transcendental dynamics is found in Bergweiler [**2**]; some more surprising properties of exponential dynamics are described in Devaney [**4**]. The topic of "curves of escaping points in transcendental dynamics" was first raised in 1926 by Fatou [**10**] and taken up more systematically by Eremenko [**8**]. In the special case of exponential dynamics, it was first investigated by Devaney and coauthors [**5**], [**6**] and completed in [**25**], [**11**]. In more general settings, there are existence results in [**7**], and the current state of the art can be found in the recent thesis of Rottenfußer [**22**]. Among current work on Hausdorff dimension in transcendental dynamics, we would like to mention the survey papers by Stallard [**27**] and by Kotus and Urbański [**14**]. We apologize to those whose work we have not mentioned here.

2. CONCEPTS OF "FRACTAL" DIMENSION. The fundamental idea that leads to "fractal" dimensions is to investigate interesting sets at different scales of size. Consider a regular three-dimensional cube, say of side-length 1. We can subdivide this cube into many small cubes of side-length $s = 1/k$ for any positive integer k. Obviously, the number of little cubes we obtain is $N(s) = k^3 = s^{-3}$. However, if we subdivide a unit square into small squares of side-length $1/k$, we obtain $N(s) = s^{-2}$ little squares. The exponent here is the dimension: *if a set X in $\mathbb{R}^n$ can be subdivided into some finite number $N(s)$ of subsets, all congruent (by translations or rotations) to one another and each a rescaled copy of X by a linear factor s, then the "self-similarity dimension" of X is the unique value d that satisfies $N(s) = s^{-d}$, i.e.,*

$$d = \log(N(s))/\log(1/s) \ .$$

This simple idea can be applied to a number of interesting sets. Consider, for example, the "snowflake" curve of Figure 1a: we only look at the top third of the snowflake, above the triangle that we have inscribed for easier description. The detail above the snowflake shows that this top third can be disassembled into $N = 4$ pieces, each of which is a rescaled version of the entire top third with a rescaling factor $s = 1/3$. The associated dimension must satisfy $3^d = 4$ (i.e., $d = \log 4/\log 3 \approx 1.26\ldots$). The snowflake is a curve (and thus has topological dimension 1), but its self-similarity dimension is greater than that of a straight line: when subdividing a straight line into pieces of one-third the original size, we obtain three pieces; for the snowflake, we get four (and for a square we get nine). Continued refinement has the same dimension: we can break up the four pieces into four pieces each, so that all are rescaled by a factor $s = 1/9$; and again $d = \log(4^2)/\log(3^2) = 1.26\ldots$.

Let us explore this idea for the standard middle-third Cantor set as shown in Figure 1b. It is constructed by starting with a unit interval, removing the (open) middle third, so as to yield two closed intervals of length $1/3$ each; removing the middle third from these and continuing inductively yields the standard middle-third Cantor set. This set consists of $N = 2$ parts (left and right) that both are rescaled versions of the original set with a factor $s = 1/3$. This Cantor set has dimension $\log 2/\log 3 \approx 0.83\ldots$: less than a curve, but more than a discrete set of points.

Here is one last example, depicted in Figure 1c: a unit square is subdivided into nine equal subsquares of size $s = 1/3$, and only the $N = 4$ subsquares at the vertices are kept and further subdivided. The dimension is $\log 4/\log 3 \approx 1.26\ldots$ as for the snowflake curve. This set is simply the Cartesian product of the middle-third Cantor set with itself.

We can play with the dimension of the Cantor set. For instance, we can start with a unit interval and remove a shorter or longer interval in the middle so as to leave $N = 2$ intervals of arbitrary length s in $(0, 1/2)$. In the next generations, we always remove an interval in the middle with the same fraction of length, so that the resulting Cantor set is self-similar again. Its dimension is $d = \log 2/\log(1/s)$, and it can assume any real value in $(0, 1)$.

What we have exploited so far is *linear self-similarity* of our sets: they consist of a finite number of pieces, each a linearly rescaled version of the entire set. It is only for such sets that the self-similarity dimension applies. Later, we define two further concepts of “fractal” dimension, box-counting dimension and Hausdorff dimension, which make sense for more general sets than the self-similarity dimension; but for the examples we have considered so far, all three dimensions apply and have the same value.

Here is a variation of the construction that leaves the realm of linearly self-similar sets: take the unit interval, replace it with two subintervals of length $s_1 \in (0, 1/2)$; each of these two intervals is replaced with two further subintervals of length $s_1 s_2$ (with s_2 in $(0, 1/2)$), and so on. If all scaling factors s_i are the same, we have a self-similar Cantor set of dimension $d = \log 2/\log(1/s_i)$ as earlier. If the first k scaling factors are arbitrary, but $s_{k+1} = s_{k+2} = \cdots = s$, then our Cantor set consists of 2^k small Cantor sets, and these small Cantor sets are linearly self-similar and have dimension $\log 2/\log(1/s)$. If the sequence s_i is not eventually constant, we need a more general concept of dimension. We would expect that the dimension would be 0 if $s_i \to 0$ and 1 if $s_i \to 1/2$. This will be true for the box-counting dimension that we define at the end of this section.

We can even construct a Cantor set within $[0, 1]$ that has positive 1-dimensional Lebesgue measure, so its dimension should certainly be 1: in the first step, we remove the middle interval of length $1/10$, say; from the remaining two intervals, we remove the central intervals of length $1/200$; then we remove four intervals of length $1/4000$, etc.. As a result, the total length of all removed intervals is $1/10+2/200+4/4000+\cdots = 0.1111\ldots = 1/9$, so the Cantor set left at the end of the process has 1-dimensional Lebesgue measure $8/9$ (note that we always remove open intervals, which ensures that the remaining set is compact, hence has well-defined Lebesgue measure).

All these Cantor sets are homeomorphic. There is even a homeomorphism of the unit interval to itself whose restriction to one Cantor set (say of dimension 0) yields the other (say of positive Lebesgue measure). (In general, a nonempty subset of a topological space is called a *Cantor set* if it is compact, totally disconnected, and without isolated points; any two metric Cantor sets are homeomorphic [**12**, Theorem 2.97]).

By taking Cartesian products of linear Cantor sets, we obtain Cantor subsets of the unit square. We can manufacture these so that they have dimension 0, positive 2-dimensional Lebesgue measure, or anything in between.

In order to define the dimensions of more general sets like the fern or the Julia set in Figure 1, we need a more general approach than self-similarity dimension. For a bounded subset X of $\mathbb{R}^n$ the idea is as follows: partition $\mathbb{R}^n$ by a regular grid of cubes of side-length s and count how many of them intersect X; if this number is $N(s)$, then we define the "box-counting dimension" (or "pixel-counting dimension") of X to be $\lim_{s\to 0} \log(N(s))/\log(1/s)$. For example, if X is a bounded piece of a d-dimensional subspace of $\mathbb{R}^n$, then $N(s) \approx c(1/s)^d$ and the dimension is d. This is what a computer can do most easily: draw the set X on the screen, count how many pixels it intersects, then draw X in a finer resolution and count again.... Of course, the limit will not exist in many cases, so the box-counting dimension is not always well-defined. It is, however, well-defined for the linearly self-similar sets discussed earlier, and for these the self-similarity dimension and the box-counting dimension coincide. Another drawback of box-counting dimension is that every countable dense subset X of $\mathbb{R}^n$ has dimension n, although a countable set should be very "small." More generally, this concept of dimension does not behave well under countable unions. The underlying reason is that all the cubes used to cover X were required to have the same size. Giving up this preconception leads to the definition of Hausdorff dimension.

3. HAUSDORFF DIMENSION. Let X be a subset of a metric space M. We define the d-dimensional Hausdorff measure $\mu_d(X)$ of X for any d in $\mathbb{R}_0^+ = [0, \infty)$ as follows:

$$\mu_d(X) = \lim_{\varepsilon \to 0} \inf_{(U_i)} \sum_i (\operatorname{diam}(U_i))^d \ , \qquad (*)$$

where the infimum is taken over all countable covers (U_i) of X such that $\operatorname{diam}(U_i) < \varepsilon$ for all i. The idea is to cover X with small sets U_i as efficiently as possible (thus the infimum) and to estimate the d-measure of X as the sum of the $(\operatorname{diam}(U_i))^d$. Smaller values of ε restrict the set of available covers, so the infimum can only increase as ε decreases. Therefore, the limit always exists in $\mathbb{R}_0^+ \cup \{\infty\}$. The measure μ_d is an outer measure on M for which all Borel sets are measurable. (Can the reader figure out the meaning of $\mu_0(X)$?)

If d is a positive integer and X is a subset of $M = \mathbb{R}^d$ with its Euclidean metric, then the d-dimensional Hausdorff measure and the d-dimensional Lebesgue measure of X coincide up to a scaling constant (a ball in $\mathbb{R}^d$ of diameter s has d-dimensional Hausdorff measure s^d). Also, countable sets have Hausdorff measure 0 for all $d > 0$. The dependence of the d-dimensional measures is governed by the following rather simple lemma:

Lemma 1 (Dependence of d-Dimensional Measure). *For any d in $\mathbb{R}_0^+$ the following statements hold:*

(1) *If $\mu_d(X) < \infty$ and $d' > d$, then $\mu_{d'}(X) = 0$.*
(2) *If $\mu_d(X) > 0$ and $d' < d$, then $\mu_{d'}(X) = \infty$.*
(3) *For each bounded set X in a given metric space there is a unique value $d =: \dim_H(X)$ in $\mathbb{R}_0^+ \cup \{\infty\}$ such that $\mu_{d'}(X) = 0$ if $d' > d$ and $\mu_{d'}(X) = \infty$ if $d' < d$.*

The first two assertions of the lemma follow directly from the definition of Hausdorff measure in (*), and together they imply the third assertion.

The value $\dim_H(X)$ in Lemma 1 is called the *Hausdorff dimension of* X. The Hausdorff measure $\mu_d(X)$ with $d = \dim_H(X)$ may be zero, positive, or even infinite.

A few remarks might help to elucidate this concept. First, the definition yields upper bounds for the dimension more easily than lower bounds: to establish an upper bound for the dimension, it suffices to find an appropriate covering for each ε; to give lower bounds, it is necessary to estimate all possible coverings. For example, the Hausdorff dimension is clearly bounded above by the box-counting dimension (if the latter exists), but the freedom to use coverings of varying sizes sometimes yields much smaller Hausdorff dimension (as mentioned earlier, any countable set has Hausdorff dimension zero).

As an example, let X be a bounded subset of a d-dimensional subspace of $\mathbb{R}^n$; to fix ideas, say X is a d-dimensional cube. For positive s let $N(s)$ be the number of open Euclidean balls in $\mathbb{R}^n$ of diameter s needed to cover X. Then $N(s) \leq c(1/s)^d$ for some constant c, hence $\mu_{d'}(X) \leq c(1/s)^d s^{d'} = cs^{d'-d}$. As $s \to 0$, the latter bound tends to 0 if $d' > d$, so $\mu_{d'}(X) = 0$ when $d' > d$ and thus $\dim_H(X) \leq d$. It is not hard to see that coverings of varying sizes would not change the dimension, so indeed $\dim_H(X) = d$. This example also shows why we need to take the limit $\varepsilon \to 0$: if $d < \dim_H(X)$, then coverings using large pieces would seem to be more efficient, whereas the limit $\varepsilon \to 0$ implies that $\mu_d(X) = \infty$ as it should be.

The equivalence between Lebesgue and Hausdorff measures implies that any set in $\mathbb{R}^d$ with finite positive d-dimensional Lebesgue measure has Hausdorff dimension d. This is another indication that Hausdorff dimension is the "right" concept.

It might be instructive to see that for linearly self-similar sets as discussed in section 2, the Hausdorff dimension never exceeds the self-similarity dimension. Indeed, if X is a bounded self-similar set of diameter R with the property that X is the union of N subsets, each similar to X and scaled by a factor $s < 1$, then X can be covered by N balls of diameter sR, or by N^2 balls of diameter s^2R, and so on. Since $s < 1$, the diameters tend to zero as $k \to \infty$. According to the definition in (*), this sequence of finite covers of X yields an upper bound for $\mu_d(X)$ of $\lim_{k\to\infty} N^k(s^kR)^d = \lim_{k\to\infty}(Ns^d)^k R^d$, and this is zero if $Ns^d < 1$ or $d > \log N/\log(1/s)$. Therefore, X has Hausdorff dimension at most $\log N/\log(1/s)$. As described earlier, upper bounds for Hausdorff dimension are easier to give than lower bounds. After all, X might well be countable and thus have Hausdorff dimension 0, even though it is linearly self-similar.

The following result collects useful properties of Hausdorff dimension that are not hard to derive directly from the definition.

Theorem 2 (Elementary Properties of Hausdorff Dimension). *Hausdorff dimension has the following properties:*

(1) *if* $X \subset Y$*, then* $\dim_H(X) \leq \dim_H(Y)$*;*
(2) *if* X_i *is a countable collection of sets with* $\dim_H(X_i) \leq d$*, then* $\dim_H \left(\bigcup_i X_i\right) \leq d$*;*
(3) *if* X *is countable, then* $\dim_H(X) = 0$*;*
(4) *if* $X \subset \mathbb{R}^d$*, then* $\dim_H(X) \leq d$*;*
(5) *if* $f : X \to f(X)$ *is a Lipschitz map, then* $\dim_H(f(X)) \leq \dim_H(X)$*;*
(6) *if* $\dim_H(X) = d$ *and* $\dim_H(Y) = d'$*, then* $\dim_H(X \times Y) \geq d + d'$*;*
(7) *if* X *is connected and contains more than one point, then* $\dim_H(X) \geq 1$*; more generally, the Hausdorff dimension of any set is no smaller than its topological dimension;*
(8) *if a subset* X *of* $\mathbb{R}^n$ *has finite positive* d*-dimensional Lebesgue measure, then* $\dim_H(X) = d$*.*

For linearly self-similar sets, the Hausdorff dimension coincides with the self-similarity dimension. Thus Hausdorff dimension is *not* preserved under homeomorphisms, as we observed in the case of linear Cantor sets in section 2. Indeed, topology and Hausdorff dimension (or measure theory in general) sometimes have a tenuous coexistence.

Some people like the word "fractal". One possibility is to define a set X to be a "fractal" if its Hausdorff dimension is not an integer (X has "fractal dimension"). The problem with this definition is that, for example, in $\mathbb{R}^d$ one can have a Cantor set whose Hausdorff dimension is an arbitrary real number in $[0, d]$ (recall our examples). A curve in $\mathbb{R}^d$ can have any dimension in $[1, d]$, and so on. Why should a curve X in $\mathbb{R}^d$ be a "fractal" when its dimension is 1.001 or 1.999, but not when its dimension is 2? A better definition is this: X is a "fractal" if its Hausdorff dimension strictly exceeds its topological dimension. More information on "fractal sets" and Hausdorff dimension can be found in [**9**].

4. KARPIŃSKA'S EXAMPLE. Here we give a beautiful and surprising example due to Karpińska.

Example (Karpińska). There exist sets E and R in the complex plane $\mathbb{C}$ with the following properties:

(1) E and R are disjoint;
(2) E is totally disconnected but has finite positive 2-dimensional Lebesgue measure (hence E has topological dimension 0 and Hausdorff dimension 2);

(3) each connected component of R is a curve connecting a single point of E to ∞;
(4) R has Hausdorff dimension 1.

Why is this surprising? Each connected component of R is a single curve connecting one point of E to ∞, so each connected component of $E \cup R$ contains one point of E and a whole curve in R. The set $E \cup R$ is an uncountable union of such things, a union so large that the union of all these single points of E acquires positive 2-dimensional Lebesgue measure, hence Hausdorff dimension 2. In the same union, the dimension of R stays 1, so a 1-dimensional set can be big enough to connect each point in the 2-dimensional set E to ∞ via its own curve, all curves and endpoints being disjoint!

Once this phenomenon is discovered (which happened unexpectedly in complex dynamics [**13**]), its proof is surprisingly simple. For the set E we use a Cantor set made from an initial closed square, which is replaced with four disjoint closed subsquares, each of which is in turn replaced with four smaller disjoint subsquares, etc. It is quite easy to arrange the sizes of the squares so that the resulting Cantor set has positive area: one simply has to make sure that the area lost at each stage is small enough so that the cumulative area lost is less than, say, half the area of the initial square. This leaves a Cantor set with positive area (which is simply a product of two one-dimensional Cantor sets with positive 1-dimensional measure).

The construction of the curves is indicated in Figure 2. We start with an initial rectangle that terminates at the initial square. When the square is refined into four closed subsquares, the rectangle is subdivided into four parallel closed subrectangles and extended through the initial square so that the four extended subrectangles reach the four subsquares. This process can be repeated at each subsequent stage to create a collection of "rectangular tubes" connecting the 4^n squares in the nth subdivision step with the right side of the original square. The nth refinement step yields 4^n squares, each of which has a "rectangular tube" attached to it, so that we have 4^n connected components. Let X_n be the set constructed in step n (consisting of 4^n squares together with their "rectangular tubes"). Then X_{n+1} is a subset of X_n. More precisely, each step refines each of the 4^n connected components of X_n into four connected components of X_{n+1}.

It is clear that the countable intersection $\bigcap X_n$ yields a compact set X with the following properties: each connected component of X consists of one point of E and a curve connecting that point to the right end of the initial rectangle. Set $R = X \setminus E$. All that remains to

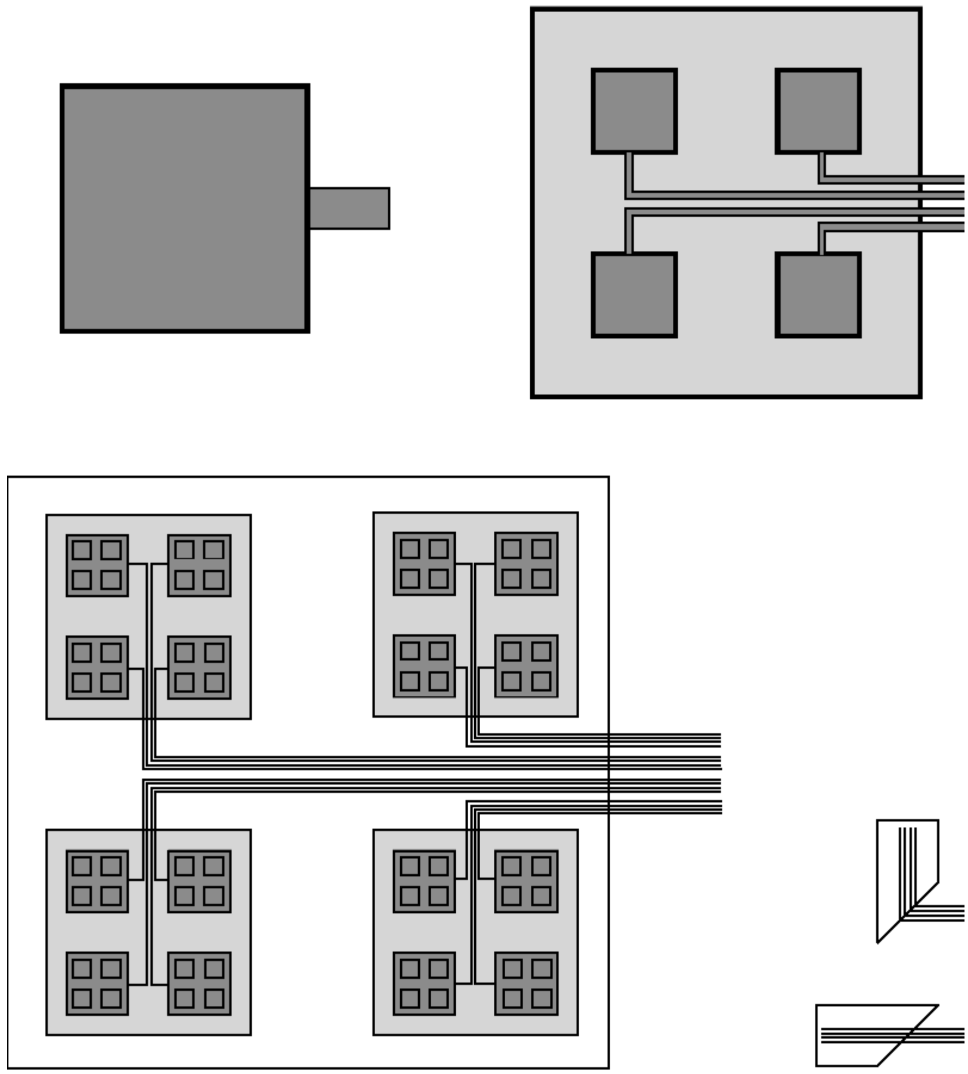

FIGURE 2. The construction of Karpińska's example. Shown are the initial square and the initial rectangle, as well as two refinement steps. In each step, we keep the dark shaded area, so we have a nested sequence of compact sets (the area of the previous refinement step is shown in a lighter shade). The detail in the lower right shows that a Cantor set of curves can be given a right-angled turn by replacing a subset with its mirror-image, not changing the dimension.

show is that R has Hausdorff dimension 1. Observe that R restricted to the initial rectangle is a product of an interval (in the horizontal direction) with a Cantor set (in the vertical direction). We can arrange things so that the vertical Cantor set has Hausdorff dimension 0, so the subset of R within the initial rectangle has Hausdorff dimension 1. Next consider the subset of R within the original square but outside of the first generation subsquares. This looks like a Cantor set of curves as before, but with a right-angled turn in the middle. If half of this curve is replaced with its mirror-image, we obtain a proper Cantor set of curves with dimension 1 (see the detail in Figure 2), and this reflection does not change the Hausdorff dimension. The entire set R is a countable union of such 1-dimensional Cantor sets of curves, each with one turn, that become smaller as they approach E. Therefore, R still has dimension 1.

The last small issue is that the curves in R do not connect E to ∞, for they terminate at the right end of the initial rectangle. This shortcoming can be cured by extending the initial rectangle to the right by countably many copies of itself.

Certainly, one might find this result surprising. Is it an artifact of the concept of Hausdorff dimension, indicating that its definition is problematic? The answer is no: a weaker form of this surprise occurs even from the point of view of planar Lebesgue measure. Our construction assures that R has zero planar measure, whereas E has strictly positive planar measure. Hausdorff dimension is a way of making the surprise more precise and stronger; the surprise lies in the sets E and R, not in any definition.

We conclude this section with an example of an "impossible" set that was brought to our attention by Adam Epstein: Larman [**15**] defines a compact set in $\mathbb{R}^n$ (for any $n \geq 3$) that is the disjoint union of closed line segments and has positive n-dimensional Lebesgue measure. However, removing the two endpoints from each segment, a set with zero measure remains (this is impossible in $\mathbb{R}^2$). In other words, we have a bunch of uncooked spaghetti in n-space so that all the nutrition lies in the endpoints. We now proceed to show how much better we can do, even in $\mathbb{R}^2$, when using cooked spaghetti and complex dynamics.

5. DYNAMICS OF COMPLEX SINE MAPS. In the rest of this article, we describe how a much stronger result arises quite naturally in the study of very simple dynamical systems, such as the one given by iterating as simple a map (apparently!) as $z \mapsto \pi \sinh z$ on $\mathbb{C}$ (but recall that Karpińska developed her example of section 4 only after she had discovered an analogous phenomenon in the dynamics of exponential maps). We again have sets E and R as in Karpińska's example, but this time $E \cup R = \mathbb{C}$. As before, each path component of R is a curve connecting one point in E to ∞, and R still has Hausdorff dimension 1, but now the set $E = \mathbb{C} \setminus R$ has infinite Lebesgue measure, even full measure in $\mathbb{C}$, and is so big that its complement has dimension 1—nevertheless, each point of E can be connected to ∞ by one or even several curves in R!

We set up the construction as follows. Let $f\colon \mathbb{C} \to \mathbb{C}$ be given by $f(z) = k\pi \sinh z = (k\pi/2)(e^z - e^{-z})$ with a nonzero integer k. We study the dynamics given by iteration of f: by $f^{\circ n}$ we denote the nth iterate of f (i.e., $f^{\circ 0} = \mathrm{id}$ and $f^{\circ(n+1)} = f \circ f^{\circ n}$). Of principal interest is the set of "escaping points," meaning the set

$$I := \{z \in \mathbb{C}\colon\, f^{\circ n}(z) \to \infty \text{ as } n \to \infty\}$$

consisting of those points that converge to ∞ under iteration of f (in the sense that $|f^{\circ n}(z)| \to \infty$). Here, I stands for "infinity"; this set plays a fundamental role in the iteration theory of polynomials [**20**, sec. 18] and is just beginning to emerge as equally important for transcendental entire functions. Eremenko [**8**] has shown that for every transcendental entire function the set I is nonempty, and he asked whether every path component of I was unbounded. An affirmative answer to this question is currently known only for functions of the form $z \mapsto \lambda e^z$ [**25**] or $z \mapsto ae^z + be^{-z}$ [**23**], where λ, a, and b are nonzero complex numbers. The latter family includes our functions f. (Recently, this question was answered affirmatively in greater generality in [**22**], [**21**], and [**1**]. However, Eremenko's question is not true for all transcendental functions; counterexamples are constructed in [**22**], [**21**]). The following is a special case of what is known for this family [**24**]:

Theorem 3 (Dynamic Rays of Sine Functions).

(1) *For the function $f(z) = k\pi \sinh z$ with a nonzero integer k each path component of I is a curve $g\colon (0,\infty) \to I$ or $g\colon [0,\infty) \to I$ such that $\lim_{t\to\infty} \operatorname{Re} g(t) = \pm\infty$. Each curve is contained in a horizontal strip of height π. (These curves are called "dynamic rays.")*

(2) *For each such curve g the limit $z := \lim_{t\searrow 0} g(t)$ exists in $\mathbb{C}$ and is called the "landing point" of g ("the dynamic ray g lands at z"). If $t > t' > 0$, then the two points $g(t)$ and $g(t')$ escape in such a way that*

$$|\mathrm{Re} f^{\circ k}(g(t))| - |\mathrm{Re} f^{\circ k}(g(t'))| \longrightarrow \infty\ .$$

(3) *Conversely, every point z of $\mathbb{C}$ either is on a unique dynamic ray or is the landing point of one, two, or four dynamic rays (i.e., either $z = g(t)$ for a unique dynamic ray g and a unique $t > 0$, or $z = \lim_{t \searrow 0} g(t)$ for up to four rays g).*

We will indicate in section 7 why these results are not too surprising, even though the precise proofs are technical. This leads quite naturally to a decomposition $\mathbb{C} = E \dot{\cup} R$ as required for our result:

$$R := \bigcup_{\text{rays } g} g((0,\infty)) \ , \qquad E := \bigcup_{\text{rays } g} \lim_{t \searrow 0} g(t) \ .$$

If your intuition for the complex sine map is better than for the hyperbolic variant, then you may use the former instead: the situation is exactly the same, except that the complex plane is rotated by 90^0. We prefer to use the sinh map because in half-planes far to the left or far to the right it is essentially the same as $z \mapsto e^{-z}$ and $z \mapsto e^z$, respectively (up to a factor of 2). Note also that the parametrization of our rays $g\colon (0,\infty) \to I$ differs from the one used in [**23**] and [**24**].

6. THE PARABOLA CONDITION. The driving force behind our results is a fundamental lemma of Karpińska [**13**], adapted to fit our purposes. For real numbers ξ in $(0,\infty)$ and p in $(1,\infty)$ consider the sets

$$P_{p,\xi} := \left\{x + iy \in \mathbb{C}\colon |x| > \xi, |y| < |x|^{1/p}\right\}$$

(the "p-parabola," restricted to real parts greater than ξ). Also let $I_{p,\xi}$ be the subset of I consisting of those escaping points z for which $f^{\circ n}(z)$ is in $P_{p,\xi}$ for all n (the set of points that escape within $P_{p,\xi}$). The results in this section hold for all maps $f(z) = ae^z + be^{-z}$ with a and b nonzero complex numbers.

Lemma 2 (Dimension and the Parabola Condition). *For each p in $(1,\infty)$ and each sufficiently large ξ, the set $I_{p,\xi}$ has Hausdorff dimension at most $1 + 1/p$.*

Proof. First observe that we seek only an upper estimate for the Hausdorff dimension. Therefore it suffices to find a family of covers whose sets have diameters less than any specified $\varepsilon > 0$ so that their combined d-dimensional Hausdorff measure is bounded for each d with $d > 1 + 1/p$. For bounded subsets of $I_{p,\xi}$, we construct a finite cover in "generations" zero, one, two, ... so that each set in the nth generation is refined into finitely many smaller sets in the $(n+1)$th generation. We do this in such a way that the diameters of all sets tend to zero as the number n of generations tends to infinity, and so that the combined

d-dimensional Hausdorff measure of all sets in the nth generation decreases as n tends to infinity provided that $d > 1+1/p$. In view of the definition in (*), this implies that the d-dimensional Hausdorff measure of $I_{p,\xi}$ is finite whenever $d > 1+1/p$, hence that the Hausdorff dimension of $I_{p,\xi}$ is at most $1+1/p$.

We first outline the proof while making a number of simplifications; we then argue that these do not matter. The first simplification is that when $\operatorname{Re} z > \xi$, we write $f(z) = ae^z$ (ignoring the exponentially small error term be^{-z}), and when $\operatorname{Re} z < -\xi$, we write $f(z) = be^{-z}$. For simplicity, we ignore certain bounded factors: we do not distinguish between side-lengths and diameters of squares, and we suppress factors like $\pi/|a|$ or $\pi/|b|$ that appear all over the place but influence only Hausdorff measure, not dimension.

For the purposes of this proof, "standard square" means a closed square of side-length π with sides parallel to the coordinate axes. The image $f(Q)$ of a standard square Q is a semiannulus bounded by two semicircles and two straight radial boundary segments. If the imaginary parts of Q are varied while the real parts are kept fixed, then the semiannulus $f(Q)$ rotates around the origin. We always adjust the imaginary parts of our standard squares so that $f(Q)$ is entirely contained in the right or the left half-plane, which is equivalent to the condition that the two straight radial boundary segments of $f(Q)$ are contained in the imaginary axis.

Cover $P_{p,\xi}$ by a countable collection of standard squares with disjoint interiors. Fix any particular square Q_0 with real parts in $[x, x+\pi]$, where $x \geq \xi$ and ξ is sufficiently large (the case where $x \leq -\xi$ is analogous). Now $f(Q_0)$ intersects $P_{p,\xi}$ in an approximate rectangle with real parts between $\pm|a|e^x$ and $\pm|a|e^{x+\pi}$ and imaginary parts at most $(|a|e^{x+\pi})^{1/p} = (|a|e^\pi)^{1/p}e^{x/p}$. Therefore, the number of standard squares of side-length π needed to cover $f(Q_0) \cap P_{p,\xi}$ is approximately $ce^x \cdot e^{x/p} = ce^{x(1+1/p)}$, where

$$c = |a|(e^\pi - 1) \cdot 2(|a|e^\pi)^{1/p}/\pi^2 = 2(e^\pi - 1)e^{\pi/p}|a|^{1+1/p}\pi^{-2} \ .$$

Transporting these squares back into Q_0 via f^{-1}, we cover not all of Q_0, but all those points z of Q_0 with $f(z)$ in $P_{p,\xi}$ (see Figure 3). Since $|f'(z)| > |a|e^x$ on Q_0, the covering sets are approximate squares of side-length at most $(\pi/|a|)e^{-x}$, hence diameter at most $(\sqrt{2}\pi/|a|)e^{-x}$. Ignoring bounded factors, we simplify this value to e^{-x}. We call this covering the "first generation covering" within Q_0 (while $\{Q_0\}$ itself is the zeroth generation covering).

Let us see what effect this refinement has on the d-dimensional Hausdorff measure. The covering at generation zero is a standard square and has constant measure. In generation one, the covering of Q_0 has measure $\sum(\mathrm{diam}(U_i))^d \approx ce^{x(1+1/p)}(e^{-x})^d = ce^{x(1+1/p-d)}$. Since $d > 1 + 1/p$, this is small for large x in (ξ, ∞), so this first refinement reduces the measure.

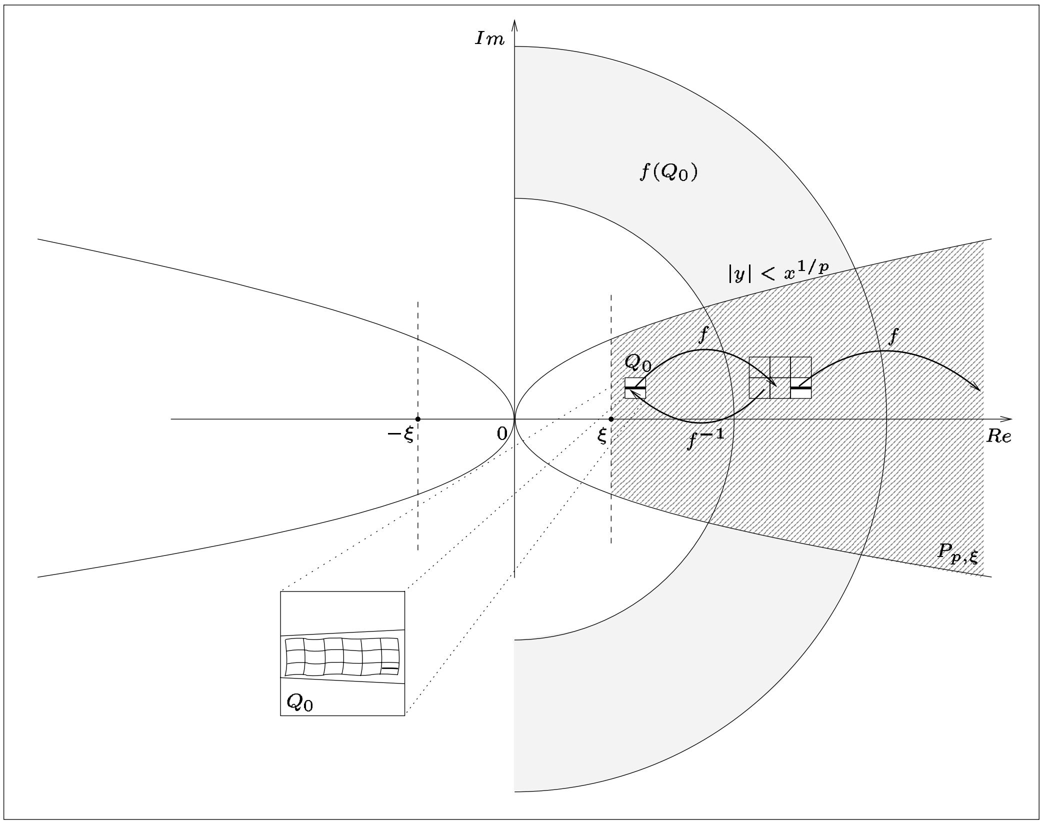

FIGURE 3. Calculating the Hausdorff measure of $I_{p,\xi}$ involves a partition by iterated preimages of a square grid, as well as refinements of such a partition.

We continue to refine our coverings so that the diameters of the covering sets tend to zero, while the d-dimensional Hausdorff measure does not increase. Each approximate square of generation n gets replaced with some number of much smaller approximate squares of generation $n + 1$. What brings the dimension down is that we consider only orbits in $P_{p,\xi}$, throwing away everything that leaves this parabola under iteration. We may thus maintain the inductive claim that all approximate squares of generation n have images under f, $f^{\circ 2}$, ..., $f^{\circ n}$ that intersect $P_{p,\xi}$; moreover, if Q' is an approximate square of generation

n, then $f^{\circ n}(Q')$ is a standard square whose points have very large real parts, say in $[y, y+\pi]$ for some y satisfying $y \geq \xi$.

Let $\lambda := |(f^{\circ n})'(z)|$ for some z in Q' (this derivative is essentially constant on Q', as noted later). Then Q' is an approximate square of side-length π/λ, so it contributes approximately π^d/λ^d to the d-dimensional Hausdorff measure. We now determine what happens to this measure under refinement.

Just as in the first step, $f^{\circ(n+1)}(Q') \cap P_{p,\xi}$ is covered by $N_y := ce^{y(1+1/p)}$ standard squares of side-length π, so the standard square $f^{\circ n}(Q') \cap f^{-1}(P_{p,\xi})$ is covered by N_y approximate squares of side-length $(\pi/|a|)e^{-y}$ or $(\pi/|b|)e^{-y}$. Ignoring constants again, we simplify this to e^{-y}. We need N_y very small approximate squares to cover those points in Q' that remain in $P_{p,\xi}$ for $n+1$ iteration steps. These N_y approximate squares within Q' have side-lengths approximately e^{-y}/λ, so their contribution to the d-dimensional Hausdorff measure within Q' is roughly $N_y \cdot (e^{-y}/\lambda)^d = ce^{y(1+1/p-d)}\lambda^{-d}$, whereas the contribution of Q' before refinement was $\pi^d\lambda^{-d}$. Therefore, if $d > 1+1/p$, each refinement step reduces the d-dimensional Hausdorff measure (at least when ξ is large). It follows that the d-dimensional Hausdorff measure of $Q_0 \cap I_{p,\xi}$ is finite whenever $d > 1+1/p$, so Lemma 1 implies that

$$\dim_H(Q_0 \cap I_{p,\xi}) \leq 1 + 1/p \; .$$

Since $I_{p,\xi}$ is a countable union of sets of dimension at most $1+1/p$, the claim follows.

There are two main inaccuracies in this proof: we have ignored constants, and we have ignored the geometric distortions caused by the mapping f and its iterates. The latter are induced by two problems: we have disregarded one of the two exponential terms in f, and the continued backward iteration of standard squares under a finite iterate of f might distort the shape of the squares because f' or $(f^{\circ n})'$ is not exactly constant on small approximate squares. However, this distortion problem is easily cured by a useful lemma usually called the *Koebe Distortion Theorem* [**19**, Theorem 2.7] for conformal mappings: *for $r \geq 1$ let $\mathbb{D}_r := \{z \in \mathbb{C}\colon |z| < r\}$, and let K_r be the family of injective holomorphic mappings $g\colon \mathbb{D}_1 \to \mathbb{C}$ that have extensions to $\mathbb{D}_r$ as injective holomorphic mappings. Then for each $r > 1$ all maps g in K_r have distortions (on $\mathbb{D}_1$) that are uniformly bounded in terms only of r.* Here the precise definition of distortion is irrelevant: any quantity can be used that measures the deviation of g from being an affine linear map. A more precise way of stating this result is as follows: if we normalize so that $g(0) = 0$ and $g'(0) = 1$, then the space K_r is compact (in the topology of uniform convergence). You may want to remember this fact as the "yellow of the egg theorem": when you spill an egg into a frying pan, the whole egg can assume any shape (this represents the Riemann map from the disk of radius $r > 1$ onto a simply connected domain in $\mathbb{C}$), but its smaller yolk (the yellow of the egg, represented by the unit disk) is not distorted too much (it remains essentially a round disk, and derivatives at any two points differ at most by a bounded factor).

In our context, the maps are easily seen to have bounded distortion, so we may assume that the nth iterate $f^{\circ n}$, which maps an nth generation approximate square to a standard square, is a linear map with constant complex derivative. All this does is to introduce a bounded factor in the diameters and in the number of sets in the coverings. These factors do not increase under repeated refinement.

The second simplification was that at several stages we ignored certain bounded factors. For example, in the calculation of Hausdorff measures, we replaced diameters with side-lengths. This introduces a factor of $\sqrt{2}$ into the measure estimates, but it has no impact on the dimension. Similarly, we have ignored factors like $\pi/|a|$ or $\pi/|b|$, we have counted the number of necessary squares only approximately, ignoring boundary effects, and we have assumed that the derivative of $f^{\circ n}$ is constant on small approximate squares. Each of these simplifications might lead to a change in the Hausdorff measure by a bounded factor, but the dimension remains unaffected. The crucial fact is that refinements do not increase the d-dimensional measure when $d > 1 + 1/p$ and x is sufficiently large, and this fact is correct. □

We have now shown that escaping orbits that spend their entire lives within the truncated parabolas $P_{p,\xi}$ form a very small set. It is easy to see that the same is true for the set of points that spend their entire orbits within $P_{p,\xi}$ except for finitely many initial steps (see Corollary 1). Nonetheless, the surprising fact is that from a different (topological) point of view, *most* orbits do exactly that: after finitely many initial steps, they enter $P_{p,\xi}$ and remain there. All this is based on the following result.

Lemma 3 (Horizontal Expansion). *For each $h > 0$ there is an $\eta > 0$ with the following property: if (z_k) and (w_k) are two orbits such that $|\mathrm{Im}(z_k - w_k)| < h$ for all k and $|\mathrm{Re}\, z_1| > |\mathrm{Re}\, w_1| + \eta$, then for each pair p and ξ there is an N such that z_k belongs to $P_{p,\xi}$ whenever $k \geq N$.*

Sketch of proof. We do not give a precise proof, which involves easy but lengthy estimates. Instead, we outline the main idea, again ignoring bounded factors. Let $c := \max\{|a|, |b|\}$ and $c' := \min\{|a|, |b|\}$, where $f(z) = ae^z + be^{-z}$. We start by estimating $\mathrm{Re} f(w)$ for sufficiently large $|\mathrm{Re}\, w|$:

$$|\mathrm{Re} f(w)| + c \leq |f(w)| + c \leq c \exp|\mathrm{Re}\, w| + c < \exp(|\mathrm{Re}\, w| + c) \ ,$$

which yields $|\mathrm{Re}\, w_{k+1}| \leq |w_{k+1}| < \exp^{\circ k}(|\mathrm{Re}\, w_1| + c)$ by induction. Therefore

$$|\mathrm{Im}\, z_{k+1}| \leq |\mathrm{Im}\, w_{k+1}| + h \leq |w_{k+1}| + h \leq \exp^{\circ k}(|\mathrm{Re}\, w_1| + c) + h \ .$$

If $|\mathrm{Re}\, z| > |\mathrm{Re}\, w| + \eta$ and both are sufficiently large, then

$$|f(z)| \geq c' \exp|\mathrm{Re}\, z| > c' \exp(|\mathrm{Re}\, w|) \exp\eta \approx |f(w)| e^{\eta} \ ,$$

hence $|f(z)| \gg |f(w)|$ if η is large. Since the imaginary parts of $f(z)$ and $f(w)$ are approximately equal, the absolute value of $f(z)$ must come mainly from its real part, so

$$|\mathrm{Re}\, f(z)| - 1 \geq \frac{1}{e}|f(z)| \approx \exp(|\mathrm{Re}\, z| - 1) \ ,$$

and we get the inductive relation $|\mathrm{Re}\, z_{k+1}| - 1 \geq \exp^{\circ k}(|\mathrm{Re}\, z_1| - 1)$.

Now if η is sufficiently large, then indeed there exist T and t with $T > t > 0$ such that

$$|\mathrm{Re}\, z_{k+1}| > \exp^{\circ k}(T) > \exp^{\circ k}(t) > |\mathrm{Im}\, z_{k+1}|$$

for almost all k. Once k is so large that $\exp^{\circ k}(T) > p \exp^{\circ k}(t)$, we have $\exp^{\circ(k+1)}(T) > (\exp^{\circ(k+1)}(t))^p$. The assertion of the lemma follows. □

We can finally prove that the set R of dynamic rays has Hausdorff dimension 1:

Corollary 1 (Hausdorff Dimension of the Union of Dynamic Rays). *The set R consisting of all dynamic rays has Hausdorff dimension* 1.

Proof. Consider an arbitrary point z of R, say $z = g(t)$ for some ray g and some $t > 0$. Let $w := g(t')$ for some t' in $(0, t)$. Then by Theorem 3 there is an h not exceeding π such that $|\mathrm{Im}(f^{\circ k}(z) - f^{\circ k}(w))| \leq h$ for all k, and $|\mathrm{Re} f^{\circ k}(z)| - |\mathrm{Re} f^{\circ k}(w)| \to \infty$ as $k \to \infty$.[1] Fix p with $p > 1$. For each choice of $\xi > 0$ Lemma 3 implies that there is an N such that $f^{\circ N}(z)$ lies in $I_{p,\xi}$.

[1] Strictly speaking, we have stated Theorem 3 only for certain maps $z \mapsto ae^z + be^{-z}$ as specified in the theorem, and only such maps will be used in the following sections, so one can read this entire paper with only the maps $z \mapsto k \sinh z$ in mind. However, the results in this section are true for all maps $z \mapsto ae^z + be^{-z}$ with a and b in $\mathbb{C} \setminus \{0\}$.

We have thus shown that $R \subset \bigcup_{N \geq 0} f^{-N}(I_{p,\xi})$. If ξ is sufficiently large, Lemma 2 ensures that $\dim_H(I_{p,\xi}) \leq 1+1/p$. Now for each N the set $f^{-N}(I_{p,\xi})$ is a countable union of holomorphic preimages of $I_{p,\xi}$, so parts 2 and 5 of Theorem 2 imply that $\dim_H(f^{-N}(I_{p,\xi})) \leq 1 + 1/p$. It follows that $\dim_H(R) \leq 1 + 1/p$. Since this is true for every p greater than 1, we conclude that $\dim_H(R) \leq 1$. Equality follows because R contains curves. $\square$

Now we have our dimension paradox complete for $f(z) = k\pi \sinh z$, using Theorem 3 (which still requires proof): every point z of $\mathbb{C}$ either lies on a dynamic ray, and thus is in R, or it is a landing point of one or several dynamic rays in R that connect z to ∞. Since the set R has Hausdorff dimension 1 (hence planar Lebesgue measure zero), the set $E = \mathbb{C} \setminus R$ has full measure and is in fact everything but the one-dimensional set R. This proves Theorem 1 (further details can be found in [**24**]).

7. DYNAMICAL FINE-STRUCTURE OF THE HYPERBOLIC SINE MAP. We now proceed to explain why Theorem 3 is true, and why it is interesting from the perspective of dynamical systems. For simplicity, we restrict attention to maps $f(z) = k\pi \sinh z = (k\pi/2)(e^z - e^{-z})$ with k a positive integer (see Figure 4).

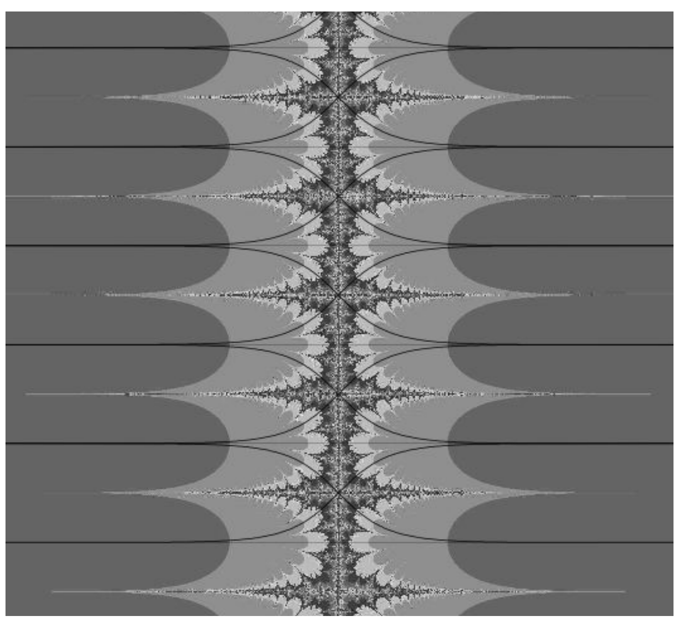

FIGURE 4. The dynamical plane of the map $f\colon z \mapsto \pi \sinh z$. Several dynamic rays are shown.

First observe that f is periodic with period $2\pi i$ (f is the rotated sine function) and maps $i\mathbb{R}$ onto the interval $[-k\pi i, k\pi i]$. Notice also that $f\colon \mathbb{R} \to \mathbb{R}$ is a homeomorphism with $f(0) = 0$ and $f'(x) \geq \pi$ for all x in $\mathbb{R}$, from which it follows that $\mathbb{R} \setminus \{0\}$ is contained in the escape set I. In fact, $\mathbb{R}^+$ and $\mathbb{R}^-$ are two of the path components of I: they are both dynamic rays, and they connect each of their points to ∞ through I. Since $f(z + i\pi) = -f(z)$, other dynamic rays include the curves $i\pi n + \mathbb{R}^+$ and $i\pi n + \mathbb{R}^-$ for integers n. These map under f onto $\mathbb{R}^+$ or $\mathbb{R}^-$. This gives a useful partition for the dynamics: for n in $\mathbb{Z}$ set

$$\begin{aligned} U_{n,R} &:= \{z \in \mathbb{C}\colon \operatorname{Re} z > 0, \operatorname{Im} z \in (2\pi n, 2\pi(n+1))\} , \\ U_{n,L} &:= \{z \in \mathbb{C}\colon \operatorname{Re} z < 0, \operatorname{Im} z \in (2\pi n, 2\pi(n+1))\} . \end{aligned}$$

(This is an ad hoc partition for our special maps f that uses the symmetry given by the invariant real and imaginary axes. In [**24**], a different partition is used that works for more general maps f.)

The geometry of the mapping f is such that its restrictions are conformal isomorphisms

$$f\colon U_{n,R} \to \mathbb{C} \setminus (\mathbb{R}^+ \cup [-k\pi i, k\pi i])$$

and

$$f\colon U_{n,L} \to \mathbb{C} \setminus (\mathbb{R}^- \cup [-k\pi i, k\pi i]) ,$$

so the image of each $U_{n,\times}$ is a one-sheeted covering of $U_{n,\times}$. This is a useful property, called the *Markov property*, that aids in reducing many dynamical questions to questions about symbolic dynamics.

Let $\mathbb{Z}_R := \{\dots, -2_R, -1_R, 0_R, 1_R, 2_R, \dots\}$ and $\mathbb{Z}_L := \{\dots, -2_L, -1_L, 0_L, 1_L, 2_L, \dots\}$ be two disjoint copies of $\mathbb{Z}$, and let $\mathcal{S}\colon = (\mathbb{Z}_R \cup \mathbb{Z}_L)^{\mathbb{N}}$ be the space of sequences with elements in $\mathbb{Z}_R \cup \mathbb{Z}_L$. To each z in $\mathbb{C}$ we assign an *itinerary* $\underline{s} = s_1 s_2 s_3 \dots$ in $\mathcal{S}$ such that $s_k = n_R$ if $f^{\circ(k-1)}(z)$ is in $\overline{U}_{n,R}$ and $s_k = n_L$ if $f^{\circ(k-1)}(z)$ is in $\overline{U}_{n,L}$. There are ambiguities if the orbit of z ever enters $\mathbb{R}$ or $[-k\pi i, k\pi i]$, but such points are easy to understand anyway, and we admit all itineraries in such cases (the number of possible itineraries for a given point z can be as large as four; see the discussion in the proof of Theorem 3). The following lemma furnishes a mechanism for understanding the detailed dynamics of f:

Lemma 4 (Symbolic Dynamics and Curves). *For each sequence $\underline{s}$ in $\mathcal{S}$ the set of all points z in $\mathbb{C}$ with itinerary $\underline{s}$ is either empty or a curve that connects ∞ to a well-defined landing point in $\mathbb{C}$. For each such curve each of its points other than the landing point escapes.*

Sketch of proof. For each positive N let $U_{\underline{s},N}$ be the set of points z such that the first N entries in the itinerary of z coincide with the first N entries of $\underline{s}$. With the aid of the Markov property it is quite easy to see that each $\overline{U}_{\underline{s},N}$ is a closed, connected, and unbounded subset of $\mathbb{C}$. Moreover, in the topology of the Riemann sphere, adding the point ∞ to these sets yields compact and connected sets containing ∞. Let

$$C_{\underline{s}} := \bigcap_{N\in\mathbb{N}} (\overline{U}_{\underline{s},N} \cup \{\infty\}) \ .$$

This is obviously a nested intersection, so $C_{\underline{s}}$ is compact and connected and contains ∞. If $C_{\underline{s}} = \{\infty\}$, then we have nothing to prove. Otherwise, we can show that f is expanding enough so that for any two points z and w in $C_{\underline{s}}$ and any $\eta > 0$ there is an n such that $||\mathrm{Re} f^{\circ n}(z)| - |\mathrm{Re} f^{\circ n}(w)|| > \eta$. Lemma 3 implies then that at least one of the points z and w escapes. (The expansion comes from the fact that $U := \mathbb{C} \setminus \{-i\pi, 0, i\pi\}$ carries a unique normalized hyperbolic metric and that $f^{-1}(U) \subset U$. With respect to this metric on U, every local branch of f^{-1} is contracting, which makes f locally expanding. This argument requires nothing but the fact that the universal cover of U is $\mathbb{D}$, plus the Schwarz lemma on holomorphic self-maps of $\mathbb{D}$.)

It follows that all points in $C_{\underline{s}} \setminus \{\infty\}$ escape, with at most one exception; the estimates in Lemma 3 imply that these points escape extremely fast. This means that for almost all z in $C_{\underline{s}} \setminus \{\infty\}$ we have $f^{\circ n}(z) \to \infty$ very fast, hence $|(f^{\circ n})'(z)| \to \infty$ very fast. Thus the forward iterates of z are very strongly expanding. Conversely, if $z_n := f^{\circ n}(z)$, then the branch of f^{-n} sending z_n to z is strongly contracting. This implies that the boundaries of the $U_{\underline{s},N}$, which are curves, converge locally uniformly to $C_{\underline{s}}$. This ensures that $C_{\underline{s}}$ is a curve. $\square$

This lemma is all we need to establish the two main results about the dynamics of the function f.

Proof of Theorem 3. Every point z in $\mathbb{C}$ has at least one associated itinerary. If it has more than one, then under iteration it must map into $i\mathbb{R}$ or into $\mathbb{R} + 2\pi i\mathbb{Z}$. In the latter case, the next iteration lands in $\mathbb{R}$, so the orbit reaches either the fixed point 0 or one of the two dynamic rays $\mathbb{R}^+$ or $\mathbb{R}^-$. If the orbit reaches $i\mathbb{R}$, then from that iteration on it spends its entire forward orbit in the interval $[-k\pi i, k\pi i]$; in particular, the orbit is bounded. Therefore, a point has four itineraries if and only if its orbit eventually terminates at 0. A point has two itineraries if it lands in the invariant interval $[-k\pi i, k\pi i]$ (and has bounded orbit), or if it lands in $\mathbb{R}^+ \cup \mathbb{R}^-$ and escapes. Every other point has a single itinerary.

Recall that the set of points with a given itinerary is a single dynamic ray consisting of escaping points, together with the unique landing point of the ray (Lemma 4). This implies that every point in $\mathbb{C}$ either lies on a unique dynamic ray or is the landing point of one, two, or four dynamic rays. This proves statements 2 and 3 in the theorem.

For statement 1, we have constructed rays consisting of escaping points, and the partition makes it clear that every ray has real parts tending to $\pm\infty$, while the imaginary parts are constrained to some interval of length π. It is clear that each escaping point either is on a unique ray or is the landing point of a ray; if a ray lands at an escaping point, then the landing point neither lies on any other ray nor is the landing point of another ray. Therefore, each ray (possibly together with its endpoint) is contained in a path component of I. It is also true that each path component of I consists of a single ray, possibly together with its endpoint. The proof of this fact requires some ingredients from continuum theory (see [**11**, sec. 4]). $\square$

8. LEBESGUE MEASURE AND ESCAPING POINTS. From the point of view of dynamical systems, an important question to ask is the following: *What do most orbits do under iteration?* From a topological vantage point, most points are on dynamic rays, rather than being endpoints of rays. On the other hand, since the union of the rays has Hausdorff dimension 1, measure theory says that most points are endpoints of rays. However, as we will now see, even measure theory asserts that most points in $\mathbb{C}$ escape (for our maps $z \mapsto k\pi \sinh z$): this assertion is a combination of results of McMullen [**18**] and Bock [**3**]. As a result, almost all points are escaping endpoints of rays. Along the way, we visit a result of Schubert [**26**] that settles a conjecture of Milnor [**20**, sec. 6] in the affirmative.

Theorem 4 (Lebesgue Measure of Escaping Points).

(1) *For every map $z \mapsto \lambda e^z$ with $\lambda \neq 0$ the set I of escaping points has two-dimensional Lebesgue measure zero but Hausdorff dimension* 2 [**18**].

(2) *However, for every map $z \mapsto ae^z + be^{-z}$ with $ab \neq 0$ the set I has infinite two-dimensional Lebesgue measure* [**18**]. *For every strip $S = \{z \in \mathbb{C}\colon \alpha \leq \operatorname{Im} z \leq \beta\}$ in $\mathbb{C}$ the two-dimensional Lebesgue measure of $S \setminus I$ is finite* [**26**].

Sketch of proof. Choose $\xi > 0$, and set $\mathbb{H}_\xi = \{z \in \mathbb{C}\colon \operatorname{Re} z > \xi\}$. We show that for every map $E(z) = \lambda \exp(z)$ and sufficiently large ξ the set $Z_\xi := \{z \in \mathbb{C}\colon \operatorname{Re} E^{\circ n}(z) > \xi \text{ for all } n\}$ has measure zero. In fact, for each square Q in $\mathbb{H}_\xi$ of side-length 2π with sides parallel to the coordinate axes the image $E(Q)$ is a large annulus in $\mathbb{C}$, and the probability that a point z in Q has $E(z)$ in $\mathbb{H}_\xi$ is approximately $1/2$. The chance of surviving n consecutive iterations in Z_ξ is then 2^{-n} (assuming independence of probabilities in the consecutive steps). Hence, the set of points z in Q whose entire orbits lie in $\mathbb{H}_\xi$ has two-dimensional Lebesgue measure zero, and thus all of Z_ξ has two-dimensional Lebesgue measure zero. But since $|E(z)| = |\lambda| \exp(\operatorname{Re} z)$, for every point z in I there must be an N such that $E^{\circ n}(z)$ belongs to Z_ξ for *all* n with $n \geq N$. Since I is a subset of $\bigcup_{n\geq 0} E^{\circ -n}(Z_\xi)$, it has measure zero for exponential maps $z \mapsto \lambda e^z$.

The situation is different for $E(z) = ae^z + be^{-z}$ with $ab \neq 0$: instead of throwing away half of the points in every step, we can "recycle" (in a literal sense) most of them: this time $|E^{\circ n}(z)| \to \infty$ implies that $|\operatorname{Re} E^{\circ n}(z)| \to \infty$. We use another parabola (or rather the complement thereof), namely,

$$P := \{x + iy \in \mathbb{C} \colon |y| < |x|^2\} \ .$$

If $z = x + iy$ with $|x|$ sufficiently large is such that $E(z)$ lies in P, then

$$|\operatorname{Re} E(z)| \geq |E(z)|^{1/2} \approx e^{|x|/2} \gg |x| \ ,$$

so points that escape to ∞ within P do so quite rapidly. On the other hand, the image of a square Q as in the first part (with real parts x greater than ξ or less than $-\xi$) is again an annulus, but the fraction of $E(Q)$ within P is approximately $1 - e^{-|x|/2}$, so most of the points survive the first step. Among these, a fraction of $1 - e^{-(e^{-|x|/2})/2}$ survives the second step, and so on. The total fraction of points within Q that "get lost" from P under iteration is less than 1, from which we infer that $I \cap Q$ has positive two-dimensional Lebesgue measure [**18**]. To be more precise, we recursively define a sequence (ξ_n) by $\xi_0 = \xi$ and $\xi_{n+1} = e^{|\xi_n|/2}$ for $n = 1, 2, \ldots$. Then $\xi_n > 2^{n-1}\xi_1$ for all n, provided that ξ_0 is sufficiently large. If

$$Q_n := \{z \in Q \colon E^{\circ k}(z) \in P \text{ for } k = 0, 1, 2, \ldots, n\} \ ,$$

then each z in Q_n has $|\operatorname{Re} E^{\circ n}(z)| > \xi_n$. This means that of all the points in Q_n, a fraction of at least $1 - e^{-\xi_n/2} = 1 - 1/\xi_{n+1}$ survives one more iteration within P. Thus, denoting two-dimensional Lebesgue measure by μ, we get

$$\frac{\mu(Q_n)}{\mu(Q)} > 1 - \frac{1}{\xi_1} - \frac{1}{\xi_2} - \ldots \frac{1}{\xi_n} > 1 - \frac{2}{\xi_1} = 1 - 2e^{-\xi_0} \ .$$

Since $\bigcap_n Q_n$ is contained in I, it follows that

$$\mu(I \cap Q) > (1 - 2e^{-\xi_0/2})\mu(Q) \ :$$

the set of escaping points has positive density in Q, hence I has positive (even infinite) two-dimensional Lebesgue measure.

In fact, we have shown much more: $\mu(Q \setminus I) < 2e^{-\xi_0/2}\mu(Q)$ [**26**]. Therefore, for each horizontal strip S of height 2π the complement of I in S has finite Lebesgue measure:

$$\mu\left(z \in S \setminus I\colon |\mathrm{Re}\, z| > \xi_0\right) < 2\pi \int_{\xi_0}^{\infty} 2e^{-x/2}\, dx = 4\pi e^{-\xi_0/2} \ .$$

This proves the result. □

Remark. Milnor [**20**, sec. 6] conjectured that for $f(z) = \sin z$ the set of points converging to the fixed point $z = 0$ has finite Lebesgue area in every strip $S' = \{z \in \mathbb{C}\colon \alpha \leq \mathrm{Re}\, z \leq \beta\}$. Since sine and hyperbolic sine represent the same map in rotated coordinate systems, this follows from Schubert's result.

We conclude with another special case in which the set I is so large that $\mathbb{C} \setminus I$ has measure zero [**24**]:

Corollary 2 (Escaping Set of Full Measure). *For maps $z \mapsto k\pi \sin z$ or $z \mapsto k\pi \sinh z$ with a nonzero integer k the set $I \cap E$ has full two-dimensional Lebesgue measure* (*i.e., the measure of $\mathbb{C} \setminus (I \cap E)$ is zero*).

Proof. We invoke a theorem of Bock [**3**]: for an arbitrary transcendental entire function at least one of the following two statements holds: (i) almost every orbit is dense in $\mathbb{C}$ or (ii) almost every orbit converges to ∞ or to one of the critical orbits (a critical orbit is the orbit of one of the two critical values $\pm k\pi i$). But since I has positive measure, case (i) cannot hold, so statement (ii) follows. For the map $E : z \mapsto k\pi \sinh z$ (or, equivalently, $z \mapsto k\pi \sin z$) the two critical values map to the fixed point 0. However, since $|E'(0)| > 1$ (i.e., the fixed point 0 is "repelling"), the only points whose orbits can converge to 0 are those countably many points that land exactly on 0 after finitely many iterations. Therefore, almost every orbit must escape. □

ACKNOWLEDGMENTS. I would like to express my gratitude to Bogusia Karpińska for many interesting discussions and for allowing me to include her example in section 4. I would also like to thank Cristian Leordeanu and Günter Rottenfußer for their help with the illustrations in this article.

References

[1] K. Barański, Trees and hairs for entire maps of finite order, (2005, preprint); available at `http://www.mimuw.edu.pl/~baranski/publ.html`.

[2] W. Bergweiler, Iteration of meromorphic functions, *Bull. Amer. Math. Soc. (N.S.)* **29** (1993) 151–188; updates available at `http://analysis.math.uni-kiel.de/bergweiler/bergweiler.engl.html`.

[3] H. Bock, On the dynamics of entire functions on the Julia set, *Results Math.* **30** (1996) 16–20.

[4] R. Devaney, Se^x: Dynamics, topology, and bifurcations of complex exponentials, *Topology Appl.* **110** (2001) 133–161.

[5] R. Devaney, L. Goldberg, and J. Hubbard, A dynamical approximation to the exponential map by polynomials, MSRI Preprint (1986).

[6] R. Devaney and M. Krych, Dynamics of $\exp(z)$, *Ergodic Theory Dynam. Systems* **4** (1984) 35–52.

[7] R. Devaney and F. Tangerman, Dynamics of entire functions near the essential singularity, *Ergodic Theory Dynam. Systems* **6** (1986) 489–503.

[8] A. Eremenko, On the iteration of entire functions, in *Dynamical Systems and Ergodic Theory*, Banach Center Publications, Polish Scientific Publishers, Warsaw, 1989, 339–345.

[9] K. Falconer, *Fractal Geometry. Mathematical Foundations and Applications*, John Wiley, Chichester, UK, 1990.

[10] P. Fatou, Sur l'itération des fonctions transcendantes entières, *Acta Math.* **47** (1926) 337–370.

[11] M. Förster, L. Rempe, and D. Schleicher, Classification of escaping exponential maps (2004, preprint); available at `http://arxiv.org/abs/math.DS/0311427`.

[12] J. G. Hocking and G. S. Young, *Topology*, Dover, Mineola, NY, 1988.

[13] B. Karpińska, Hausdorff dimension of the hairs without endpoints for $\lambda \exp(z)$, *C. R. Acad. Sci. Paris Sér. I Math.* **328** (1999) 1039–1044.

[14] J. Kotus and M. Urbański, Fractal measures and ergodic theory of transcendental meromorphic functions, *Transcendental Dynamics and Complex Analysis*, volume in honour of Professor I. N. Baker, LMS Lecture Note Series (to appear); available at `http://www.math.unt.edu/~urbanski`.

[15] D. G. Larman, A compact set of disjoint line segments in E^3 whose end set has positive measure, *Mathematika* **18** (1971) 112–125.

[16] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1983.

[17] Y. I. Manin, The notion of dimension in geometry and algebra, *Bull. Am. Math. Soc.* New Ser. **43** 2 (2006), 139–161.

[18] C. McMullen, Area and Hausdorff dimension of Julia sets of entire functions, *Trans. Amer. Math. Soc.* **300** (1987) 329–342.

[19] — — —, *Complex Dynamics and Renormalization*, Princeton University Press, Princeton, 1994.

[20] J. Milnor, *Dynamics in One Complex Variable. Introductory Lectures*, 2nd ed., Vieweg Verlag, Wiesbaden, 2000.

[21] G. Rottenfußer, J. Rückert, L. Rempe, D. Schleicher, *On questions of Fatou and Eremenko concerning escaping sets of entire functions*, Manuscript, in preparation.

[22] G. Rottenfußer, *On the Dynamical Fine Structure of Entire Transcendental Functions*, Ph.D. thesis, International University Bremen, 2005.

[23] G. Rottenfußer and D. Schleicher, Escaping points of the cosine family, *Transcendental Dynamics and Complex Analysis*, volume in honour of Professor I. N. Baker, LMS Lecture Note Series (to appear); available at `http://arxiv.org/abs/math.DS/0403012`.

[24] D. Schleicher, The dynamical fine structure of iterated cosine maps and a dimension paradox, *Duke Math. J.* (to appear); available at `http://arxiv.org/abs/math.DS/0406255`.

[25] D. Schleicher and J. Zimmer, Escaping points of exponential maps, *J. London Math. Soc.* (2) **67** (2003) 380–400.

[26] H. Schubert, *Über das Maß der Fatoumenge trigonometrischer Funktionen*, Diplomarbeit, Universität Kiel, 2003.

[27] G. Stallard, Dimensions of Julia sets of transcendental meromorphic functions, *Transcendental Dynamics and Complex Analysis*, volume in honour of Professor I. N. Baker, LMS Lecture Note Series (to appear).

[28] M. Urbański, Measures and dimensions in conformal dynamics, (2003, preprint); available at `http://www.math.unt.edu/~urbanski`.

Jacobs University Bremen (formerly: International University Bremen), Research I, Postfach 750 561, D-28725 Bremen, Germany, `dierk@jacobs-university.de`