The Entropy of the Fibonnaci Code

J. C. Alexander*
Department of Mathematics
University of Maryland
College Park, MD 20742

February, 1989

* Supported in part by N.S.F.

1. Introduction.

The Fibonacci sequence

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

defined by

$$f_0 = 1, \quad f_1 = 1, \quad f_r = f_{r-1} + f_{r-2} \text{ for } r \geq 2,$$

is certainly one of the best known sequences of mathematics. In this paper we consider its use as a basis for representing integers. For any string $a_n a_{n-1} \cdots a_n$ of zeroes and ones, the integer represented by the string is $\sum_{r=1}^{n} a_r f_r$. For example, 010110 represents $2 + 3 + 8 = 13$. We call such a representation a *Fibonacci representation.* We consider this representation as a code and ask for the value of the information-theoretic entropy of this code, especially its asymptotic (in n) behavior. The question is phrased more precisely in the next section.

Some integers have more than one such representation — for example 13 can also be written 100000. There have been a number of investigations of the number of Fibonacci representations for integers; see for example Carlitz [1968]. Entropy measures in some sense the lumpiness of the representation — whether the integers are more or less uniformly represented or whether some numbers have many more representations than others. If Fibonacci representations are considered as a code, asking for the value of the entropy is rather natural, yet it seems not to have been investigated. In fact, the entropy is asymptotically strictly smaller than it needs to be for general reasons, and the reason seems to be associated in some mysterious way with the behavior of the Euclidean algorithm.

That the entropy is smaller than it might be has been proved earlier in another context — the study of certain probability measures on the real line called ICBMs, which have to do with β-adic expansions of real numbers for $\frac{1}{2} < \beta < 1$ — however without any more precise estimate on its value or any indication of how to compute it. In turn these measures are naturally associated with certain dynamical systems on the square called baker's transformations, and the entropy is related to a metric dimension of the attractor. The fact that the entropy is asymptotically smaller than it might be means that the attractor is a strange attractor in some sense. The author's interest in the problem came from trying to estimate the value of the dimension in some way.

In section 2, we define entropy and phrase the question more precisely. In section 3, we discuss some of the combinatorics of the problem. Here there is a connection with another fascinating, but lesser-known, sequence of elementary number theory — Stern's diatomic sequence, which is discussed in section 4. It is here the connection with the Euclidean algorithm appears. In section 5, we develop a generating function for the entropy. In section 6, using the generating function, we make some asymptotic numerical estimates, and in particular, establish rigorous bounds on the entropy. These first sections of the paper are self-contained and elementary. For the interested reader, the connection with probability measures is discussed in section 7. In this context, the fact that the entropy is small is a consequence of the fact that the golden ratio is what is called a PV (Pisot-Vijayarghavan) number. In section 8, we mention some relations with other areas of mathematics and discuss the generalization to hyper-Fibonacci numbers.

The author would like to acknowledge the help of Don Zagier, who taught him how to manipulate the generating functions of sections 5 and 6 and who gave the manuscript a critical reading.

2. Entropy.

The entropy of a code was introduced by Claude Shannon in his seminal papers on information theory. Any introductory text on information theory will have a complete exposition. Here we need only first definitions. Consider first a finite probability space X, which is a finite set of points $x_1, x_2, \ldots, x_N$ weighted with non-negative numbers (probabilities) $p_1, p_2, \ldots, p_N$ such that $\sum_{i=1}^{N} p_i = 1$. The *information* of x_i is $-\log_2 p_i$. Conventionally 2 is chosen as the base of logarithms, but in fact our main concepts are independent of the choice of base. The *entropy* is the expectation of the information:

$$H(x) = -\sum_{i=1}^{N} p_i \log_2 p_i$$

(where $0 \cdot \log_2 0 = 0$).

It is a standard fact, proved for example with Lagrange multipliers, that $H(x) \leq \log_2 N$ and that this bound is obtained precisely when all the p_i are equal. At the other extreme, if one $p_i = 1$ and all the others are 0, then $H(x) = 0$. The entropy measures the uniformity of X; the more evenly the points are weighted, the larger the entropy.

We next define the entropy of a *code* C of length n. For simplicity we consider only binary codes. A *codeword* is one of the 2^n strings of length n of zeroes and ones. Each such codeword represents a (plaintext) *word*. The set of words is finite. Each word x_i is weighted by the number of different representations it has by codewords. Thus if x_i is represented by r_i different codewords, set $p_i = 2^{-n} r_i$. The entropy $H(C)$ of the code is $H(X)$.

If a word x_i has several representatives, C has redundancy. The more redundancy, the smaller the entropy. If there are N words represented by the code, $H(C) \leq \log_2 N$. To take into account this crudest of constraints, we define the *relative entropy*

$$H_R(C) = H(C)/\log_2 N,$$

which is bounded by 1 and is independent of the choice of base for logarithms.

Consider some examples:

1. *Binary code* Bin_n. This is the usual representation of integers in binary notation. We consider representations of length n where we fill in with zeroes on the left if necessary. The $N = 2^n$ integers from 0 through $2^n - 1$ are the plaintext words. Each is represented once, so $H_R(\mathrm{Bin}_n) = 1$, the maximum possible.

2. *Half-binary code* $\frac{1}{2}\mathrm{Bin}_n$. For a string of length n, the odd positions (counting from the right) are the binary representations of an integer. The even positions are fillers and mean nothing. Thus $N = 2^{\left[\frac{n+1}{2}\right]}$ and each word has $2^{\left[\frac{n}{2}\right]}$ representations. Thus $H(\frac{1}{2}\mathrm{Bin}_n) = \left[\frac{n+1}{2}\right]$, and $H_R(\frac{1}{2}\mathrm{Bin}_n) = 1$. The redundancies are evenly distributed.

3. *Fibonacci code* Fib_n. Recall that $\sum_{r=1}^{n} f_r = f_{n+2} - 2$. Thus the number of integers that can be represented by the length n Fibonacci code is $f_{n+2} - 1$. Clearly there is redundancy in

the representations and it is unevenly distributed. For example, with strings of length 3, the number 3 has 2 representations 011 and 100. The other 6 numbers from 0 through 6 each have one representation. Thus

$$H(\mathrm{Fib}_3) = -\frac{6}{8}\log_2\frac{1}{8} - \frac{1}{4}\log_2\frac{1}{4} = \frac{11}{4},$$

and

$$H_R(\mathrm{Fib}_3) = \frac{11}{4}\Big/\log_2 7 \approx .98.$$

In the Fibonacci representation, any occurrence of 011 in a string can be replaced with 100 and *vice-versa*, leading to redundancies. For example, considering 13,

$$010110 \equiv 011000 \equiv 100000.$$

It is true, and is shown in the next section, that two strings in Fib_n represent the same integer if and only if one can be obtained from the other by a sequence of interchanges of substrings 011 and 100.

In Table 1, we list $H_R(\mathrm{Fib}_n)$ for n from 1 to 38. For $n = 38$, there are $2^{38} \approx 2.75 \times 10^{11}$ codewords representing 165,580,140 integers. The table is computer generated, not by formulae, but by counting aided by the combinatorics of the next section.

Table 1: Entropy and Relative Entropy

Level	Number of Represented Numbers	Entropy $H(\mathrm{Fib}_n)$	Relative Entropy $H_R(\mathrm{Fib}_n)$	Level	Number of Represented Numbers	Entropy $H(\mathrm{Fib}_n)$	Relative Entropy $H_R(\mathrm{Fib}_n)$
1	2	1.0000000000	1.0000000000	21	46 367	15.2805148110	.9857881012
2	4	2.0000000000	1.0000000000	22	75 024	15.9717806490	.9862128420
3	7	2.7500000000	.9795697645	23	121 392	16.6630464471	.9866029267
4	12	3.5000000000	.9763003098	24	196 417	17.3543122453	.9869623656
5	20	4.2028195311	.9724408733	25	317 810	18.0455780344	.9872945925
6	33	4.9056390622	.9724932165	26	514 228	18.7368438236	.9876025608
7	54	5.5992384298	.9729535856	27	832 039	19.4281096107	.9878888190
8	88	6.2928377974	.9742092136	28	1 346 268	20.1193753978	.9881555757
9	143	6.9845854170	.9755182864	29	2 178 308	20.8106411844	.9884047526
10	232	7.6763330366	.9768836348	30	3 524 577	21.5019069711	.9886380289
11	376	8.3677002066	.9781534042	31	5 702 886	22.1931727576	.9888568774
12	609	9.0590673766	.9793270408	32	9 227 464	22.8844385441	.9890625954
13	986	9.7503548474	.9803840842	33	14 930 351	23.5757043306	.9892563294
14	1 596	10.4416423183	.9813347701	34	24 157 816	24.2669701171	.9894390964
15	2 583	11.1329128094	.9821859627	35	39 088 168	24.9582359036	.9896118019
16	4 180	11.8241833006	.9829496193	36	63 245 985	25.6495016901	.9897752543
17	6 764	12.5154501207	.9836359343	37	102 334 154	26.3407674766	.9899301777
18	10 945	13.2067169408	.9842549083	38	165 580 140	27.0320332631	.9900772226
19	17 710	13.8979829569	.9848150911				
20	28 656	14.5892489730	.9853240185				

For $n \geq 3$, $H_R(\mathrm{Fib}_n) < 1$. Some integers have more representations than others. However, it may be that the redundancies even out as $n \to \infty$. We define the *asymptotic relative entropy*

$$H_R(\mathrm{Fib}) = \liminf_{n\to\infty} H_R(\mathrm{Fib}_n),$$

and ask in particular whether $H_R(\mathrm{Fib}) < 1$. Is the Fibonacci code terminally lumpy? Or are the entries in Table 1 converging to 1? It turns out that $H_R(\mathrm{Fib})$ exists as a limit (not just lim inf) and that $H_R(\mathrm{Fib}) < 1$.

This value is a number as intimately attached to the Fibonnaci sequence as the golden ratio. One can reasonably ask if it can be expressed in some more-or-less closed form. In section 6, $H_R(\mathrm{Fib})$ is expressed as the sum of an infinite series. With this series, we can estimate $H_R(\mathrm{Fib}) \approx .995713126686$. In fact, we set the bounds

$$.997161165488 > H_R(\mathrm{Fib}) > .995458787137.$$

3. The Fibonacci graph.

We would like to represent the Fibonacci code in graphical form to get an overall view of it. The node (vertices) of the graph are in levels $n = 0, 1, 2, \ldots$ and the edges connect nodes in level n with nodes in level $n + 1$. The nodes represent words (integers). The level is the length of a string. We define the graph inductively, starting with a single node at level 0 (the empty) string. From each node two edges descend to nodes at the next level, a right edge and a left edge, subject to the following rule. The node obtained by one right descent followed by two left descents is the same as the node obtained by one left descent followed by two right descents.

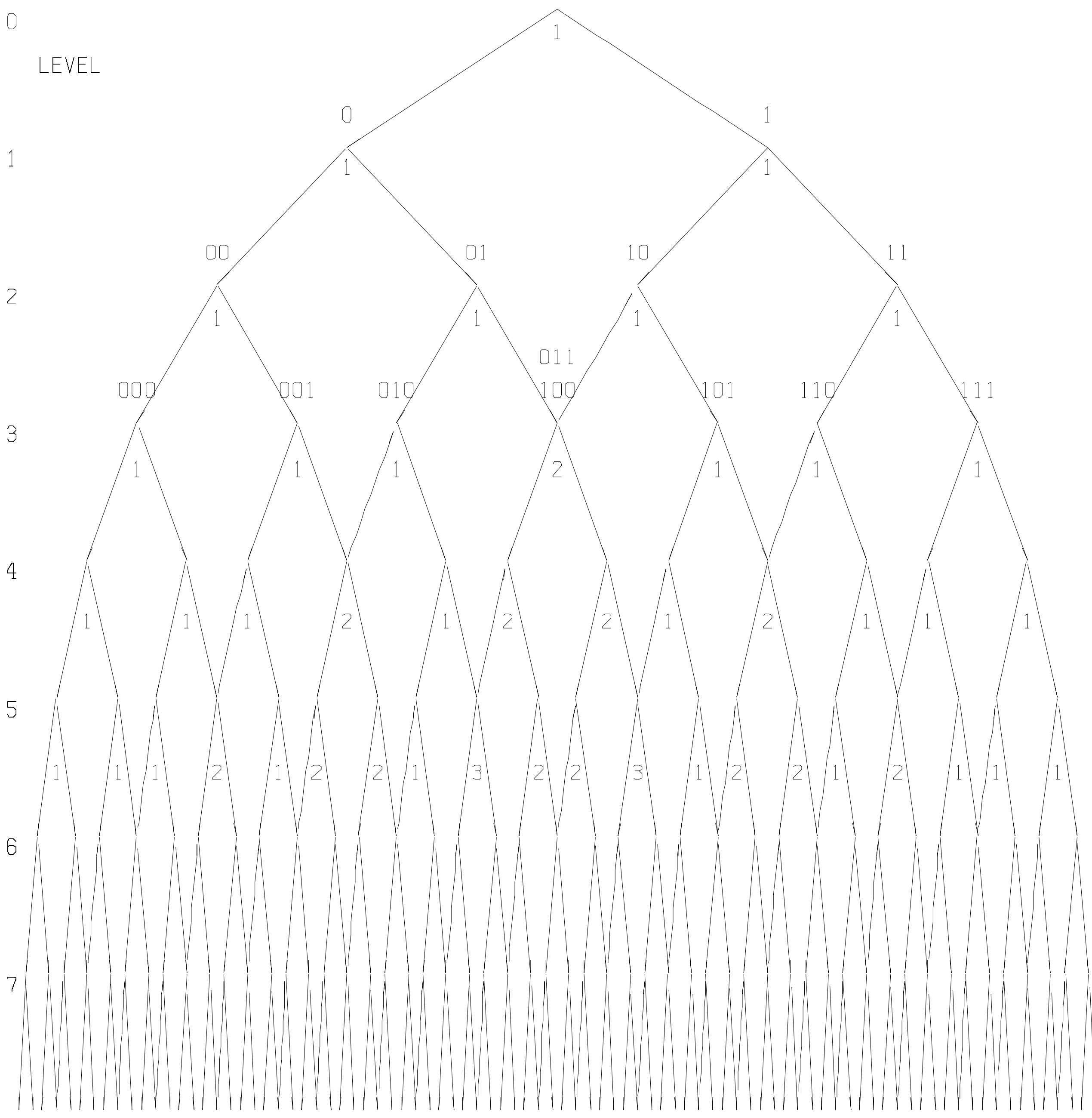
The nodes are labelled with codewords as follows. The node at level zero is labelled with the empty string and called the *root node*. Inductively, if a node is labelled 's,' it right descendent is labelled '$s1$' and its left descendent is labelled '$s0$.' The rule above means that some nodes have more than one label. Alternatively, there is an equivalence of labels; that equivalence is generated by the relation: any substring $011 \equiv 100$. The level is the length of the codeword.

In Table 2, the graph is pictured through level 8. The labelling is indicated above the nodes through level 3. We call this graph the *Fibonacci graph*. We claim that if we number the nodes at level n from left to right with integers $0, 1, 2, \ldots$, the node numbered k is labelled with the set of length-n Fibonacci representations of k (in Table 2, the number k, as a decimal number, is not shown — only the Fibonacci representations). In particular, *the number of Fibonacci representations of length n of an integer k is the number of descending paths from the root node to the node in level n numbered k.* The number of such paths is called the *count* at the node; it is indicated in Table 2 below each node through level 5.

The claim above requires proof. By construction the graph represents strings subject to the equivalence relations generated by $011 \equiv 100$. There is thus a well-defined set map

$$\{\text{nodes}\} \to \{\text{non-negative integers}\}$$

Table 2: The Fibonacci graph



The strings above each node are the Fibonacci representations corresponding to that node. The number below each node is the number of such representations, equivalently, the number of paths from the top (root) node to the particular node. The level (length of string) is indicated at the left.

given by

$$\text{node} \to \text{integer with Fibonacci representation given by label.}$$

The claim is that at each level this map is one to one. If not, at some level n, there are less than $f_{n+2} - 1$ nodes. Let g_n denote the number of nodes at level n. The claim is proved if we show $g_n = f_{n+2} - 1$. To this end, note that $g_n = 2g_{n-1} - g_{n-3}$ since there is one equivalence for each node in level $n - 3$. Note that $f_{n+2} - 1$ satisfies this difference equation, and since $g_n = f_{n+2} - 1$ for $n = 1, 2, 3$ (by inspection), $g_n = f_{n+2} - 1$ for all n. Thus the claim is proved.

As a corollary, we have shown that two Fibonacci representations of an integer differ by the equivalence generated by $011 \equiv 100$, as mentioned in section 2. The counts on the Fibonacci graph can be matched up with tables in Carlitz [1968]. Elementary facts about Fibonacci representations can be deduced from the graph. For example, by induction one can show there are $2n$ integers with one Fibonacci representation of length n.

4. The Stern graph.

At level 38 of the Fibonacci graph, there are 165,580,140 nodes. For numerical calculation, even with a computer it is not a good strategy to run through all these nodes to count; some shortcuts are needed. We turn now to some factor graphs of the Fibonacci graph which provide us with the shortcuts. The structure we discuss does not show up well until level 7 or so. It is suggested the reader continue the counts on the Fibonacci graph through at least level 7.

Suppose we erase all nodes from the Fibonacci graph with count 1 together with the the two edges descending from each such node. The remainder of the graph falls apart into disjoint, but isomorphic graphs. Each such subgraph has a *top* node (the one with the least level) with a count of 2. In Table 2, 11 such tops are visible through level 8; the first is in the center position of level 3. For each level larger than 3, there are two tops. Suppose the top of such a subgraph is at level r. At level $r + 2$, there are two nodes with count 3. At level $r + 4$, there are two nodes with count 4 and two with count 5. We want to understand the structure of these counts.

Table 3: Stern Diatomic Series

Level																																	
0	1																																1
1	1																2																1
2	1								3								2								3								1
3	1				4				3				5				2				5				3				4				1
4	1		5		4		7		3		8		5		7		2		7		5		8		3		7		4		5		1
5	1	6	5	9	4	11	7	10	3	11	8	13	5	12	7	9	2	9	7	12	5	13	8	11	3	10	7	11	4	9	5	6	1
⋮																																	

In the mid-1800's, M. Stern introduced the "diatomic series" of Table 3. Starting with two 1's at level 0, an entry is made at level n either as (i) a copy of an entry in level $n - 1$ in the same column or (ii) the sum of two adjacent entries in level $n - 1$ in an intermediate column. This series has a number of interesting properties. Any integer r appears precisely $\phi(r)$ times in each level $\geq r-1$; here $\phi(r)$ is Euler's totient function. The largest entry in level n is f_{n+1}. Any two adjacent elements are coprime; conversely any two ordered coprime numbers appear adjacently exactly once in the series. For more complete discussions, the reader is referred to Lehmer [1929], Williams & Browne [1947], Lind [1969]. An equivalent construction appears in Knuth [1969] under the name Stern-Pierce tree, where it is used to analyze a numerical rounding procedure. This is related to a connection between the nth level of the Stern series and the $(n-1)$st Farey series.

Stern called the entries at the top of any column *dyads.* The dyads are precisely the numbers occurring as counts in the subgraphs of the Fibonacci graph discussed above. By induction, or by known results (see Lehmer [1929]), the sum of the dyads at level n is $2 \cdot 3^{n-1}$. For descriptive purposes, we construct the following labelled and rooted tree (in close analogy with the construction in Knuth [1969]), which we call the *Stern dyad tree.* It has one node (the root) at level 0, and one edge from this node to a node at level 3 which has label 2 (the first dyad). Starting from this node we have a binary tree (two edges descend from each node and their bottom nodes are disjoint from all other nodes); each edge is two levels long and the nodes are labelled with the Stern dyads. This is the Stern dyad tree. The Stern dyad tree is obtained from Table 3 as follows: (i) erase the columns of 1's at the edges, (ii) erase all of each column except the dyad at the top, (iii) from each dyad draw an edge to the two nearest dyads in the next level, (iv) draw an edge from the 2 at level 1 to an unlabelled root at level 0, (iv) change the levels from $0,1,2,3,4,\ldots$ to $0,3,5,7,9,\ldots$ The result looks like a mobile with a hanger. Finally we multiply all the labels of a Stern dyad tree by the positive integer d; we call this a *Stern* $d\times$*dyad tree.*

We next make a connection between the Stern dyad graph and the Euclidean algorithm. The *simple Euclidean algorithm* is the Euclidean algorithm without division. Given a pair $\langle a,b\rangle$ of positive integers with $a \geq b$, let $\langle a^{(1)}, b^{(1)}\rangle = \big(\max(a-b,b), \min(a-b,b)\big)$. This is iterated until $a^{(n)} = b^{(n)}$, which is then the greatest common divisor of a and b. The number n is the *length* of the pair. For example

$$\langle 11,3\rangle \mapsto \langle 8,3\rangle \mapsto \langle 5,3\rangle \mapsto \langle 3,2\rangle \mapsto \langle 2,1\rangle \mapsto \langle 1,1\rangle$$

(length 5). The length is denoted $e(a,b)$. The length is extended to pairs $\langle b,a\rangle$ with $b < a$ by $e(b,a) = e(a,b)$. The length function can be defined inductively by:

$$e(a,a) = 0, \qquad e(a,b) = e(b,a), \qquad e(a+b,a) = e(a+b,b) = e(a,b)+1.$$

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 $\langle 1,1\rangle$ and one node at level 1 labelled with the pair $\langle 2,1\rangle$. Inductively, given a node at level n labelled with the pair $\langle a,b\rangle$, there are two descending edges (left and right) to nodes at level $n+1$ labelled with pairs $\langle a+b,a\rangle$ and $\langle a+b,b\rangle$. At level n, there are 2^{n-1} nodes. For example, at level 3, the pairs are

$$\langle 5,3\rangle, \quad \langle 5,2\rangle, \quad \langle 4,3\rangle, \quad \langle 4,1\rangle.$$

Start with any pair $\langle a,b\rangle$ 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 $\langle a,b\rangle$, 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 $\langle a, b\rangle$, it is routine to locate the pair in the tree. Thus each pair of coprime integers $\langle a, b\rangle$ appears exactly once in this tree at level $e(a, b)$. We call this tree the *Euclidean tree*. Consider this tree labelled with the first of the pair of integers, for each node. It is not hard to see by induction that the tree with these labels is precisely the Stern dyad tree, although the labels are not in the same order. Any dyad d which appears in the Stern dyad tree as the sum $d_1 + d_2$ of dyads, also appears as the sum $d_2 + d_1$. In the correspondence with the Euclidean tree, these nodes correspond to nodes labelled $\langle d, d_1\rangle$ and $\langle d, d_2\rangle$. The Euclidean algorithm is encoded in the Stern tree. Note that this correspondence "explains" a number of the results in the references on the Stern series.

We next embed the Stern tree in the Fibonacci graph. Start with a graph consisting of the outer edges of the Fibonacci graph. That is, it has one node at level 0 and two nodes at each succeeding level, all with count 1. From each of these nodes a Stern dyad tree is hung from its root. There is more. From each node in each hanging Stern dyad tree labelled by a dyad d, a Stern $d\times$dyad tree is hung. This process is continued to exhaustion (at any level the process is finite). At this point we have a tree with three edges descending from each node except the original nodes which were labelled 1. One of the edges drops straight down three levels to a new Stern tree, the other two go down two levels to the left and right. Add more nodes. For each node of level r in this tree with label $l > 1$, two nodes of label l are appended at each level $> r$. Finally add $2n - 2$ nodes with label 1 at each level > 2. We have not kept track of the edge structure, but we claim the labelled nodes we have constructed match up with the nodes of the Fibonacci graph labelled with their counts and the isomorphism respects levels.

Like many combinatorial constructions, this one is best understood by drawing pictures in private. Once the construction is understood, the claim can be formally verified using for example the generating function for the Stern series developed in Lind [1969] and the generating function $\prod_{r=1}^{n}(1 + x^{f_r})$ for the nth level of the Fibonacci graph.

The point is that the Stern tree is the irreducible part of the Fibonacci graph; everything about Fibonacci representations is somehow encoded in the Stern tree. For example, the dyads can be computed in terms of continuants of continued fractions. The formulae in Carlitz [1968] involving continuants can be derived from those coming from the Stern tree. For direct numerical calculations, the Stern dyads can be computed directly and combined combinatorially to determine the labels of the Fibonacci graph. This is how Table 1 was generated. A frequency table was constructed. Let $F_n(k)$ denote the number of integers with k Fibonacci representations of length n. Then

$$H(\mathrm{Fib}_n) = -\sum_k kF_n(k)2^{-n}\log_2 2^{-n}k = n - \sum_k kF_n(k)2^{-n}\log_2 k.$$

The number $F_n(k)$ is the number of nodes of the Fibonacci graph at level n with count k. Let $S_n(k)$ denote the corresponding frequency count of Stern dyads. The discussion above shows that $F_n(k)$ can be derived in a simple manner from $S_n(k)$. To calculate $H(\mathrm{Fib}_{38})$ requires $S_n(k)$ through $n = 18$ ($2^{18} = 262,144$), not several hundred million.

5. A generating function.

In this section, we develop a generating function for the frequency count of the Fibonacci graph, and the entropy.

Let $F_n(k)$ equal the number of integers having exactly k Fibonnaci expansions of length n — the frequency count. Let $H_n = H(\mathrm{Fib}_n)$. Note that

$$\sum_{k=1}^{\infty} F_n(k) = f_{n+2} - 1,$$

$$\sum_{k=1}^{\infty} kF_n(k) = 2^n,$$

$$\sum_{k=1}^{\infty} \frac{k}{2^n} F_n(k) \log_2 \frac{2^n}{k} = H_n.$$

Let $f_k(x) = \sum_{n=1}^{\infty} F_n(k)x^n$. Let $\hat{\alpha}_k(n)$ be the number of times the integer k appears in the Stern dyad tree at level n, and let $\hat{\alpha}_k(x) = \sum_{n=1}^{\infty} \hat{\alpha}_k(n)x^n$. From the description of the Euclidean graph given in section 4, it is apparent that

$$\hat{\alpha}_k(x) = \sum_{\substack{i=1,\ldots,\infty \\ i \leq k \\ (k,i)=1}} x^{e(k,i)}.$$

When the Stern graph is embedded in the Fibonacci graph, the levels are shifted; this shift leads us to define

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

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

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

A short list of $\alpha_k(x)$ and $l_k(x)$ is given in Table 4. Next define functions of two variables x and s:

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

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

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

The formulae above imply

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

Note also that

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

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

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

so that

$$\mathcal{H}(x) = \sum_{n=0}^{\infty} H_n x^n = \frac{x}{(1-x)^2} - \frac{1}{\ln 2}\left.\frac{\partial\Phi(x/2;s)}{\partial s}\right|_{s=1}.$$

Since the sum of the Stern dyads at level n is $2\cdot 3^{n-1}$,

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

Contemplation of the Fibonacci graph as a union of Stern $d\times$dyad trees leads to the expression

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

To see this, note that $l_k(x)$ is the generating function for the tree constructed by the process which starts with the Stern dyad tree and iteratively hangs Stern $d\times$dyad trees from each node with label d. One of these trees is hung from each of the node at level 0 and 2 nodes at each level ≥ 1. This leads to the generating function

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

For each of these nodes, there are two nodes with the same label at every larger level. This involves multiplying the generating function by another $(1+x)/(1-x)$. The count for $k=1$ is handled separately, giving the above expression. Hence

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

Thus

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

On the other hand

$$\frac{\partial \mathcal{A}(x;s)}{\partial s}\Big|_{s=1} = -\sum_{\substack{k>i>0\\(k,i)=1}} x^{2e(k,i)+1} k \ln k = \sum_{n=1}^{\infty}\Big(\sum_{\substack{k>i>0\\(k,i)=1\\e(k,i)=n}} k\ln k\Big) x^{2n+1}.$$

Hence

$$\mathcal{H}(x) = \frac{x}{(1-x)^2} - \frac{(4-3x^2)^2}{(4-x^2)^2(1-x)^2}\sum_{n=1}^{\infty}\Big(\sum_{\substack{k>i>0\\(k,i)=1\\e(k,i)=n}} k\log_2 k\Big)\left(\frac{x}{2}\right)^{2n+1} = \frac{x}{(1-x)^2}\mathcal{T}\left(\frac{x^2}{4}\right) \tag{1}$$

where

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

with

$$\kappa_n = \sum_{\substack{k>i>0\\(k,i)=1\\e(k,i)=n}} k\log_2 k. \tag{3}$$

This is the generating function for the entropy. Note that κ_n is the sum of $k\log_2 k$ over the dyads at level n in the Stern dyad tree.

As a formal corollary of (1), note that $(1-x)^2\sum_{n=0}^{\infty} H(\mathrm{Fib}_n)x^n$ is an odd function of x. Thus for n odd

$$H(\mathrm{Fib}_{n-1}) - 2H(\mathrm{Fib}_n) + H(\mathrm{Fib}_{n+1}) = 0. \tag{4}$$

This may be checked in Table 1.

The series (2), as written, converges too slowly for worthwhile estimates. It is not hard to see that $2\cdot 3^{n-1}\log_2(n+1) < \kappa_n < 2\cdot 3^{n-1} n\log_2\phi$, where ϕ is the golden ratio. Accordingly we define coefficients μ_n and λ_n by the formulae

$$\frac{1-3x}{(1-x)^2}\sum_{n=1}^{\infty}\mu_n x^n = \frac{1}{2}\left(\frac{1-3x}{1-x}\right)^2\sum_{n=1}^{\infty}\kappa_n x^n, \tag{5}$$

$$\sum_{n=1}^{\infty}\lambda_n x^n = \frac{1}{2}\left(\frac{1-3x}{1-x}\right)^2\sum_{n=1}^{\infty}\kappa_n x^n. \tag{6}$$

We can put useful bounds on the μ_n. We claim that

$$3^{n-1}\log_2 1.5 < \mu_n < 2\cdot 3^{n-2} \tag{7}$$

(and in particular, $\mu_n > 0$ and it grows geometrically). That is

$$1.75\cdot 3^{n-2} < \mu_n < 2\cdot 3^{n-2}.$$

To prove (6), we consider a node at level $n > 1$ with label $\langle a, b\rangle$, $a > b$ in the Stern dyad tree. It has a "sibling" pair $\langle a, a-b\rangle$ (both descending from the pair $\langle a-b, a\rangle$). These two spawn pairs labelled

$$\langle a+b, a\rangle, \quad \langle a+b, b\rangle, \quad \langle 2a-b, a\rangle, \quad \langle 2a-b, a-b\rangle$$

at level $n+1$. Accordingly $\mu_{n+1} = \frac{1}{2}(\kappa_{n+1} - 3\kappa_n)$ can be written

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

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

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

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

This proves the first inequality of (7). On the other hand, the function $x \mapsto x\log_2 x + (3-x)\log_2(3-x)$ is convex on the interval $[1,2]$ and thus takes it maximum at one or both endpoints. Letting $x = (a+b)/a$, we find that μ_{n+1} is bounded by

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

This proves the second inequality of (7).

The behavior of $\mathcal{H}(x)$ as a meromorphic function depends on the rate of growth of the coefficients. From (7) we see that $\sum_{n=1}^{\infty}\mu_n x^n < 2x/3(1-3x)$, so that $\mathcal{T}(x)$ converges is some disk of radius larger than $1/3$. Consequently, $(1-x)^2\mathcal{H}(x)$ converges in the disk of radius at least $\sqrt{4/3}$; in particular at $x = 1$. Thus $\mathcal{H}(x)$ has a double pole at $x = 1$, a fact consistent with the known rate of growth of $H(\mathrm{Fib}_n)$. In a disk of radius larger than $\sqrt{4/3}$,

$$\mathcal{H}(x) = H_\infty \frac{x}{(1-x)^2} + \frac{\hat{H}}{1-x} + O(1), \tag{8}$$

where $O(1)$ is standard notation for a bounded function. Equivalently, as $n \to \infty$,

$$H(\mathrm{Fib}_n) = nH_\infty + \hat{H} + O(c^{-n}) \tag{9}$$

for some $c > \sqrt{4/3}$.

6. The asymptotic relative entropy.

In this section, we prove the asymptotic relative entropy exists and express it as the sum of an infinite series. We make the computations to establish the estimates stated at the end of section 2.

Recall that

$$\begin{aligned} H_R(\mathrm{Fib}_n) &= H(\mathrm{Fib}_n)/\log_2(f_{n+2}-2) \\ &= H(\mathrm{Fib}_n)\Big/\log_2\left(\frac{\phi^{n+3}+(-1)^{n+2}\phi^{-n-3}}{\sqrt{5}}\right) \\ &= \frac{H(\mathrm{Fib}_n)}{(n+3)\big(\Lambda+O(\phi^{-2n})\big)-\frac{1}{2}\log_2 5}, \end{aligned}$$

which is asymptotic to $H(\mathrm{Fib}_n)/n\Lambda$. Thus from equation (9),

$$H_R(\mathrm{Fib}) = \Lambda^{-1}H_\infty. \tag{10}$$

Table 5: Evaluation of asymptotic relative entropy

N	κ_N	$\Lambda^{-1}\left[\frac{4}{9}4^{-N}\mu_N\right]$	$\Lambda^{-1}\left[1-\frac{4}{9}\sum_{n=1}^{N}4^{-n}\mu_n\right]$
1	1.3862943611	.0770163534	1.2803734137
2	2.4327906486	.0337887590	1.2101574355
3	7.4097128173	.0257281695	1.1566920630
4	22.296725925	.0193547968	1.1164711129
5	66.933951273	.0145255971	1.0862856596
6	200.83142955	.0108958024	1.0636432351
7	602.51491583	.0081721316	1.0466608354
8	1807.5595075	.0061291487	1.0339239316
9	5422.6893160	.0045968707	1.0243712348
10	16268.075992	.0034476547	1.0172067086
11	48804.234074	.0025857414	1.0118333133
12	146412.706916	.0019393061	1.0078032667
13	439238.124411	.0014544796	1.0047807317
14	1317714.37613	.0010908597	1.0025138304
15	3953143.13070	.0008181448	1.0008136545
16	11859429.39400	.0006136086	.9995385225
17	35578288.18347	.0004602064	.9985821736
18	106734864.5517	.0003451548	.9978649119

To obtain an expression for $H_R(\mathrm{Fib})$, we multiply equation (8) by $(1-x)^2$ to obtain

$$(1-x)^2\mathcal{H}(x) = x\mathcal{T}\left(\frac{x^2}{4}\right) = H_\infty + (1-x)\hat{H} + (1-x)^2 O(c^{-n}).$$

This is convergent at $x=1$, so setting $x=1$, we obtain

$$H_R(\mathrm{Fib}) = \Lambda^{-1}\mathcal{T}\left(\frac{1}{4}\right). \tag{11}$$

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

$$H_R(\mathrm{Fib}) = \Lambda^{-1}\bigl(1 - \frac{1}{18}\sum_{n=1}^{\infty}\frac{\kappa_n}{4^n}\bigr) \tag{12}$$

$$= \Lambda^{-1}\bigl(1 - \frac{4}{9}\sum_{n=1}^{\infty}\frac{\mu_n}{4^n}\bigr) \tag{13}$$

$$= \Lambda^{-1}\bigl(1 - \sum_{n=1}^{\infty}\frac{\lambda_n}{4^n}\bigr) \tag{14}$$

The series (12) converges too slowly for effective computation. We use series (13). The partial sums of the series are exhibited in Table 5. Since each $\mu_n > 0$, the values in the third column are upper bounds for $H_R(\mathrm{Fib})$; hence $H_R(\mathrm{Fib})$ is clearly seen to be less than 1. However, using (7), we can bound $H_R(\mathrm{Fib})$. If we truncate at step N, the error E_N is bounded by

$$\frac{\frac{1}{9}\left(\frac{3}{4}\right)^{N-1}\log_2 1.5}{\log_2 \phi} < E_N < \frac{\frac{2}{9}\left(\frac{3}{4}\right)^{N-1}}{\log_2 \phi}.$$

These bounds lead to the stated bounds of section 2.

Table 6: Asymptotic relative entropy

N	λ_N	$\Lambda^{-1}\left[1-\sum_{n=1}^{N}4^{-n}\lambda_n\right]$	N	λ_N	$\Lambda^{-1}\left[1-\sum_{n=1}^{N}4^{-n}\lambda_n\right]$
1	1.000000000000	1.080315067809	10	.186750168736	.995713200862
2	.754887502163	1.012355372552	11	.167131013759	.995713143466
3	.590090465783	.999074463770	12	.150897353618	.995713130510
4	.474047485496	.996407168762	13	.137306231932	.995713127563
5	.389580409352	.995859161488	14	.125804507821	.995713126888
6	.326447711849	.995744361234	15	.115974644527	.995713126733
7	.278194391267	.995719903422	16	.107496997512	.995713126697
8	.240588130845	.995714615519	17	.100125510671	.995713126688
9	.210767793089	.995713457400	18	.093668993028	.995713126686

It is also interesting to consider the computations from (14). This series seems to converge much more rapidly; the results are exhibited in Table 6. There are a couple of surprises in this table. The λ_n are obtained by summing and differencing large numbers. There is no reason to expect them to (a) be positive and (b) be small and decreasing. There is obviously something deeper occurring here. If (a) and (b) are true for all N, the series of Table 6 converges faster than $O(4^{-n})$. From the tabulations, it is converging faster than $O(4.5^{-n})$. Also (6) converges for $|x| < C$ for some $C > 2$. This is surprising, since there is ostensibly a pole of $\mathcal{T}(x)$ at $x = 1$. Other calculations indicate that the full set of digits of Table 6 is uncontaminated by machine roundoff error. If the indicated convergence is valid, the last entry is the value of $H_R(\mathrm{Fib})$, except possibly for the the last digit, which may be a 5.

7. ICBM's and their probability theory.

In this section, we develop the context in which the question of the value of the relative entropy first arose. It has nothing to do with codes *per se*; but rather certain probability measures on the real line, called *infinitely convolved Bernoulli measures* (ICBMs). The question is whether these measures are continuous or singular. This problem is over 50 years old and the entropy was introduced over 25 years ago by A. Garsia in an investigation of that question. The question has arisen again in the context of dynamical systems; 1 plus the relative entropy is a certain metric dimension of the attractor of a dynamical system on the plane.

We begin by considering how the Fibonacci graph of Table 2 was generated. Let $\beta = \phi^{-1} = \frac{1}{2}(\sqrt{5}-1)$. At each level, consider a horizontal line — a copy of the reals — so that the nodes of the graph at that level are a finite set of real numbers. For normalization, suppose the point at level 0 is at the origin and the two points at level 1 are at ± 1. If x is a node at level n, the two nodes descending from it are at points $x \pm \beta^{n-1}$. Thus the four points at level 2 are at $\pm 1 \pm \beta$ and the seven points at level 3 are at $\pm 1 \pm \beta \pm \beta^2$. Note that since $-1+\beta+\beta^2 = 1-\beta-\beta^2 = 0$, the point 0 at level three is a "double point;" this is precisely the relation $011 \equiv 100$ in the Fibonacci representation. Here the relation might be better phrased '$-1, 1, 1$' $\equiv$ '$1, -1, -1$' in terms of the coefficients of powers of β.

At level n, the nodes are at points $\sum_{r=0}^{n-1} a_r \beta^r$, $a_r = \pm 1$. The count at a node is the number of ways it can be represented as such a sum, due to the relation $-1+\beta+\beta^2 = 1-\beta-\beta^2$. Note that the width of the graph is compressed into the interval $[-(1-\beta)^{-1}, (1-\beta)^{-1}]$. We have converted the Fibonacci code into "β-adic" expansions. In fact, it is equivalent to consider expansions $\sum_{r=0}^{n-1} \alpha_r \beta^r$, $\alpha_r = 0, 1$, by letting $\alpha_r = 2a_r - 1$. The count at each node defines a measure. We normalize this measure by dividing through by 2^n at level n, so the total measure is 1. More precisely, for any interval E on the real line, let

$$\mu_\beta^{(n)}(E) = \frac{1}{2^n} \#\{x \in E \colon x = \sum_{r=0}^{n-1} a_r \beta^r, a_r = \pm 1\},$$

where the "#" means the cardinality, counted with multiplicity. The graph of Table 2 is drawn to scale and it is not hard to visualize that there is some kind of limit measure space as the level $n \to \infty$, which measures somehow the density of infinite β-adic expansions $\sum_{r=0}^{\infty} a_r \beta^r$, $a_r = \pm 1$ (or $\sum_{r=0}^{\infty} \alpha_r \beta^r$, $\alpha_r = 0, 1$).

From this point of view, there is no reason to restrict to $\beta = \frac{1}{2}(\sqrt{5}-1)$. We can use any value of β, $0 < \beta < 1$, and construct a graph of β-adic expansions. The individual levels of the graphs can be constructed with a simple computer program. We denote the counting measure at each level by $\mu_\beta^{(n)}$, and the limiting measure (which is known to exist) by μ_β. Consider some other examples.

1. Suppose β is transcendental, so that it satisfies no algebraic equation. Then there are no nodes with count > 1, and at each level n, there are 2^n nodes, each of count 1. Of course, the distribution of the points on the real line is different and the limit measure spaces need not be the same. (More generally, there can be nodes of count > 1 only if β satisfies an equation $\sum_{r=0}^{n} a_r \beta^r = 0$, $a_r = 0, \pm 1$.)

2. Suppose $\beta = \frac{1}{2}$. As above, we can make a translation of the measure $\mu_\beta^{(n)}$ to $\tilde{\mu}_\beta^{(n)}$ by letting $\alpha_r = \frac{1}{2}(a_r + 1)$. We consider points $y = \sum_{r=0}^{n-1} \alpha_r \left(\frac{1}{2}\right)^r$ where $\alpha_r = 0, 1$. Thus we are considering ordinary binary expansions, and the measure $\tilde{\mu}_{\frac{1}{2}}$ is the uniform probability measure on the unit interval and the measure $\mu_{1/2}$ is the uniform probability measure on the interval $[-1, 1]$.
3. Let $\beta = \frac{1}{3}$. We make the translation $\alpha_r = a_r + 1$. Thus we consider points $y = \sum_{r=0}^{n-1} \alpha_r \left(\frac{1}{3}\right)^r$ where $\alpha_r = 0, 2$. We are considering triadic expansions without coefficient 1. Such expansions lead to the Cantor set and $\mu_{1/3}$ is (a translate of) the classical Cantor measure. For any $\beta < \frac{1}{2}$, the support of μ_β is a Cantor type set, and in fact all the μ_β for $\beta < \frac{1}{2}$ are isomorphic.

A measure with support two points each of which has measure $\frac{1}{2}$ is called a *Bernoulli measure* (choosing a point from this space is a Bernoulli trial — a flip of a fair coin). Convolution of measures on the line amounts to adding the supports. Thus the measure $\mu_\beta^{(n)}$ is the convolution $*_{r=0}^{n-1} \nu_\beta^{(n)}$ and $\mu_\beta = *_{r=0}^{\infty} \nu_\beta^{(n)}$. Accordingly μ_β is an *infinitely convolved Bernoulli measure* or ICBM. In this form, these measures are interesting from the point of view of harmonic analysis. The characteristic function (Fourier transform) $\hat{\nu}_\beta^{(n)}(\omega)$ of $\nu_\beta^{(n)}$ is easily seen to be $\cos(\beta^{n-1}\omega)$. The characteristic function of a convolution is the product of the characteristic functions of the components. Thus

$$\hat{\mu}_\beta^{(n)}(\omega) = \prod_{r=1}^{n} \cos(\beta^{r-1}\omega), \qquad \hat{\mu}_\beta(\omega) = \prod_{r=1}^{\infty} \cos(\beta^{r-1}\omega).$$

Among other things, this shows that μ_β is well-defined. For a general discussion of these matters, the reader could see for example Kawata [1972].

These measures and some variants were studied in the 1930s, because their characteristic functions have interesting asymptotic properties. Recall that a measure μ on the reals is *absolutely continuous* is it has a density f so that $\mu(E) = \int_E f \, dx$. At the other extreme, a measure is *totally singular* if it is supported on a Lebesgue null set, e. g., the Cantor measure. In general, a measure is the (essentially unique) sum of an absolutely continuous and a totally singular measure. A measure is *continuous* if every single point has measure 0. The following facts were proved about the μ_β. They are continuous and *pure*, i. e., either absolutely continuous or totally singular (Jessen & Wintner [1938]). For β a root of $\frac{1}{2}$, μ_β is absolutely continuous (Wintner [1935]). For $\beta < \frac{1}{2}$, μ_β is totally singular (since its support is a Cantor set). For $\beta > \frac{1}{2}$, every interval contained in $[-(1-\beta)^{-1}), (1-\beta)^{-1})]$ has strictly positive measure (the measure is dense).

Thus it was generally supposed that for $\beta > \frac{1}{2}$, the μ_β are absolutely continuous. However P. Erdös, who was at the Institute for Advanced Study, visited Johns Hopkins, where A. Wintner was, learned of the question, and using number theory, showed that there are $\beta > \frac{1}{2}$ for which μ_β is totally singular (Erdös [1939]). The property he needed was that β^{-1} is an algebraic integer — that is it satisfies an integral polynomial with lead coefficient 1 — all of whose conjugates (the other roots) lie inside the unit circle. Such numbers β^{-1} are called *Pisot-Vijayarghavan* (PV) numbers. This property allowed him to estimate $\hat{\mu}_\beta(\omega)$ as $|\omega| \to \infty$ and show it does not converge to zero. By the Riemann-Lebesgue lemma, such a measure cannot be absolutely continuous. The simplest such β^{-1} (the only quadratic one) is $\frac{1}{2}(\sqrt{5}+1)$, a root of $z^2 - z - 1$, since its conjugate is $-\beta$. In fact, this was the explicit example of Erdös. Thus for $\beta = \frac{1}{2}(\sqrt{5}-1)$, μ_β is totally singular.

Some years later A. Garsia turned to this question of which μ_β are totally singular (Garsia [1963]). He introduced an entropy. Let $h_n(\beta)$ denote the entropy of the finite probability space defined by $\mu_\beta^{(n)}$. Garsia considered the quantities

$$G_n(\beta) = \frac{h_n(\beta)}{-n\log_2\beta}$$

and showed *inter alia* that

$$G(\beta) = \lim_{n\to\infty} G_n(\beta) = \liminf_{n\to\infty} G_n(\beta)$$

exists and that for β^{-1} a PV number, $G(\beta) < 1$. This result depended on the earlier results about singularity. Note that if there are no multiplicities in the β-adic expansions, then $G(\beta) = -(\log_2\beta)^{-1} > 1$ for $\frac{1}{2} < \beta < 1$. The proof is rather technical, and we do not discuss it here; the interested reader is referred to the original paper. Garsia's method does not give a precise estimate of the value. This in fact was the motivation for the present paper — to independently estimate the value of $G(\phi^{-1})$, the "simplest" case where Garsia's entropy is known to be less than 1.

8. Comments.

1. Algebraic integers, all of whose conjugates lie inside the unit circle, are called *Pisot-Vijayarghavan* (PV) numbers, although the concept evidently goes back to G. Hardy (Vijayarghavan was a student of Hardy). There is an extensive literature on them, mostly in the context of number theory. There are an infinite number of PV numbers between 1 and 2, so by Erdös' result, there are an infinite number of μ_β, $\frac{1}{2} < \beta < 1$ for which μ_β is totally singular. For an introduction to PV numbers, Salem [1963] is particularly recommended. For a recent application of PV numbers to a different coding problem, see Wilf [1987].

2. The Euclidean tree has occurred in other studies; in particular in estimates of Markov numbers (Cohn [1979]; Zagier [1982]). These numbers have to do with how well irrational numbers can be approximated by rationals.

3. It is a standing question to determine for which β, $\frac{1}{2} < \beta < 1$, the ICBM μ_β is absolutely continuous or totally singular. Of course, for $\beta < \frac{1}{2}$, μ_β is totally singular because it is a Cantor measure. Erdös' method establishes that if β^{-1} is a PV number, then μ_β is totally singular. It can be seen directly that if β is a root of $\frac{1}{2}$, then μ_β is absolutely continuous. Erdös established (again by estimating $\lim_{|\omega|\to\infty}\hat{\mu}_\beta(\omega)$) that μ_β is absolutely continuous for lots of β: namely that μ_β is absolutely continuous for almost all β near 1 (Erdös [1940]). Salem (see Salem [1963]) showed that Erdös' original technique works only for β the inverse of a PV number; $\lim_{|\omega|\to\infty}\hat{\mu}_\beta(\omega) \not\to 0$ only for such β. Garsia determined some other algebraic values of β for which μ_β is absolutely continuous (Garsia [1962]). However to this date, there is no effective characterization of which μ_β are totally singular and which are absolutely continuous. In particular, the only β, $\frac{1}{2} < \beta < 1$ for which μ_β is known to be totally singular are the inverses of PV numbers.

4. More recently, these measures have appeared in some examples in dynamics, concerned with metric dimensions and strange attractors. We define the following piecewise-linear (discontinuous) map T_β on the unit square

$$\{(x,y)\colon |x| \le 1, |y| \le 1\}$$

for any β between 0 and 1. Cut the square in half vertically at the line $y = \frac{1}{2}$. The lower half is stretched double vertically and compressed by β horizontally with the right edge remaining on the edge of the square. Thus $(x, y) \to (\beta x + (1 - \beta), 2y + 1)$. The same is true of the upper half, except the left edge remains on the edge of the square: $(x, y) \to (\beta x - (1 - \beta), 2y - 1)$. When $\beta = \frac{1}{2}$, this is the classical bakers' transformation. It so happens that there is a natural invariant measure on the square. The measure is uniform in the vertical direction and is the ICBM μ_β (up to scale) in the horizontal direction (Alexander & Yorke [1984]). The entropy is some kind of metric dimension of the attractor. In particular, the fact that $H_R(\mathrm{Fib}) < 1$ means that the "essential attractor" for $\beta = \frac{1}{2}(\sqrt{5} - 1)$ is some kind of strange attractor — in some sense it has a fractal structure. In fact, Table 6 appears in Alexander & Yorke [1984], but with no theory to back it up. We might also mention that there is another measure associated with β-adic expansions (Gel'fond [1959]; Parry [1960]), and that this measure and μ_β have some relation to each other in the context of dynamics (Alexander & Parry [1988]). The book Billingsley [1979] discusses relations between measure theory and metric dimensions.

5. The rth hyper-Fibonacci numbers are defined by the recursion

$$f_l = \begin{cases} 2^{l-1} & \text{for } l = 1, \ldots, r+1, \\ f_{l-1} + \cdots + f_{l-r+1} & \text{for } l > r+1, \end{cases}$$

and a code exists based on such numbers. The graph for such a code is similar to that of Table 2. The Stern graphs sit inside the graph analogously. In terms of β-adic expansions, the code corresponds to expansions for $\beta = \beta_r$, a root of the polynomial

$$x^r + x^{r-1} + x^{r-2} + \cdots + x - 1 = 0.$$

Indeed the analogy goes further; the inverse of β_r is a PV number, and so μ_{β_r} is totally singular. In fact, Garsia's result works for any β with β^{-1} a PV number, and so the asymptotic relative entropy of this code is strictly less than one. Here we indicate how to show that β_r^{-1} is a PV number (a result which is probably well known).

The proof is based on Rouche's theorem which counts how many roots a function has inside a simple closed curve in the complex plane. Note that β_r^{-1} satisfies the polynomial

$$x^r - x^{r-1} - x^{r-2} - \cdots - x - 1 = 0.$$

We claim this polynomial has $r - 1$ roots inside the unit circle. Append a further root $x = 1$ by multiplying the polynomial by the factor $x - 1$ to obtain the polynomial

$$x^{r+1} - 2x^r + 1.$$

Consider the unit circle. We claim that for x on this circle

$$|-2x^r + 1| > |x^{r+1}| = 1,$$

except for x the rth roots of 1. To see this, let $x = e^{i\theta}$, and compute

$$|-2x^r + 1|^2 - |x^{r+1}|^2 = 4 - 4\cos r\theta,$$

which is positive unless $\cos r\theta = 1$. By direct substitution, the only rth root of 1 which is a root of $x^{r+1} - 2x^r + 1$ is 1 itself. By a corollary of Rouche's theorem, the polynomial $x^{r+1} - 2x^r + 1$ has one less root inside the unit circle as does the polynomial $-2x^r + 1$, namely $r - 1$. Hence $x^r - x^{r-1} - x^{r-2} - \cdots - x - 1 = 0$ has $r - 1$ roots inside the unit circle, as claimed.

It is left to the reader to estimate the value of the asymptotic relative entropy.

References

J. C. Alexander & W. Parry [1988], "Discerning fat baker's transformations," in *Dynamical Systems: Proceedings, University of Maryland 1986–87*, J. C. Alexander, ed., Lect. Notes in Math. #1342, Springer-Verlag, New York–Heidelberg–Berlin, 1–6.

J. C. Alexander & J. A. Yorke [1984], "Fat baker's transformations," *Ergodic Theory Dynamical Systems* 4, 1–23.

P. Billingsley [1979], *Probability and Measure*, John Wiley & Sons, New York, NY.

L. Carlitz [1968], "Fibonacci representations," *Fibonacci Quart.* 6, 193–220.

H. Cohn [1979], "Growth types of Fibonacci and Markoff," *Fibonacci Quart.* 17, 178–183.

P. Erdös [1939], "On a family of symmetric Bernoulli convolutions," *Amer. J. Math.* 61, 974–976.

P. Erdös [1940], "On the smoothness properties of a family of Bernoulli convolutions," *Amer. J. Math.* 62, 180–186.

A. M. Garsia [1962], "Arithmetic properties of Bernoulli convolutions," *Trans. Amer. Math. Soc.* 102, 409–432.

A. M. Garsia [1963], "Entropy and singularity of infinite convolutions," *Pacific J. Math.* 13, 1159–1169.

A. O. Gel'fond [1959], "A common property of number systems," *Izv. Akad. Nauk SSSR Ser. Mat.* 23, 809–814.

B. Jessen & A. Wintner [1938], "Distribution functions and the Riemann zeta function," *Trans. Amer. Math. Soc.* 38, 48–88.

T. Kawata [1972], *Fourier Analysis in Probability Theory*, Academic Press, New York, NY.

D. E. Knuth [1969], *The Art of Computer Programming II, Seminumerical Algorithms*, Addison Wesley, Reading, MA.

D. H. Lehmer [1929], "On Stern's diatomic series," *Amer. Math. Monthly* 36, 59–67.

D. A. Lind [1969], "An extension of Stern's diatomic series," *Duke Math. J.* 36, 55–60.

W. Parry [1960], "On the β-expansions of real number," *Acta Math. Sci. Acad. Hungar.* 11, 401–416.

R. Salem [1963], *Algebraic Numbers and Fourier Analysis*, Heath.

H. S. Wilf [1987], "Strings, substrings, and the 'nearest integer' function," *Amer. Math. Monthly* 94, 855–860.

G. T. Williams & D. H. Browne [1947], "A family of integers and a theorem on circles," *Amer. Math. Monthly* 54, 534–536.

A. Wintner [1935], "On convergent Poisson convolutions," *Amer. J. Math.* 57, 827–838.

D. Zagier [1982], "On the number of Markoff numbers below a given bound," *Math. Comp.* 39, 709–723.