Journal of Integer Sequences, Vol. 7 (2004),
Article 04.3.5

Generalizing the Conway-Hofstadter $10,000 Sequence

John A. Pelesko
Department of Mathematical Sciences
University of Delaware
Newark, DE 19716
USA
pelesko@math.udel.edu

Abstract

We introduce a generalization of the Conway-Hofstadter $10,000 sequence. The sequences introduced, called *k-sequences*, preserve the Conway-Hofstadter-Fibonacci-like structure of forming terms in the sequence by adding together two previous terms, equidistant from the start and end of the sequence. We examine some particular k-sequences, investigate relationships to known integer sequences, establish some properties which hold for all k, and show how to solve many of the defining nonlinear recursions by examining related underlying sequences termed *clock* sequences.

1 Introduction

In a talk at AT&T Bell Labs [1] in 1988, J.H. Conway introduced the sequence (A004001 in the On-Line Encyclopedia of Integer Sequences)

$$1, 1, 2, 2, 3, 4, 4, 4, 5, 6, 7, 7, 8, 8, 8, 8, 9, \ldots$$

defined by the recursion

$$c(n) = c(c(n-1)) + c(n - c(n-1)), \tag{1}$$

with initial conditions $c(1) = c(2) = 1$. Conway had proven that $c(n)/n \to 1/2$, but was unable to establish the rate of convergence. Somewhat overestimating the difficulty of the question he offered a prize of $10,000 to the first person who could. The challenge was answered by C.L. Mallows [2] shortly thereafter. Mallows not only established the rate of convergence, but uncovered additional structure in the sequence as well. The exchange caught the attention of the popular press inspiring an entertaining article in the *New York Times* [3]. The popularization of A004001 generated by this exchange also led to the study of Kubo and Vakil [4] where much of the combinatorial structure of the sequence was unveiled. The use of a *compression operation* to characterize the sequence allowed for simple proofs of many of A004001's interesting properties. We also note that unbeknownst to Conway and Mallows the sequence had previously been introduced by Hofstadter [5] and had also appeared in the problems section of the American Mathematical Monthly [6]. Today A004001 is known either as the "Conway-Hofstadter $10,000 sequence" or as the "Conway-Newman" sequence. We will refer to it as the "Conway-Hofstadter" sequence.

Many of the properties that inspired interest in (1) are nicely enumerated by Kubo and Vakil [4]. For convenience of the reader we list those relevant to this paper here:

1. $c(n) \leq n$.
2. $c(n) - c(n-1) = 0$ or 1, for all $n \geq 1$.
3. $c(n) \geq n/2$, with equality iff n is a power of 2 and $n \neq 1$.
4. $c(n)/n \longrightarrow 1/2$ as $n \longrightarrow \infty$.
5. $c(2n) \leq 2c(n)$ for all n.

In this paper we generalize the Conway-Hofstadter sequence. Our generalized class of sequences shows much of the structure of (1), but also exhibits interesting new behavior. The generalization leads to new representations of old sequences and to new solvable nonlinear recursions.

2 The Generalization: *k-Sequences*

In reading the work of Mallows [2] or Kubo and Vakil [4] one is immediately struck by the statements following the presentation of the first property listed above of the Conway-Hofstadter sequence. Both authors note that $c(n) \leq n$ and then go on to say "so that $c(n)$ is well-defined by the recurrence." To understand this comment and to appreciate the motivation for our generalization it is worth visualizing how terms in the Conway-Hofstadter sequence are formed. Consider the first five terms of the Conway-Hofstadter sequence:

$$1, 1, 2, 2, 3.$$

To form the sixth term, we note that the fifth term is equal to 3, count forward from the beginning of the sequence three terms, backwards from the end of the sequence three terms, and add the results to find $c(6) = 2 + 2 = 4$. This procedure generates all terms in the sequence. Note that there is a beautiful symmetry in this construction process; in forming the nth term, one term from the first half of the sequence is added to a term from the second half of the sequence. These terms are always equidistant from the start and end of the sequence. The observation of Mallows or Kubo and Vakil is equivalent to noting that $c(n) \leq n$ assures that we never count past the end (or the beginning) of the sequence. Of course, if we consider clock or modular arithmetic, counting beyond the start or end of the sequence is no longer a problem. This immediately suggests our generalization.

Definition 1. *We say that* $\{c_k(n)\}$ *is a Conway-Hofstadter-like sequence of order* k *when defined by the recursion*

$$c_k(n) = c_k(kc_k(n-1) \bmod (n-1)) + c_k(n - kc_k(n-1) \bmod (n-1)) \tag{2}$$

with $c_k(1) = c_k(2) = 1$. *We also call such sequences, k-sequences, and denote them* c_k.

Note that this generalization preserves the symmetry of the Conway-Hofstadter sequence. That is, the only modifications to the construction process above are that the last term in the sequence in multiplied by k, and the count from the start and end of the sequence is done using modular arithmetic. However, in forming the nth term, one term from the first half of the sequence is still added to one term from the second half of the sequence; these terms are again equidistant from the start and end of the sequence. Throughout this paper, we observe the convention that a zero in modular arithmetic mod n is replaced by n.

2.1 A Glance at Some k-Sequences

Computing the first few terms of k-sequences for various k reveals some familiar sequences hiding among the k's as well as some new surprises. The observed behavior of the first fifteen k-sequences is summarized in the following table:

k-sequence	First twenty terms
c_1	1,1,2,2,3,4,4,4,5,6,7,7,8,8,8,8,9,10,11,12 ...
c_2	1,1,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,11 ...
c_3	1,1,2,3,4,4,5,6,6,7,8,8,9,10,10,11,12,12,13,14 ...
c_4	1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10 ...
c_5	1,1,2,3,3,4,4,5,6,6,7,7,8,9,9,10,10,11,12,12 ...
c_6	1,1,2,3,3,4,5,5,6,7,7,8,9,9,10,11,11,12,13,13 ...
c_7	1,1,2,2,3,4,4,5,6,6,7,7,8,8,9,10,11,12,11,12 ...
c_8	1,1,2,3,4,4,5,6,7,7,8,9,10,10,11,12,13,13,14,15 ...
c_9	1,1,2,3,3,4,5,5,6,7,7,8,9,9,10,11,11,12,13,13 ...
c_{10}	1,1,2,2,3,4,4,5,5,6,7,7,8,8,9,10,10,11,11,12 ...
c_{11}	1,1,2,3,4,4,5,5,6,7,8,9,9,9,10,12,12,13,13,14 ...
c_{12}	1,1,2,3,4,4,5,6,7,7,8,9,10,10,11,12,13,13,14,15 ...
c_{13}	1,1,2,2,3,3,4,5,6,5,6,7,7,8,9,9,10,10,11,10 ...
c_{14}	1,1,2,3,3,4,4,5,6,6,7,7,8,9,10,10,10,12,12,13 ...
c_{15}	1,1,2,3,4,5,5,6,6,7,9,8,9,10,11,12,12,13,14,15 ...

The sequence c_1 is, of course, the Conway-Hofstadter sequence. Other familiar sequences are lurking in this list. The sequence c_4 is the nice sequence $\lfloor (n+1)/2 \rfloor$ which is equivalent to A004526. Both c_6 and c_9 appear to follow the pattern "one even followed by two odd" and hence appear equivalent to A004396. The sequences c_8 and c_{12} appear equivalent to A037915, or more simply $\lfloor (3n+4)/4 \rfloor$. These suggested equivalences require proof. We do not present such proof at this point. Rather, we will first establish some general results about the c_k's, then, demonstrate a method for uncovering the hidden structure of many c_k's, and finally show how to prove an equivalence suggested by the table above. We do at this point note that direct computation of various k-sequences highlights interesting similarities and differences between c_1 and other c_k's. Properties (1) and (3) of c_1 appear to be satisfied for all k. Property (2) however is violated for most c_k's. This is seen both through failure of monotonicity and in the "skipping" of integers in sequences such as c_{11} and c_{13}. Of course, the other striking feature of the c_k is the apparent new representation of some familiar sequences such as A004526 or A004396.

3 Properties of the c_k

Observations suggest that all c_k satisfy upper and lower bounds on growth similar to the bounds on c_1. This is indeed true and we have

Proposition 1. *$c_k(n) \leq n$ for all $n \geq 1$, $k \geq 1$.*

Proof. For any fixed k, we proceed by induction on n. Note that $c_k(1) = c_k(2) = 1$ and $c_k(3) = 2$ and hence $c_k(n) \leq n$ for $1 \leq n \leq 3$. Now, assume $c_k(j) \leq j$ for all j satisfying $1 \leq j \leq n$ and consider $c_k(n+1)$. We have

$$c_k(n+1) = c_k(kc_k(n) \bmod n) + c_k(n+1-kc_k(n) \bmod n).$$

Let $j = kc_k(n) \bmod n$ and observe that $1 \leq j \leq n$. Hence

$$c_k(n+1) = c_k(j) + c_k(n+1-j) \leq j + n + 1 - j = n + 1$$

and

$$c_k(n+1) \leq n+1$$

as desired. □

A similar argument yields the lower bound

Proposition 2. *$c_k(n) \geq n/2$ for all $n \geq 1$, $k \geq 1$.*

A key difference between c_1 and a general c_k is that property (2) need not hold. This allows a particular c_k to be non-monotone and to "skip" integers. For example, $c_7(19) - c_7(18) = -1$ demonstrating the non-monotone property while c_{11} does not contain the number 11, as we shall soon see. The bounds above immediately provide a means to prove that integers can indeed be skipped. We have

Proposition 3. *Let $m > 0$ and suppose m does not appear in the first N terms of $c_k(n)$ where $N > 2m$, then m never appears in $c_k(n)$.*

Proof. $c_k(N) \geq \frac{N}{2} > m$. □

Notice that this proposition, along with computation of the first 23 terms of c_{11} establishes that c_{11} is indeed "missing" 11. We can also easily bound the maximum number of occurrences of a particular integer in the sequence c_k.

Proposition 4. *Let $f_k(m)$ denote the total number of occurrences of m in the sequence c_k. Then, $f_k(m) \leq m+1$ for all k and m.*

Proof. By our lower bound on c_k we have $c_k(2m) \geq m$. By our upper bound we have $c_k(m) \leq m$. Hence m can only appear amongst the $m+1$ terms $c_m, c_{m+1}, \ldots c_{2m}$. $\square$

Another natural question is whether or not c_k and c_j can be "equivalent." We consider two notions of equivalence.

Definition 2. *We say that c_k and c_j are numerically equivalent to order N iff $c_k(n) = c_j(n)$ for all n satisfying $1 \leq n \leq N$. We say that c_k and c_j are structurally equivalent to order N iff $kc_k(n) \bmod n = jc_j(n) \bmod n$ for all n satisfying $3 \leq n \leq N$.*

Structural equivalence tracks the process of forming a k-sequence. It decides whether or not two k-sequences were formed by adding together terms located at the same point in each sequence. Structural equivalence clearly implies numerical equivalence. The converse is not true. It is possible for two sequences to be numerically equivalent, but not structurally equivalent. Of the two types of equivalence, we consider structural equivalence to be fundamental. We can compute the set of all k-sequences structurally equivalent to a given sequence, c_j, by solving a system of linear congruences. For example, consider c_2, which begins $\{1, 1\}$. Since $2 \equiv 0 \pmod 2$ we may find all k-sequences structurally equivalent to order 2 by solving the congruence $k \equiv 0 \pmod 2$. The set of even integers satisfies this congruence. At the next step, c_2 is $\{1, 1, 2\}$, and since $4 \equiv 1 \pmod 3$ we must solve the congruence $2k \equiv 1 \pmod 3$. The solutions to this are numbers in the arithmetic progression $2, 5, 8, \ldots$. Hence the k-sequences structurally equivalent to c_2 at order 3 are those in the progression with steps of length 2×3, i.e., $k = 2, 8, 14, \ldots$. Replacing 2 with j and generalizing the argument above we may show

Proposition 5. *Given any k, N, there exists a $j \neq k$, such that c_k and c_j are structurally equivalent to order N. Further, if j is the smallest such integer, $j \to \infty$ as $N \to \infty$.*

Notice that this proposition implies that no two k-sequences are structurally equivalent of infinite order. In this sense, the k-sequences are distinct. As mentioned above, two k-sequences may be numerically equivalent, but not structurally equivalent. Numerical investigation suggests that numerical equivalence is infrequent.

4 The Beat of the c_k's

The structure implicit in the notion of structural equivalence can also shed light on the behavior of particular k-sequences. The underlying structure of a given k-sequence, that is the sequence of terms used to create c_k, is tracked by the associated *clock* sequence.

Definition 3. *Associated with each k-sequence, c_k, we define a clock sequence, denoted t_k, as the sequence satisfying*

$$t_k(n) = \min(kc_k(n-1) \bmod (n-1), n - kc_k(n-1) \bmod (n-1))$$

for $n \geq 3$ with $t_k(1) = t_k(2) = 1$.

Note that the clock, $t_k(n)$, starting at $n = 3$, tracks the term from the *lower* half of the sequence of length $n-1$ that is used to compute the nth term. In terms of its clock, a k-sequence can be written

$$c_k(n) = c_k(t_k(n-1)) + c_k(n - t_k(n-1)),$$

again where $n \geq 3$. A clock sequence becomes particularly useful when it become *periodic*. For example, consider the growth of c_2. We circle the terms at level $n-1$ used to create the new term at level n:

1, 1
1,1,2
1,1,2,3
1,1,2,3,3
1,1,2,3,3,4
1,1,2,3,3,4,4
1,1,2,3,3,4,4,5

The regular visual pattern translates into periodic behavior of t_2. In particular, $t_2 = 1, 1, 1, 1, 2, 1, 2, 1, 2, 1, 2, \ldots$. If we conjecture that this pattern continues, we may extract from t_2 the simpler set of *linear* recursions satisfied by c_2

$$c_2(2n) = 1 + c_2(2n-1), \tag{3}$$

$$c_2(2n+1) = 1 + c_2(2n-1), \tag{4}$$

from which the description of c_2 in our table and properties such as (4) and (5) of c_1 may easily be established. One might conjecture (wishfully) that a clock sequence that repeats itself continues to do so. Unfortunately, the clock of c_7 already furnishes a counterexample, repeating a portion of itself, and then wandering off into apparent aperiodic behavior. However, when a c_k does behavior in a regular, if perhaps complicated, fashion, clock sequences allow us to uncover this hidden structure. To search for this structure in a given c_k, we can plot the phase portrait of the associated clock sequence. Some sequences can yield visually appealing lengthy periodic behavior. The phase portraits showing the "beat" (the clock) of c_{16}, c_{260}, and c_{138} appear in Figures 1-3. Each of these phase portraits was drawn by computing the first ten thousand terms of the sequence and the associated clock sequence. Then, the next ten thousand terms of the associated clock sequence were plotted as points $(t_k(n), t_k(n+1))$. If the clock has become periodic by this point, revealing the underlying structure of the sequence, the phase portrait then reveals a closed orbit such as those in Figures 1-3. On the other hand, if the "beat" is still irregular, no apparent order is

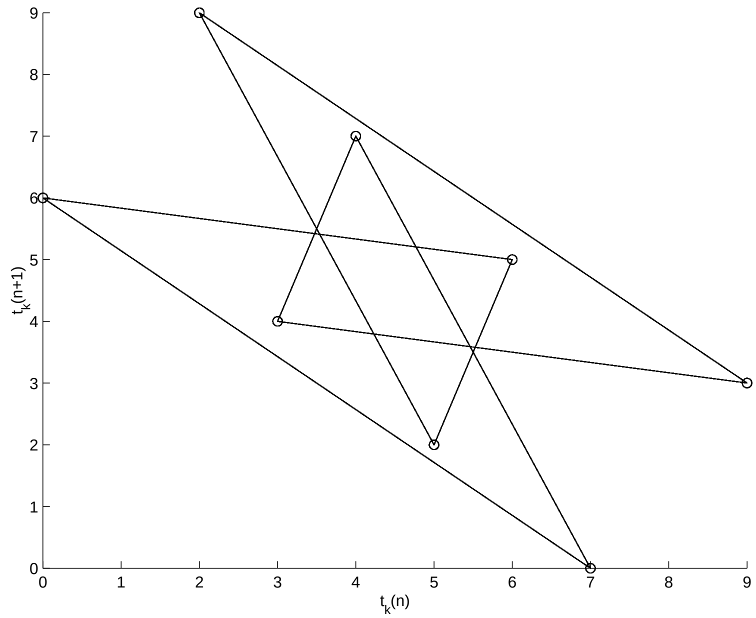

Figure 1: The ‘beat’ of c_{16}.

discernable in the phase portrait. Once the underlying structure is revealed, we may make conjectures concerning equivalences with known sequences or conjectures about the behavior of unknown sequences. These conjectures are then often easily proved (although when the period is long, the proof is tedious). As an easy example we have

Proposition 6. $c_4(n) = \lfloor \frac{n+1}{2} \rfloor$.

Proof. It is sufficient to prove that

$$c_4(n) = \lfloor \frac{n+1}{2} \rfloor.$$

We may easily verify that this is true for $n = 1$ to $n = 6$. Now, assume it is true for $k = 1 \ldots n$. Consider $c_4(n+1)$. We must show

$$c_4(n+1) = \lfloor \frac{n+2}{2} \rfloor.$$

But,

$$c_4(n+1) = c_4(4c_4(n) \bmod n) + c_4(n+1-4c_4(n) \bmod n).$$

By hypothesis

$$c_4(n) = \lfloor \frac{n+1}{2} \rfloor,$$

and hence

$$c_4(n+1) = c_4(4\lfloor \frac{n+1}{2} \rfloor \bmod n) + c_4(n+1-4\lfloor \frac{n+1}{2} \rfloor \bmod n).$$

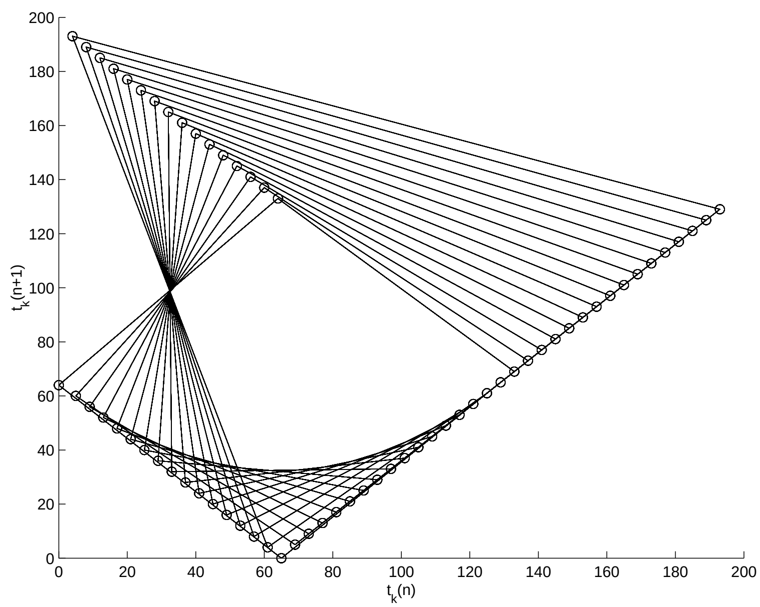

Figure 2: The 'beat' of c_{260}.

But, $4\lfloor \frac{n+1}{2} \rfloor$ mod n is 0 if n is even and 2 if n is odd. So,

$$c_4(n+1) = c_4(1) + c_4(n) = 1 + \lfloor \frac{n+1}{2} \rfloor,$$

for n even and

$$c_4(n+1) = c_4(2) + c_4(n-1) = 1 + \lfloor \frac{n}{2} \rfloor,$$

for n odd. From which it follows directly that

$$c_4(n+1) = \lfloor \frac{n+2}{2} \rfloor$$

as desired. □

Note that we implicitly used the clock sequence in this proof. In fact, the result can be restated as a result on the periodicity of t_4.

Randomly searching for k-sequences with the nice behavior of c_2, c_4 or c_{260}, we develop the feeling that many k-sequences are in fact irregular. To get a broader picture, we compute a "bifurcation diagram" for the c_k. For each k, we compute the first 5000 terms, of both c_k and t_k. Then, we compute the density of t_k in the interval $[0, 2500]$. Finally, we plot the negative log of this density versus k. Those sequences with highly ordered clocks, and hence a clear underlying structure, appear as peaks in this plot. Those with an irregular "beat" map roughly to zero. The bifurcation diagram for k ranging from one to one-thousand appears in Figure 4. Order appears to decrease with increasing k. Also the frequency of highly-ordered sequences appears to decrease with increasing k.

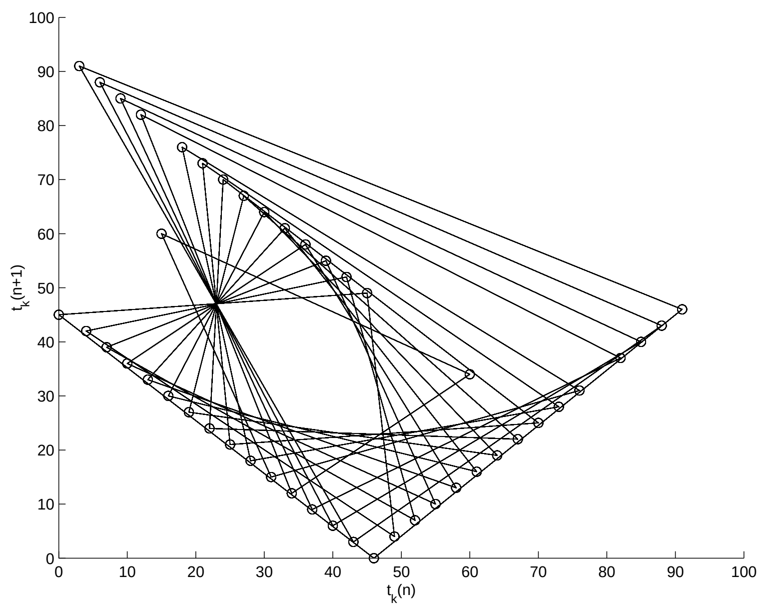

Figure 3: The 'beat' of c_{138}.

5 Open Questions and More Generalizations

We have just scratched the surface of k-sequences. Many open questions and further generalizations remain. One particularly intriguing puzzle concerns irregular sequences such as c_7 and c_{13}. Do the clocks of sequences such as c_7 or c_{13} ever become periodic or do they always beat irregularly? Does $c_k(n)/n$ tend to a limit for these sequences? Another question concerns sequences with "missing" numbers such as c_{11}. Computation reveals that c_{11} misses 11, 29, 33, 37, and 39. Does c_{11} miss infinitely many integers? What is the sequence of integers missed? Yet another less precise question concerns order. Is there a k-sequence that becomes irregular after a long period of regularity? (The reader may wish to examine c_{204} which exhibits the opposite behavior.) We may also ask: What other known sequences are lurking among the k's? Finally, we note that several authors have generalized the Conway-Hofstadter sequence in directions other than the one presented here. The generalization presented here however can be applied to those offered by Mallows [2], Newman [6], or Pinn [7]. For example, Mallows [2], introduces the sequence

$$c(n) = c(c(n-2)) + c(n - c(n-2)) \tag{5}$$

as a generalization of the Conway-Hofstadter sequence. This generalization bases the next term of the sequence on the second to last term of the sequence. Multiplying the $c(n-2)$ terms by k and computing modulo $n-1$, is a natural generalization in the spirit of the k-sequences introduced here. We hope the reader will be intrigued by the preliminary results presented in this paper and will be inspired to uncover new facts about the c_k's or generalize the work of Mallows, Newman, and Pinn, in the direction suggested here.

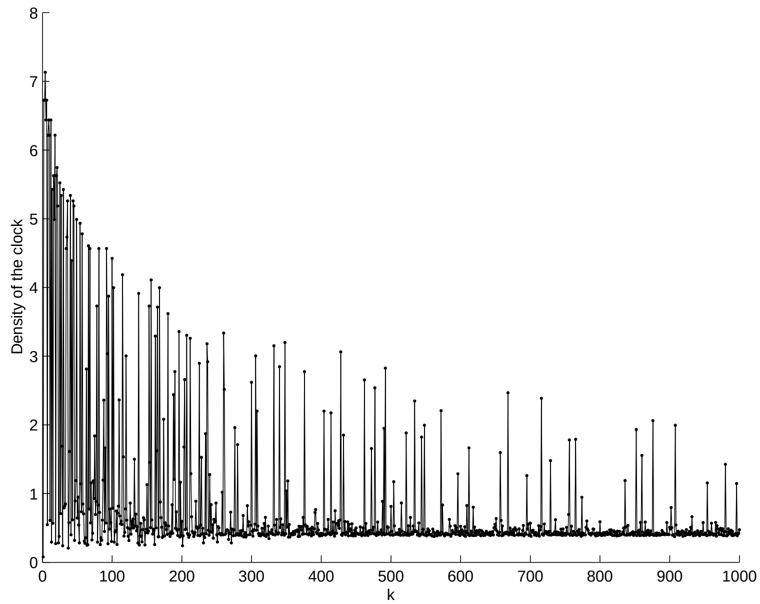

Figure 4: The bifurcation diagram for the c_k's. Those c_k's whose order is apparent by examining t_k appear as peaks.

Acknowledgment Thanks to Julia Pelesko whose interest in Logo led to this work and to M. Tempel whose article [8] first introduced us to the Conway-Hofstadter sequence. Thanks also to the anonymous referee for many useful comments and suggestions.

References

[1] J.H. Conway, Some Crazy Sequences, videotaped talk at AT&T Bell Labs, July 15, 1988.

[2] C.L. Mallows, Conway's Challenge Sequence, *Amer. Math. Monthly* **98** (1991), 5–20.

[3] M.W. Browne, Intellectual Duel: Brash Challenge, Swift Response, *The New York Times* Section C August 30 (1988), 1.

[4] T. Kubo and R. Vakil, On Conway's recursive sequence, *Disc. Math.* **152** (1996), 225–252.

[5] D.R. Hofstadter, Godel, Esher,Bach, Vintage Books, New York, 1980, 137.

[6] D. Newman, Problem E3274, *Amer. Math. Monthly* **95** (1988), 555.

[7] K. Pinn, A Chaotic Cousin of Conway's Recursive Sequence, *Exp. Math.* **9** (2000), 55–66.

[8] M. Tempel, Easy as 11223, *Online publications of the Logo Foundation*, el.media.mit.edu/logo-foundation/pubs/papers.

2000 *Mathematics Subject Classification*: Primary 11B37; Secondary 11B50.
Keywords: Conway-Hofstadter, Fibonacci sequence, nonlinear recursion.

(Concerned with sequences A004001, A004526, A004396, and A037915.)

Received January 19 2004; revised version received August 16 2004. Published in *Journal of Integer Sequences*, October 1 2004.

Return to Journal of Integer Sequences home page.