THE ENTROPY OF A CERTAIN INFINITELY CONVOLVED BERNOULLI MEASURE

J. C. ALEXANDER AND DON ZAGIER

ABSTRACT

An entropy was introduced by A. Garsia to study certain infinitely convolved Bernoulli measures (ICBMs) μ_β, and showed it was strictly less than 1 for β the reciprocal of a Pisot–Vijayarghavan number. However, it is impossible to estimate values from Garsia's work. The first author and J. A. Yorke have shown this entropy is closely related to the 'information dimension' of the attractors of fat baker transformations T_β. When the entropy is strictly less than 1, the attractor is a type of strange attractor. In this paper, the entropy of μ_β is estimated for the case when $\beta = \phi^{-1}$, where ϕ is the golden ratio. The estimate is fine enough to determine the entropy to several decimal places. The method of proof is totally unlike usual methods for determining dimensions of attractors; rather a relation with the Euclidean algorithm is exploited, and the proof has a number-theoretic flavour. It suggests that some interesting features of the Euclidean algorithm remain to be explored.

1. *Introduction*

Consider the discrete probability density on the real line with measure $\frac{1}{2}$ at each of the two points ± 1. This is called the *Bernoulli measure*. We denote it by $b(x)$. Let $a_1, a_2, \ldots, a_N > 0$. We can form the convolution

$$b(x/a_1) * b(x/a_2) * \ldots * b(x/a_N).$$

This measure is supported on points

$$\sum_{n=1}^{N} \pm a_n. \tag{1.1}$$

The measure of each such point is 2^{-N} times the number of representations of the point as a sum (1.1). This measure is a finite probability measure and is called a finitely convolved Bernoulli measure. If $a_1, a_2, \ldots$ is an infinite sequence of positive numbers decreasing to zero sufficiently rapidly, the *infinitely convolved Bernoulli measure* (ICBM)

$$b(x/a_1) * b(x/a_2) * \ldots$$

can be formed as a weak limit of finitely convolved Bernoulli measures.

These measures were studied extensively in the 1930s, largely because their characteristic functions have interesting asymptotic properties. The ICBMs with $a_t = \beta^t$ for $\beta < 1$ fixed, were singled out. We denote this measure μ_β. For example, if $\beta = \frac{1}{2}$, the measure μ_β is uniform on the interval $[-1, 1]$. If $\beta = \frac{1}{3}$, the measure μ_β is the classical Cantor measure on $[-\frac{1}{2}, \frac{1}{2}]$. The following were proved. For $\beta < 1$,

Received 5 March 1990.

1980 *Mathematics Subject Classification* (1985 *Revision*) 28D20.

Supported in part by NSF.

the measures μ_β are well defined; that is, the limiting processes converge [**8**]. The μ_β are continuous and are either absolutely continuous on the interval $[-(1-\beta)^{-1}\beta, (1-\beta)^{-1}\beta]$ or totally singular [**8**]. For β the nth root of $\frac{1}{2}$, μ_β is absolutely continuous (and indeed, progressively smoother as n is increased) [**12**]. For $\beta < \frac{1}{2}$, it is easy to see that the support of μ_β is a Cantor set and hence μ_β is totally singular. However, for $\beta \geqslant \frac{1}{2}$, the support of μ_β is the interval $[-(1-\beta)^{-1}\beta, (1-\beta)^{-1}\beta]$, and it was generally supposed that these μ_β were absolutely continuous. Thus it was somewhat surprising when Paul Erdös proved that for $\beta = \phi^{-1}$, where $\phi = \frac{1}{2}(\sqrt{5}+1)$ is the golden ratio, the measure μ_β is totally singular [**4**]. A year later he proved a result in the opposite direction, namely there is $\gamma < 1$ such that for almost all $\beta > \gamma$, the measure μ_β is absolutely continuous [**5**]. Both of these results were proved by considering the asymptotics of the characteristic function. Erdös's method was shown to work whenever β is the reciprocal of a Pisot–Vijayarghyavan (PV) number, that is, whenever β is an algebraic integer and all of its conjugates lie outside the unit disk in the complex plane [**11**].

The question of which μ_β are totally singular for $\beta > \frac{1}{2}$ remains open. The only $\beta > \frac{1}{2}$ for which μ_β are known to be totally singular are the reciprocals of PV numbers, and the only explicit β for which μ_β is known to be absolutely continuous are roots of $\frac{1}{2}$ and a countable family of similar algebraic numbers [**6**]. For other work and expositions, see [**7, 9, 10, 11**].

Garsia introduced a new concept in the study of the μ_β [**7**]. He considered the finite measure spaces G_β^N of the finitely convolved Bernoulli measures

$$b(x/\beta) * b(x/\beta^2) * b(x/\beta^3) * \ldots * b(x/\beta^N)$$

for $N = 1, 2, \ldots$, and defined a limiting *entropy* as follows. Suppose that $x_1, x_2, \ldots, x_r$ are the points of G_β^N with measures $p_1, p_2, \ldots, p_r$. The (base 2) entropy of G_β^N is given by

$$H(G_\beta^N) = -\sum_{x_i \in G_\beta^N} p_i \log_2 p_i.$$

Then H_β is defined as the limit

$$H_\beta = \lim_{N\to\infty} \frac{H(G_\beta^N)}{\mathrm{N}\log_2 \beta},$$

which he showed exists (note that H_β is independent of the choice of base of logarithms). He also proved that if $H_\beta < 1$, then μ_β is totally singular. He also showed that $H_\beta < 1$ if β is the reciprocal of a PV number (this did not give a new proof that these μ_β are totally singular, since he used that fact in the argument). His argument is non-constructive and it is impossible to estimate H_β.

The purpose here is to determine an explicit value for H_β for one β, namely the one Erdös first considered, $\beta = \phi^{-1}$. There are three reasons why such a computation might be of interest.

1. It gives a direct proof that $H_\beta < 1$ for this β. It may be possible to generalize to other β (even perhaps β which are not the reciprocals of PV numbers?). The general approach is valid for any β which satisfies a polynomial equation with all coefficients 0 or 1 (which is the case for PV numbers and their reciprocals), and it seems likely the particulars will extend to the sequence of β which satisfy the equations

$$\beta^n + \beta^{n-1} + \beta^{n-2} + \ldots + \beta^2 + \beta - 1 = 0$$

(which are the reciprocals of PV numbers).

2. James Yorke and the first author introduced dynamical systems T_β on a square which are generalizations of the classical baker's transformation [1]. These transformations have natural invariant measures which are products of the μ_β with a uniform measure. We considered a certain dimension of the attractor of T_β, and showed that for β the reciprocal of a PV number, this dimension is $1+H_\beta$. In particular, since the dimension is not an integer, the attractor is some kind of strange attractor. The motivation for the present paper was to give an explicit computation of this dimension. In [1], an estimate for the dimension was given, but was not rigorous. In the present paper, a rigorous estimate is given. The present method is completely different from standard methods of estimating metric dimensions of strange attractors. Indeed, one of the conclusions of [1] is that standard methods cannot work very well for this particular map.

3. Since $\beta = \phi^{-1}$, the structure of the G_β^N is related to representations of integers in terms of Fibonacci numbers, and through them to the Euclidean algorithm. The computation involves generating functions related to the Euclidean algorithm. In the details, it becomes clear that there are some interesting new features of the Euclidean algorithm to be explored. We do not pursue more than we need in this paper.

For the remainder of the paper, we fix $\beta = \phi^{-1}$ and let $\Lambda = 1/\log_2 \phi$. We next briefly state the final results. Consider the *simple Euclidean algorithm*, that is, the Euclidean algorithm with a sequence of subtractions replacing the usual division. Given two positive integers k and i, the *length* $e(k,i)$ of the pair $\{k,i\}$ is the number of steps in the simple Euclidean algorithm applied to k and i (formally: $e(i,i) = 0$, $e(i+k,i) = e(i+k,k) = e(i,k)+1$).

THEOREM 1. *Let*

$$\kappa_n = \sum_{\substack{k>i>0 \\ \gcd(k,i)=1,\, e(k,i)=n}} k \log_2 k, \tag{1.2}$$

and

$$\mathcal{T}(x) = 1 - \frac{1}{2}\left(\frac{1-3x}{1-x}\right)^2 \sum_{n=1}^{\infty} \kappa_n x^n. \tag{1.3}$$

The series (1.3) *converges for* $|x| < \frac{1}{3}$ *and*

$$H_\beta = \Lambda \mathcal{T}(\tfrac{1}{4}) = \Lambda\left(1 - \frac{1}{18} \sum_{\substack{k>i>0 \\ \gcd(k,i)=1}} \frac{k \log_2 k}{4^{e(k,i)}}\right). \tag{1.4}$$

By determining the growth rate of the κ_n, and making rearrangements in $\mathcal{T}(x)$, we can estimate truncation errors. In particular, by truncating one rearranged series at the 20th term, we obtain the estimates

$$0.995570 < H_{\phi^{-1}} < 0.995736. \tag{1.5}$$

Another rearrangement gives sharper estimates; however, we have not been able to justify them rigorously. This series leads to the point estimate 0.995713126686.

We also show that (1.4) is equivalent to the following, somewhat different-looking, formula for H_β.

THEOREM 1′. *There is a unique bounded function $g(t)$ on $[0, 1]$ which satisfies the functional equation*

$$g(t) = (1+t)\log_2(1+t)+(2-t)\log_2(2-t)+\frac{1+t}{4}g\left(\frac{1}{1+t}\right)+\frac{2-t}{4}g\left(\frac{1}{2-t}\right). \quad (1.6)$$

This function is analytic on $(0, 1)$ and, moreover,

$$H_\beta = \Lambda\left(\tfrac{8}{9}-\tfrac{1}{36}g(\tfrac{1}{2})\right) = \Lambda\left(1-\tfrac{1}{24}g(0)\right). \quad (1.7)$$

It is possible to compute with (1.6), (1.7); the computations are identical to those coming from (1.3).

2. *The combinatorics*

The spaces G_β^N consist of sums

$$\sum_{n=1}^{N} \pm\beta^n. \quad (2.1)$$

Since $\beta^2+\beta-1 = 0$, there are multiple sums representing one point and of course this increases the measure at that point. For example $\beta^2+\beta-1$ and $-\beta^2-\beta+1$ both

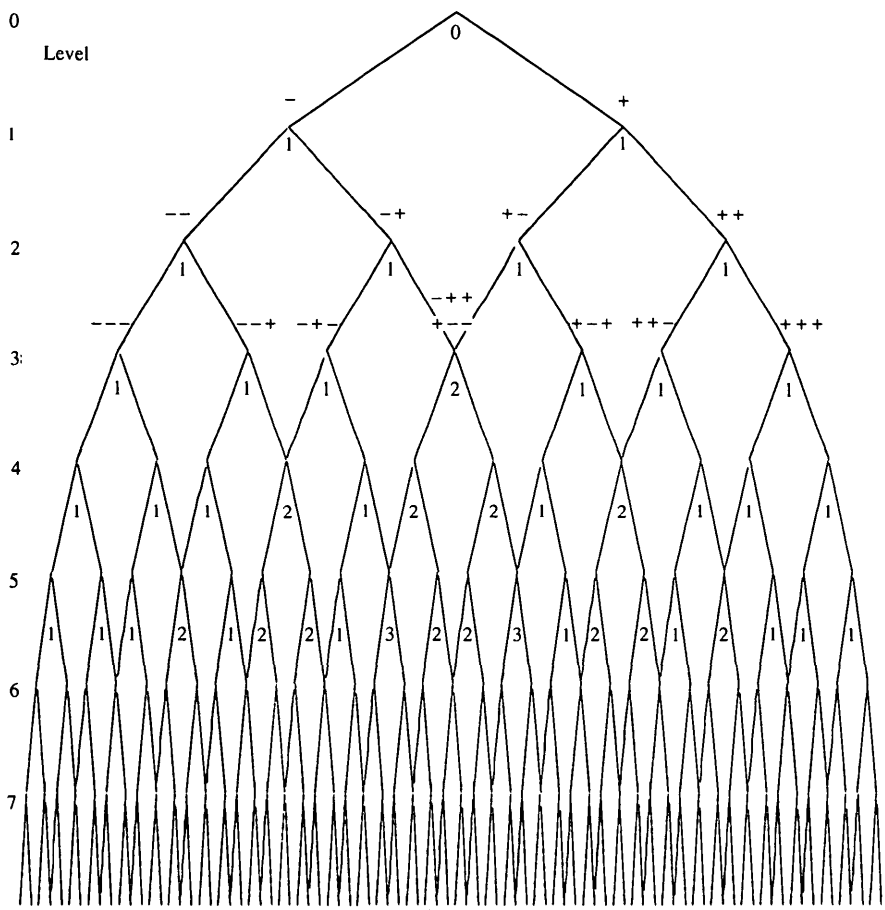

FIG. 1. The Fibonacci graph

represent 0. The whole structure of the G_β^N can be represented graphically as in Figure 1. The space G_β^N is represented by the nodes of the graph at level N, constructed from the top down. At level N, for each node, we consider the sequence of $\pm$ signs in (2.1) corresponding to that point. From each node, beginning with the node corresponding to the empty sequence at level 0, there are two descending edges. The left (respectively right) edge descends to a node corresponding to appending a $-$ (respectively $+$) sign to the sequence. Because of the relation $\beta^2+\beta-1=0$, the graph is not a tree. Above each node through level 3, we list the sequence(s) corresponding to each node. Below the nodes through level 5, we list the number of sequences corresponding to that node. We call this number the *frequency*. It is equal to the number of descending paths from the top node to the node in question. Moreover, at level N, 2^{-N} times the frequency is the measure of the point in G_β^N. Thus the entropies $H(G_\beta^N)$ can be computed from the frequencies. Let $F_N(k)$ denote the number of nodes at level N with frequency k. Then

$$H(G_\beta^N) = -\sum_k F_N(k)\,2^{-N}\log_2 2^{-N}k = N-\sum_k F_N(k)\,2^{-N}\log_2 k. \tag{2.2}$$

In fact, there is another simple interpretation of this graph and the entropies we consider. Let f_i be the ith element of the Fibonacci sequence 1, 2, 3, 5, 8,.... A *Fibonacci expansion* of length N of a non-negative integer n is a representation $n=\sum_{i=1}^N a_i f_i$, $a_i=0,1$. The graph of Figure 1 encodes Fibonacci representations of the non-negative integers. If the nodes at level N are labelled from 0 to $f_{N+2}-2$ from left to right, the frequency associated to a node is the number of Fibonacci expansions of length N of the label. For this reason, we call the graph of Figure 1 the *Fibonacci graph*. These frequencies have been studied; see for example [**2**]. The entropy $H(G_\beta^N)$ measures the average redundancy of the Fibonacci representations of length N. There is obvious redundancy since 2^N words on symbols zero and one represent $f_{N+2}-1$ numbers. If, asymptotically, the redundancy were more or less evenly distributed, the entropies $H(G_\beta^N)$ would be asymptotic to $N\log_2\beta+o(N)$. The fact that H_β is strictly less than one implies that the redundancies in the representation are not evenly distributed in some strong sense.

There is considerable and repeated structure in the Fibonacci graph. In particular, suppose we erase from that graph the node at level 0 and all the nodes with frequency 1 and the edges descending from those nodes. The remainder is the disjoint union of an infinite number of isomorphic graphs. Each of these has a 'top' node with frequency 2 in the lowest level, two nodes with frequency 2 on the next level, two with frequency 2 and two with frequency 3 two levels beyond, etc. There is one such top at level 3, and two such tops at each succeeding level. Moreover, these subgraphs themselves break apart into more basic subgraphs. In this section, we expose this decomposition and develop the structure of the most basic subgraphs. To this end we proceed through a succession of technical definitions and observations, some labelled for future reference. The reader may find it convenient to illustrate the statements with Figure 1, concentrating on the central node at level 3.

Any edge connects two nodes, one at level N and one at level $N+1$. We call the one at level N the *parent* of the one at level $N+1$ and the one at level $N+1$ the *child* of the one at level N. We extend the terminology to speak of grandchildren, etc. The *descendants* of a node are its children, their children, etc. The *ancestors* of a node are its parents, grandparents, etc.

The graph divides the plane into bounded regions we call *diamonds*. The diamonds are consequences of the relation $-1+\beta+\beta^2 = 1-\beta-\beta^2$ (or $-++ = +--$ in terms of plus and minus signs). Each diamond is bounded by six edges and six nodes. The nodes are at levels N, $N+1$, $N+2$, and $N+3$ for some N. There is one *top* node at level N. There are two nodes at level $N+1$, two at level $N+2$, and a *bottom* node at level $N+3$. The bottom node we call a *relational node*, since its position at the bottom is a consequence of the relation above. The node at the top is called the *relational parent* of the node at the bottom and the node at the bottom is called the *relational child* of the node at the top. A relational node has two parents; others have one.

In addition to a frequency d, a relational node X has a *frequency pair*. A relational node has two parents with frequencies a and b. The frequency pair of X is the unordered pair $\{a, b\}$. Clearly the frequency of X is $a+b$. Finally, for purposes of setting up inductions, we assign a *depth* to each node. This is done inductively. The node at level 0 and all nodes with frequency 1 are assigned depth 0. If a node is not relational, it is assigned the depth of its parent. If it is relational, it is assigned depth 1 plus the depth of its relational parent.

There are four edges emanating from a relational node; hence a relational node is in the boundary of four diamonds—one above, one below, one left, and one right. (a) If the top node of a diamond is at level N, neither of the boundary nodes at level $N+2$ is relational. (b) Moreover, at most one of the nodes at level $N+1$ is relational, since there is not enough room for two diamonds above two relational nodes in those positions. On the other hand, a non-relational node which is not at the extreme left or right of a level is in the boundary of three diamonds. It is the top of one diamond, at level 1 below the top of another diamond, and at level 2 below the top of a third. (c) Accordingly one of the children of a non-relational node is relational and one is non-relational; the non-relational child has the same frequency and same depth as its parent.

We next consider the structure of the graph below a relational node X. Say X has frequency pair $\{a, b\}$, level N, and depth r. There are two children at level $N+1$, neither of which is relational by statement (a) above. Accordingly, these two nodes have frequency $a+b$ and depth r. There are four grandchildren at level $N+2$. Two of these are relational, being the relational children of the parents of X, and two are not. The two which are not have frequency $a+b$ and depth r. The two which are relational have frequency pairs $\{a, a+b\}$ and $\{b, a+b\}$ and depth r, since their relational parents are the parents of X. Finally the relational child of X has frequency pair $\{a+b, a+b\}$ and depth $r+1$. In the Fibonacci graph, all nodes through level 6 have depth 0 or 1 except the central node at level 6 which has depth 2. At level 7 there are four nodes of depth 2.

The preceding paragraph describes the local structure near a relational node. We turn to the global structure. (d) First note that from every relational node, two edge paths descend; every node on these two paths has the same depth and the same frequency as the parent relational node. This follows from (a) and (c) above. Consider a relational node X of depth r both of whose parents have depth less than r. Call such a node a *primary* relational node. In the Fibonacci graph, there is one primary relational node of depth 1 at level 3, and two at each succeeding level. There is a primary relational node of depth 2 at level 6. There are two at level 7 and four at level 8. Consider the set $D(X)$ of all descendants of X which are relational and have depth r. Since each relational node has two relational grandchildren, if we insert edges in

$D(X)$ between each node and its two relational grandchildren, we make $D(X)$ into a binary tree. Note that if X and Y are both relational nodes, $D(X) \cap D(Y) = \emptyset$. This follows from statements (b) and (c) above. We next note that the frequency pair of X has the form $\{d, d\}$, where $d = 1$ if the depth of X is 1, or d is the frequency of the relational parent of X if the depth of X is greater than 1. The frequency pair of X determines the frequency pairs of all members of $D(X)$, as we study next.

The *simple Euclidean algorithm* is the Euclidean algorithm without division. We can describe one step of the algorithm as follows. For any pair of non-negative integers $\{k, i\}$ (possibly equal), let $T\{k, i\} = \{|k-i|, \min(k, i)\}$. The *length* of $\{k, i\}$ is the greatest n such that $0 \notin T^n\{k, i\}$. For example, if we use $\mapsto$ to denote the effect of T,

$$\{11,3\} \mapsto \{8,3\} \mapsto \{5,3\} \mapsto \{3,2\} \mapsto \{2,1\} \mapsto \{1,1\} \mapsto \{1,0\}$$

(length 5). The length is denoted by $e(k, i)$. This definition is clearly equivalent to the one in the introduction.

Conversely, we can make a binary tree labelled with pairs of integers as follows. Start with one node at level 0 labelled with the pair $\{1, 1\}$ and one node at level 1 labelled with the pair $\{2, 1\}$. Inductively, given a node at level n labelled with the pair $\{a, b\}$, there are two descending edges (left and right) top nodes at level $n+1$ labelled with pairs $\{a+b, a\}$ and $\{a+b, b\}$. At level n, there are 2^{n-1} nodes. For example, at level 3, the pairs are $\{5,3\}$, $\{5,2\}$, $\{4,3\}$, $\{4,1\}$. Start with any pair $\{a, b\}$ at level n. The labels on the nodes of the unique path up the tree to the node at level 1 are precisely those of the simple Euclidean algorithm of $\{a, b\}$, up to the last step. The length of the pair is n. Conversely, given the expansion of the simple Euclidean algorithm for any pair of coprime integers $\{a, b\}$, it is routine to locate the pair in the tree. Thus each pair of coprime integers $\{a, b\}$ appears exactly once in this tree at level $e(a, b)$. We call this tree the *Euclidean tree*. This tree has appeared in other contexts, see for example [**3**, **13**]. We trivially form the *d-Euclidean tree* by multiplying all the labels involved in the Euclidean tree by d.

We next observe that for any primary relational node X in the Fibonacci graph with frequency pair $\{d, d\}$, the labelled tree structure of $D(X)$ is a d-Euclidean tree. This is immediate from the way frequency pairs of relational descendants are determined. Thus the Euclidean tree is the irreducible kernel of the Fibonacci graph and, in the end, all computations will be done in terms of the Euclidean graph. Note that if a node at level $L \geqslant 1$ is labelled with $\{a, b\}$ with $a > b$ in the Euclidean tree, then a is the sum of the two labels of the parent of the node. Thus it is essentially equivalent to consider the maxima of the two integers of a label or the sum of the labels. For minor technical reasons, we work below with the former.

We can reverse the above process and, starting from the Euclidean tree, build the nodes of the Fibonacci graph. We construct a set of nodes; each node is assigned a level, a frequency, and a depth. We can speak of *appending a d-Euclidean tree of depth r, starting at level N* to an existing set. By this we mean making a disjoint union of the existing set and the nodes of levels $L \geqslant 1$ of a copy of a d-Euclidean tree, so that if a node in the d-Euclidean tree has length L, it is assigned level $N+2L$ and depth r in the set. If the node in the Euclidean tree is labelled with integers $\{a, b\}$, with $a > b$, it is given frequency a in the set.

The discussion above proves the following lemma.

LEMMA. *The following process constructs a set of points (nodes) labelled with levels, frequencies and depths, which is isomorphic to the set of nodes of the Fibonacci graph in levels greater than* 1.

(i) *Begin with two nodes of frequency* 1 *and depth* 0 *at each level* $N \geqslant 1$.

(ii) *Inductively append Euclidean trees to obtain relational nodes.*

(a) *Append one* 1*-Euclidean tree starting at level* 3 *and two* 1*-Euclidean trees starting at each level* $N > 3$.

(b) *Proceed by induction on depth. Assume that nodes up through depth r have been assigned. For each node of depth r which has, say, level N and frequency d, append a 2d-Euclidean tree of depth* $r+1$ *starting at level* $N+3$.

(iii) *Append non-relational nodes. For each node of depth r, frequency d, and level N, append two nodes at each level greater than N of depth r and frequency d.*

As a consequence of this result, many of the results about Fibonacci representations [2] can be deduced from known results about the Euclidean algorithm. Particulars are left to the interested reader.

3. *Generating functions*

In this section we translate the construction of the previous lemma into the language of generating functions. The object of the present section is to prove the following result.

PROPOSITION. *Let*

$$\mathscr{H}(x) = \sum_{N=0}^{\infty} H(G_\beta^N)\, x^N$$

be the generating function for the entropies of the finite spaces G_β^N. *Then*

$$\mathscr{H}(x) = \frac{x}{(1-x)^2}\,\mathscr{T}\left(\frac{x^2}{4}\right) \tag{3.1}$$

with $\mathscr{T}$ *as in* (1.3).

Proof. Let $F_N(k)$ be equal to the number of nodes at level N with frequency k. Note that

$$\sum_{k=1}^{\infty} F_N(k) = f_{N+2} - 1, \qquad \sum_{k=1}^{\infty} kF_N(k) = 2^N, \qquad \sum_{k=1}^{\infty} \frac{k}{2^N} F_N(k) \log_2 \frac{2^N}{k} = H(G_\beta^N).$$

Let $f_k(x) = \sum_{N=1}^{\infty} F_N(k)\, x^N$ (this and all further generating functions are to be considered formal power series in x, although all in fact converge for $|x|$ small enough). Let $\tilde{\alpha}_k(N)$ be the number of times the integer k appears as the larger of the two integers in the pairs associated to nodes of length N in the Euclidean tree, and let $\hat{\alpha}_k(x) = \sum_{N=1}^{\infty} \tilde{\alpha}_k(N)\, x^N$. From the description of the Euclidean tree in the previous section, it is apparent that

$$\hat{\alpha}_k(x) = \sum_{\substack{1 \leqslant t \leqslant k \\ \gcd(k,t)=1}} x^{e(k,t)}.$$

When the nodes of Euclidean trees are embedded in the nodes of the Fibonacci graph,

TABLE 1. *Components of generating functions*

k	$\alpha_k(x)$	$l_k(x)$
1	1	1
2	x^3	x^3
3	$2x^5$	$2x^5$
4	$2x^7$	x^6+2x^7
5	$2x^7+2x^9$	$2x^7+2x^9$
6	$2x^{11}$	$4x^8+2x^{11}$
7	$4x^9+2x^{13}$	$4x^9+2x^{13}$
8	$2x^9+2x^{15}$	$3x^9+4x^{10}+2x^{15}$
9	$4x^{11}+2x^{17}$	$4x^{10}+4x^{11}+2x^{17}$

the levels of the nodes come from the lengths, up to a shift and a stretch; this leads us to define

$$\alpha_k(x) = \sum_{i=1}^{\infty} x^{1+2e(k,i)}.$$

Starting from $l_1(x) = 1$, inductively define

$$l_k(x) = \sum_{\substack{d\mid k\\ d\neq 1}} \alpha_d(x)\, l_{k/d}(x)$$

(a short list of $\alpha_k(x)$ and $l_k(x)$ is given in Table 1). Next, define functions of two variables x and s:

$$\mathscr{L}(x;s) = 1+\sum_{k=2}^{\infty} k^s l_k(x),$$

$$\mathscr{A}(x;s) = 1-\sum_{k=2}^{\infty} k^s \alpha_k(x),$$

$$\Phi(x;s) = 1+\sum_{k=2}^{\infty} k^s f_k(x).$$

The formulae above imply that

$$\mathscr{L}(x;s) = \mathscr{A}(x;s)^{-1}.$$

Note also that

$$\Phi(x;0) = \sum_{N=0}^{\infty} (f_{N+2}-1)\, x^N = \frac{1}{(1-x)(1-x-x^2)},$$

$$\Phi(x;1) = \sum_{N=0}^{\infty} 2^N x^N = \frac{1}{1-2x},$$

$$\left.\frac{\partial\Phi(x;s)}{\partial s}\right|_{s=1} = \sum_{\substack{k\geqslant 1\\ N\geqslant 1}} kF_N(k)\ln k x^N,$$

so that

$$\mathscr{H}(x) = \sum_{N=0}^{\infty} H(G_\beta^N)\, x^N = \frac{x}{(1-x)^2} - \frac{1}{\ln 2}\left.\frac{\partial\Phi(x/2;s)}{\partial s}\right|_{s=1}.$$

By induction, it is seen that the sum of the maxima of the pairs associated to the nodes of length N in the Euclidean tree is $2\cdot 3^{N-1}$. Thus

$$\mathscr{A}(x;1) = 1-\sum_{\substack{k>i>0\\ \gcd(k,i)=1}} kx^{1+2e(k,i)} = 1-2x\sum_{N=1}^{\infty} 3^{N-1}x^{2N} = \frac{(1+x)^2(1-2x)}{1-3x^2}.$$

The lemma of the previous section leads to the expression

$$f_k(x) = \begin{cases} 2x/(1-x)^2 & \text{if } k = 1, \\ \left(\dfrac{1+x}{1-x}\right)^2 l_k(x) & \text{if } k > 1. \end{cases}$$

To see this, note that $l_k(x)$ is the generating function for the set constructed in the induction (ii)(b) of the lemma, starting from a single 1-Euclidean tree starting at level 0. By (ii)(a), the generating function for the set constructed in (ii) is

$$(1+2x+2x^2+ +2x^3+\dots)\, l_k(x) = \frac{1+x}{1-x} l_k(x).$$

By part (iii) of that lemma, to obtain the generating function for $k > 1$, this expression is multiplied by another $(1+x)/(1-x)$. The count for $k = 1$ is handled separately, giving the above expression. Hence

$$\Phi(x;s) = \frac{1+x^2}{(1-x)^2} + \left(\frac{1+x}{1-x}\right)^2 \sum_{k=2}^{\infty} k^s l_k(x) = \left(\frac{1+x}{1-x}\right)^2 \mathscr{L}(x;s) - \frac{2x}{(1-x)^2}.$$

Thus

$$\left.\frac{\partial\Phi(x;s)}{\partial s}\right|_{s=1} = \left(\frac{1+x}{1-x}\right)^2 \left.\frac{\partial\mathscr{L}(x;s)}{\partial s}\right|_{s=1} = -\left(\frac{1+x}{1-x}\right)^2 \mathscr{A}(x;1)^{-2} \left.\frac{\partial\mathscr{A}(x;s)}{\partial s}\right|_{s=1}$$

$$= -\frac{(1-3x^2)^2}{(1-x^2)^2(1-2x)^2} \left.\frac{\partial\mathscr{A}(x;s)}{\partial s}\right|_{s=1}.$$

On the other hand

$$\left.\frac{\partial\mathscr{A}(x;s)}{\partial s}\right|_{s=1} = -\sum_{\substack{k>t>0 \\ \gcd(k,t)=1}} x^{2e(k,t)+1} k \ln k = \sum_{N=1}^{\infty} \Big(\sum_{\substack{k>t>0 \\ \gcd(k,t)=1,\, e(k,t)=N}} k \ln k \Big) x^{2N+1}.$$

Hence

$$\mathscr{H}(x) = \frac{x}{(1-x)^2} - \frac{(4-3x^2)^2}{(4-x^2)^2(1-x)^2} \sum_{N=1}^{\infty} \Big(\sum_{\substack{k>t>0 \\ \gcd(k,t)=1,\, e(k,t)=N}} k \log_2 k \Big) \left(\frac{x}{2}\right)^{2N+1}$$

$$= \frac{x}{(1-x)^2} \mathscr{T}\left(\frac{x^2}{4}\right).$$

This completes the proof of the proposition.

The series (1.3), as written, converges too slowly at $t = \frac{1}{4}$ for useful estimates. It is not hard to see that $2 \cdot 3^{N-1} \log_2(N+1) < \kappa_N < 2 \cdot 3^{N-1} N \log_2 \phi$; that is, the κ_n grow faster than geometrically. It is possible to rearrange the series to obtain coefficients which grow geometrically. Accordingly we define coefficients μ_N and λ_N by the formulae

$$\sum_{N=1}^{\infty} \mu_N x^N = \tfrac{1}{2}(1-3x) \sum_{N=1}^{\infty} \kappa_N x^N, \tag{3.2}$$

$$\sum_{N=1}^{\infty} \lambda_N x^N = \frac{1}{2}\left(\frac{1-3x}{1-x}\right)^2 \sum_{N=1}^{\infty} \kappa_N x^N = 1 - \mathscr{T}(x). \tag{3.3}$$

We can put useful rigorous bounds on the μ_N; the λ_N are used to make sharper, but non-rigorous, estimates.

LEMMA.

$$0.585\ldots = \log_2 1.5 < \frac{u_N}{3^{N-1}} < \tfrac{2}{3} \tag{3.4}$$

(*in particular the* μ_N *are positive and grow geometrically*).

Proof. Consider a node of length N in the Euclidean tree associated to the pair $\{a, b\}$, $a > b$. It has a 'sibling' pair $\{a, a-b\}$ (both descending from the pair $\{a-b, a\}$. These two spawn pairs labelled

$$\{a+b, a\}, \quad \{a+b, b\}, \quad \{2a-b, a\}, \quad \{2a-b, a-b\}$$

of length $N+1$. Accordingly $\mu_{N+1} = \frac{1}{2}(\kappa_{N+1} - 3\kappa_N)$ can be written

$$\sum_{\substack{a>b>0 \\ (a,b)=1,\, e(a,b)=N}} \tfrac{1}{2}((a+b)\log_2(a+b) + (2a-b)\log_2(2a-b) - 3a\log_2 a)$$

$$= \sum_{\substack{a>b>0 \\ (a,b)=1,\, e(a,b)=N}} \tfrac{1}{2}a\left[\frac{a+b}{a}\log_2\left(\frac{a+b}{a}\right) + \frac{2a-b}{a}\log_2\left(\frac{2a-b}{a}\right)\right].$$

By the convexity of the function $x \mapsto x\log_2 x$ for $x > 0$, this expression is greater than

$$\sum_{\substack{a>b>0 \\ (a,b)=1,\, e(a,b)=N}} \frac{3a}{2}\log_2 1.5 = 3^N \log_2 1.5.$$

This proves the first inequality of (3.4). On the other hand, the function

$$x \longmapsto x\log_2 x + (3-x)\log_2(3-x)$$

is convex on the interval $[1, 2]$ and thus takes its maximum at one (actually both) of its end points. Letting $x = (a+b)/a$, we find that μ_{N+1} is bounded by

$$\sum_{\substack{a>b>0 \\ (a,b)=1,\, e(a,b)=N}} a\log_2 2 = 2\cdot 3^{N-1}.$$

This proves the second inequality of (3.4) and completes the proof of the lemma.

The behaviour of $\mathscr{H}(x)$ as a meromorphic function depends on the rate of growth of the coefficients. From (3.4) we see that $\sum_{N=1}^{\infty} \mu_N x^N$ is dominated by $2x/3(1-3x)$, so that (3.2) converges for $|x| < \frac{1}{3}$. Consequently, $\mathscr{H}(x)$ is holomorphic for $|x| < \sqrt{\frac{4}{3}}$ except for a double pole at $x = 1$. This is consistent with the known rate of growth of $\mathscr{H}(G_\beta^N)$. Thus

$$\mathscr{H}(x) = H_{-2}\frac{x}{(1-x)^2} + \frac{H_{-1}}{1-x} + O(1), \quad |x| < \sqrt{\tfrac{4}{3}} \tag{3.5}$$

for some numbers H_{-2} and H_{-1}. Equivalently, as $N \to \infty$,

$$H(G_\beta^N) = NH_{-2} + H_{-1} + O((\tfrac{3}{4})^{N/2}). \tag{3.6}$$

Note, then, that

$$H_\beta = \Lambda H_{-2}. \tag{3.7}$$

4. *Numerical computations*

In this section we prove Theorem 1 and make numerical computations. We present some tables and derive the estimates (1.5). Recall that $\Lambda = 1/\log_2 \phi$.

To obtain (1.4), we multiply equation (3.1) by $(1-x)^2$ to obtain

$$(1-x)^2\mathscr{H}(x) = x\mathscr{T}(\tfrac{1}{4}x^2) = H_{-2} + (1-x)H_{-1} + (1-x)^2O(c^{-n}).$$

This is convergent at $x = 1$, so setting $x = 1$, we obtain from (3.7) that

$$H_\beta = \Lambda\mathscr{T}(\tfrac{1}{4}). \tag{4.1}$$

Note that since $\mathscr{T}$ converges for $x < \sqrt{\tfrac{1}{3}}$, this series converges. We can evaluate it in several ways, depending on how we expand $\mathscr{T}$. Thus we obtain

$$H_\beta = \Lambda\left(1 - \frac{1}{18}\sum_{n=1}^{\infty}\frac{\kappa_n}{4^n}\right), \tag{4.2}$$

$$= \Lambda\left(1 - \frac{4}{9}\sum_{n=1}^{\infty}\frac{\mu_n}{4^n}\right), \tag{4.3}$$

$$= \Lambda\left(1 - \sum_{n=1}^{\infty}\frac{\lambda_n}{4^n}\right). \tag{4.4}$$

The series (4.2) converges too slowly for effective computation. We use series (4.3). The partial sums of the series are exhibited in Table 2. Since each $\mu_n > 0$, the values

TABLE 2. *Evaluation of entropies at levels N from equation* (4.3)

N	κ_N	$\frac{4}{9}4^{-N}\mu_N$	$\Lambda[1-\frac{4}{9}\sum_{n=1}^{N}4^{-n}\mu_n]$
1	2.00000000000	0.1111111111	1.2803734137
2	9.50977500433	0048746875 1	1.2101574355
3	39.2192809489	0.0371179026	1.1566920630
4	149.825218767	0.0279230694	1.1164711129
5	546.040935867	0.0209560069	1.0862856596
6	1927.86131506	0.0157193201	1.0636432351
7	6652.82922633	0.0117898937	1.0466608354
8	22566.2448165	0.0088424925	1.0339239316
9	75522.0214340	0.0066318825	1.0243712348
10	250035.936861	0.0049739144	1.0172067086
11	820517.437055	0.0037304362	1.0118333133
12	2672781.19736	0.0027978273	1.0078032667
13	8652030.25592	0.0020983705	1.0047807317
14	27857150.7635	0.0015737779	1.0025138304
15	89274632.2811	0.0011803334	1.0008136545
16	284933436.818	0.0008852500	0.9995385225
17	906128930.379	0.0006639375	0.9985821736
18	2872372650.92	0.0004979532	0.9978649119
19	9079075532.09	0.0003734649	0.9973269656
20	28623099334.3	0.0002800986	0.9969235058

in the third column are upper bounds for H_β; hence H_β is clearly seen to be less than 1. However, we can bound truncation errors in H_β and thus compute. The following result is immediate from (3.4).

LEMMA. *If* (4.3) *is truncated at term N, the error E_N is bounded by*

$$\tfrac{1}{9}(\tfrac{3}{4})^{N-1}\Lambda\log_2 1.5 < E_N < \tfrac{2}{9}(\tfrac{3}{4})^{N-1}\Lambda. \tag{4.5}$$

These bounds combined with the computations of Table 3 lead to the bounds (1.5).

It is very interesting to consider the computations from (4.4). This series seems to converge much more rapidly than (3.2); the results are exhibited in Table 3. There are

TABLE 3. *Estimates using series* (4.4)

N	λ_N	$\Lambda[1-\sum_{n=1}^{N}4^{-n}\lambda_n]$	N	λ_N	$\Lambda[1-\sum_{n=1}^{N}4^{-n}\lambda_n]$
1	1.000 000 000 000 00	1.080 315 067 809 42	11	0.167 131 013 5	0.995 713 143 465 75
2	0.754 887 502 163 47	1.012 355 372 552 07	12	0.150 897 353 0	0.995 713 130 510 34
3	0.590 090 465 782 94	0.999 074 463 769 65	13	0.137 306 234	0.995 713 127 563 21
4	0.474 047 485 496 02	0.996 407 168 762 16	14	0.125 804 500	0.995 713 126 888 15
5	0.389 580 409 351 5	0.995 859 161 488 27	15	0.115 974 59	0.995 713 126 732 57
6	0.326 447 711 849 7	0.995 744 361 233 72	16	0.107 497 31	0.995 713 126 696 52
7	0.278 194 391 272	0.995 719 903 421 82	17	0.100 125 7	0.995 713 126 688 12
8	0.240 588 130 815	0.995 714 615 519 34	18	0.093 666 6	0.995 713 126 686 16
9	0.210 767 793 08	0.995 713 457 399 51	19	0.087 967	0.995 713 126 685 70
10	0.186 750 168 90	0.995 713 200 862 37	20	0.082 907	0.995 713 126 685 59

a couple of surprises in this table. The λ_n are obtained by adding and subtracting large numbers. There is no reason to expect them to be positive and decreasing to zero, although the table indicates that they are. We can prove only that λ_n is $O((\frac{4}{3})^{n/2})$. There is evidently something more subtle going on in the Euclidean algorithm. If the λ_n are bounded, the data in Table 3 converge at a rate no worse than $O(4^{-N})$. From the tabulations, the data are converging faster than $O(4.5^{-N})$. Equivalently, $\sum \lambda_n x^n$ converges for $|x| < C$ for some $C \sim 1.1$. This is surprising since, from the form of (3.3), $\sum \lambda_n x^n$ could have a pole at $x = 1$. We leave the investigation of this series to the interested reader. Other calculations (higher precision) indicate that the exhibited digits of Table 3 are uncontaminated by machine roundoff error. If the indicated convergence is valid, the value of H_β is most likely between 0.995 713 126 685 and 0.995 713 126 686.

5. *A functional relation*

In this section we prove Theorem 1′. Define a sequence of functions $g_n(t)$ $(n = 1, 2, \ldots)$ by

$$g_1(t) = (1+t)\log_2(1+t)+(2-t)\log_2(2-t),$$

$$g_{n+1}(t) = 3^n g_1(t)+(1+t)\,g_n\left(\frac{1}{1+t}\right)+(2-t)\,g_n\left(\frac{1}{2-t}\right) \qquad (n \geqslant 1).$$

An easy induction shows that

$$g_n(t) = \sum_{\substack{k,i>0 \\ \gcd(k,i)=1,\, e(k,i)=n}} [kt+i(1-t)]\log_2[kt+i(1-t)].$$

Hence each g_n is analytic on the set $T = \mathbb{C}\backslash([1,\infty) \cup (-\infty, 0])$, and the identity $\sum_{k,i}(k+i) = 2\cdot 3^n$ implies that the series

$$g(t,x) = \sum_{n=1}^{\infty} g_n(t)\, x^n$$

converges absolutely and locally uniformly in $T \times \{x : |x| < \frac{1}{3}\}$ and therefore defines a jointly analytic function of t and x there. The recurrence defining the g_n translates into the functional equation

$$g(t,x) = \frac{x}{1-3x} g_1(t) + x(1+t)\, g\left(\frac{1}{1+t}, x\right) + x(2-t)\, g\left(\frac{1}{2-t}, x\right).$$

Conversely, it is easily seen by the estimate just used that this functional equation has a unique solution whenever $|x| < \frac{1}{3}$ (the crucial point is that then the sum of the factors $x(1+t)$ and $x(2-t)$ is uniformly bounded by a constant less than 1). For $x = \frac{1}{4}$ the functional equation for $g(t,x)$ reduces to (1.6), so the uniqueness implies that $g(t) = g(t, \frac{1}{4})$. On the other hand, the closed formula for $g_n(t)$ implies that $g_n(\frac{1}{2}) = \frac{1}{2}\kappa_{n+1} - 3^n$ for every $n \geqslant 1$ with κ_n as in (1.2). Thus (1.7) is indeed equivalent to (1.4), as asserted.

References

1. J. C. ALEXANDER and J. A. YORKE, 'Fat baker's transformations', *Ergod. Theory Dynam. Systems* 4 (1984) 1-23.
2. L. CARLITZ, 'Fibonacci representations', *Fibonacci Quart.* 6 (1968) 193–220.
3. H. COHN, 'Growth types of Fibonacci and Markoff', *Fibonacci Quart.* 17 (1979) 178–183.
4. P. ERDÖS, 'On a family of symmetric Bernoulli convolutions', *Amer. J. Math.* 61 (1939) 974–976.
5. P. ERDÖS, 'On the smoothness properties of a family of Bernoulli convolutions', *Amer. J. Math.* 62 (1940) 180–186.
6. A. M. GARSIA, 'Arithmetic properties of Bernoulli convolutions', *Trans. Amer. Math. Soc.* 102 (1962) 409–432.
7. A. M. GARSIA, 'Entropy and singularity of infinite convolutions', *Pacific J. Math.* 13 (1963) 1159–1169.
8. B. JESSEN and A. WINTNER, 'Distribution functions and the Riemann zeta function', *Trans. Amer. Math. Soc.* 38 (1938) 48–88.
9. J. P. KAHANE, 'Sur la distribution de certaines series aleatoires', *Bull. Soc. Math. France* 25 (1971) 119–122.
10. T. KAWATA, *Fourier analysis in probability theory* (Academic Press, London, 1972).
11. R. SALEM, *Algebraic numbers and Fourier analysis* (Heath, Lexington, 1963).
12. A. WINTNER, 'On convergent Poisson convolutions', *Amer. J. Math.* 57 (1935) 827–838.
13. D. ZAGIER, 'On the number of Markoff numbers below a given bound', *Math. Comp.* 39 (1982) 709-723.

Department of Mathematics
University of Maryland
College Park
Maryland 20742
USA