FIBONACCI WORDS - A SURVEY (*)

Jean Berstel
L. I. T. P.
Université Paris VI

INTRODUCTION

Fibonacci words have many amazing combinatorial properties. Like Fibonacci numbers they are easy to define, and many of their properties are easy to prove, once discovered. The aim of this survey is to sketch some of the combinatorial properties related to factors (subwords) of Fibonacci words, and also to describe basic arithmetic operations (i.e. normalization and addition) in the Fibonacci number system. No attempt was made to be complete.

Fibonacci words are easily defined by iterating a morphism. In fact, the Fibonacci morphism is among the absolutely simplest (more precisely shortest) conceivable morphisms ; discard the one letter alphabet, and try to define a nontrivial short morphism on two letters. It suffices, for this, that the image of one letter has length two, and you already get Fibonacci's morphism ! Fibonacci words also are "simple" because they have few subwords : as we shall see, Fibonacci words achieve the minimum for nonperiodic words. Despite of this weak number of subwords (or perhaps, on the contrary, it is a consequence of it) there are many repetitions in Fibonacci words : the number of repetitions grows like $n \log n$ with the length of the word. However, Fibonacci words do not contain high powers of words. They have cubes, but no fourth power.

Another topic that will be treated here is computation in Fibonacci base. Fibonacci numbers, as any regularly increasing sequence of natural numbers, are a candidate for a number system. Nonnegative integers are expressed as linear combinations of Fibonacci numbers, with coefficients 0 or 1. There exists a normalized representation computable by several types of transducers. Also addition and even weak addition can be described.

(*) Contrat ADI 83/695

1.- FIBONACCI WORDS, BEATTY SEQUENCES AND THE STURMIAN PROPERTY

Let $A = \{a, b\}$ be an alphabet. The Fibonacci morphism $\theta : A^* \to A^*$ is defined by

$$\theta(a) = ab$$
$$\theta(b) = a$$

Iteration of this morphism defines the Fibonacci words

$$f_0 = a$$
$$f_1 = ab$$
$$f_2 = aba$$
$$f_3 = abaab$$
$$f_4 = abaababa$$
$$f_5 = abaababaabaab$$

Since the DOL-system $\langle A, \theta, a\rangle$ is catenative [17], the sequence of Fibonacci words can also be defined by

$$f_0 = a,\ f_1 = ab$$
$$f_{n+2} = f_{n+1}f_n \quad (n \geq 0)$$

The infinite Fibonacci word

$$f = abaababaabaab\ldots$$

is obtained as a "limit" of the sequence $(f_n)_{n\geq 0}$, i.e. simply by requiring that each f_n $(n\geq 0)$ is a left factor of f.

For uniformity of exposition, the numbering of the letters in a (finite or infinite) word will start at 0. So the first letter of a word has index 0, and so on.

There is another definition of the Fibonacci word f related to the golden ratio

$$\Phi = (1 + \sqrt{5})/2$$

which is through Beatty's Theorem [1], see also Stolarsky [20].

BEATTY'S THEOREM - Let r and s be positive real numbers. The sets

$$\{[nr] - 1 \mid n \geq 1\} \quad \text{and} \quad \{[ns] - 1 \mid n \geq 1\}$$

form a partition of the set $\mathbb{N}$ or natural numbers iff r and s are irrational numbers and

$$1/r + 1/s = 1$$

Observe that ϕ and $\phi^2 = 1 + \phi$ satisfy the conditions of Beatty's Theorem, since they are irrational and

$$1/\phi + 1/\phi^2 = (\phi + 1)/\phi^2 = 1$$

PROPOSITION. - Let a be the letter at the k-th position in the infinite Fibonacci word f. Then

$$c = \begin{cases} a & \text{if } k \in \{[n\phi] \mid n \geq 1\} \\ b & \text{if } k \in \{[n\phi^2] \mid n \geq 1\} \end{cases}$$

Beatty's Theorem shows that the infinitive Fibonacci word is a very special case among an apparently nice family of infinite words. Let us give an alternative way of defining them. For this, consider the box in Figure 1 with sides of length 1. A billard ball starts at a fixed point P in a fixed direction given by the tangent, say α. Whenever it meets one of the sides of the box, the ball is perfectly reflected and continues its walk. Of course, if α is rational, the ray will eventually pass again through P. Thus assume α is irrational, and construct an infinite word on a, b as follows :
whenever the ball meets a horizontal side, write an "a", and when it meets a vertical side, write a "b". This defines an infinite word, say $w(P, \alpha)$. It is easily seen that $f = w(0, \phi)$.

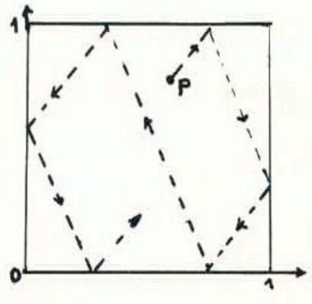


Figure 1. Defining the word abaaba...

A word $w(P,\alpha)$ defined by this geometrical construction with α irrational is called Sturmian (see Coven, Hedlund [4]). Words defined by Beatty's theorem are special cases of these words, with origin $P = 0$. For related topics, see Rauzy [16].

Sturmian words have two nice additional characterizations. First, consider any finite word w over the alphabet $A = \{a, b\}$, and defined its cost to be the absolute value of the difference of the number of a's and of the number b's occuring in w. Thus for instance, the cost of abaaba is 2. Then an infinite word is Sturmian if and only if it is not ultimately periodic, and if any two factors of the same length have costs which differ at most by 1.

The second characterization is through the number of factors. Given an infinite word x, denote by $F(x)$ the set of finite words having at least one occurence in x. For the Fibonacci word f, this set starts with the empty word ϵ, and contains

a, b	(length 1)
aa, ab, ba	(length 2)
aab, aba, baa, bab	(length 3)

and so on. Let $p_n(x)$ denote the number of factors of length n in x, i.e. $p_n(x) = \#(A^n \cap F(x))$. It is not too difficult to show that if x is not ultimately periodic, then $p_n(x) \geqslant n+1$ for all $n \geqslant 0$. Thus the minimum realizable for a nonperiodic word x is $p_n(x) = n+1$ for all n. This is precisely the characterization of Sturmian words.

THEOREM (Coven, Hedlund [4]) – An infinite word x over the alphabet $A = \{a, b\}$ is Sturmian if, and only if $p_n(x) = n+1$ for all $n \geqslant 0$.

2. FACTORS OF THE FIBONACCI WORD

As a consequence of Coven and Hedlund's theorem stated above, the infinite Fibonacci word f has exactly $n+1$ factors of length n, for all $n \geqslant 0$. It is interesting to know more on these factors.

PROPOSITION. - If w is a factor of f, then its reversal w^R also is a factor of f.

This is an immediate consequence of the following observation : consider any Fibonacci word f_n, and delete its two final letters. For $n = 5$, one obtains for instance

abaababaaba

Then the resulting word is palindrom. (A similar property is given in Knuth, Morris, Pratt [15], see also A. de Luca [7]).

PROPOSITION (Karhumaki [13]) - The Fibonacci word f has factors which are cubes, but no fourth power

Indeed, as underlined below, $(aba)^3$ is a factor of f.

f = abaababaabaababaababa...

The result can be strengthened as follows (the statement seems to be folklore, a proof can be found in Seebold [19]).

PROPOSITION - If u^2 is a factor of f, then u is conjugate to a Fibonacci word f_n.

(Two words u, v are called conjugate if they are cyclic permutations one of each other). Despite the facts that there are no fourth powers in f, and there are only few distinct factors in f, one has the astonishing

PROPOSITION (Crochemore [5]) - The number of occurences of maximal repetitions in a factor of length n of f grows like $n \log n$.

A repetition is a word of the form $u^k u'$, with $k \geqslant 1$ and u' a nonempty left factor of u. A repetition is maximal if it cannot be extended.

After these global results, we consider more specifically the structure of the factors of f. Consider an integer $n \geqslant 0$, and the $n+1$ factors of f of length n. Each of these factors can be extended to the right, and gives at least one factor of f of length $n+1$. However, since there are only $n+2$ factors of length $n+1$, only one of the length n factors can be extended to the right in two different ways. This factor will be called special : a factor w of f is <u>special</u> if wa, wb are both factors of f.

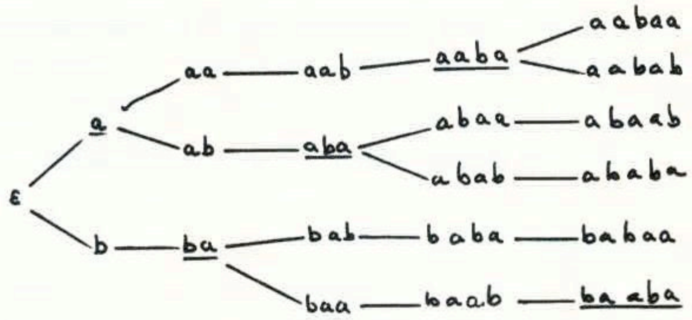


Figure 2. Factors of f, special factors are underlined.

As already stated, there is exactly one special factor of length n. Inspection of Figure 2 suggests

THEOREM – <u>The special factor of length</u> n <u>of</u> f <u>is the reversal of the left factor of length</u> n <u>of</u> f.

3. FIBONACCI NUMBER SYSTEM

One of the nicest applications of Fibonacci numbers is the Fibonacci number system. Many other number systems are described by Knuth [14]. More recent results on ambiguity in number systems are given by Culik, Salomaa [6], Honkala [12], De Luca, Restivo [8]. Generalizations of the Fibonacci arithmetic are investigated by Fraenkel [10] and Frougny [11].

The Fibonacci numbers are defined by

$$F_0 = 1, \; F_1 = 2, \; F_{n+2} = F_{n+1} + F_n \quad (n \geqslant 2)$$

ZECKENDORF'S THEOREM [21]. Every integer $n \geqslant 0$ admits a representation as sum of distinct Fibonacci numbers, i.e.

$$n = F_{k_T} + F_{k_{T-1}} + \ldots + F_{k_1} \quad (k_T > k_{T-1} > \ldots > k_1)$$

Furthermore, this representation is unique if, for each i, $k_{i+1} \geqslant k_i + 2$.

Many other related results are given by Carlitz [2] and Carlitz, Hoggatt, Scoville [3]. To any representation

$$n = F_{k_T} + \ldots F_{k_1}$$

we associate the word

$$a_{k_T} \ldots a_1 a_0$$

with

$$a_{k_T} = \ldots = a_{k_1} = 1, \; a_i = 0$$

otherwise. The representation satisfying the additional condition will be called the normalized representation and will be denoted by $\langle n \rangle$.

Example : for $n = 128$ (in base 10), the (words of the) Fibonacci representations are

1010001000
1010000110
1001101000
1001100110
1001011110
111101000
111100110
111011110

The first is the normalized representation ‹128›.

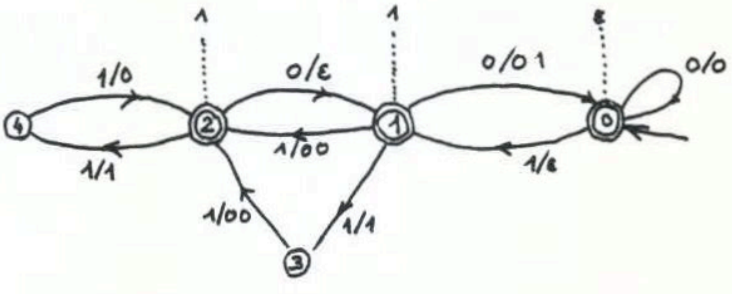
The relation between the infinite Fibonacci word and Fibonacci representation is the following (see e.g. Knuth [14]).

PROPOSITION - The letter in position n in the Fibonacci word f is a or b according to the word ‹n› finishes with 0 or 1.

This result has been considerably extended by Carlitz, Scoville and Hoggatt [3]. We are interested here in the complexity of constructing the normalized representation from a given one. The problem clearly consists of replacing adjacent "1" in a representation, more precisely of replacing a bloc 011 by a bloc 100 (and a leading bloc 11 by 100, but we may agree that a representation starts with enough leading 0 if necessary). Thus one has to compute a canonical element in the congruence class of a word, for the congruence generated by

$$011 = 100$$

It appears that this can be done by a finite transducer, but not by an arbitrary one. General theory of rational transductions says that an unambiguous transducer exists (Eilenberg [9]). The following subsequential transducer has been given by M.P. Schützenberger (private communication):



The input is read from right to left, and when it is completely read, the word on the dotted line is output at the end. Final states are doubly circled.

Example : Consider the word

1101110011

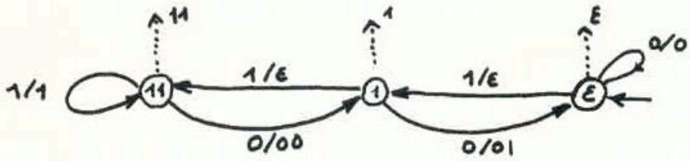
The path in the transducer is as follows

$$
\begin{array}{ccccccccccccccccccccc}
 & 1/00 & & 1/1 & & 0/\epsilon & & 1/00 & & 1/1 & & 1/\epsilon & & 0/01 & & 0/\epsilon & & 1/00 & & 1/\epsilon & \\
② & \leftarrow & ③ & \leftarrow & ① & \leftarrow & ② & \leftarrow & ③ & \leftarrow & ① & \leftarrow & ⓪ & \leftarrow & ① & \leftarrow & ② & \leftarrow & ① & \leftarrow & ⓪
\end{array}
$$

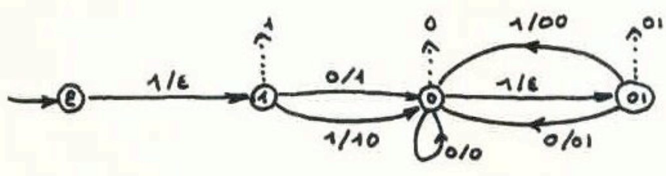
Thus the result is, as desired

10010010100

A general result on transductions says that a rational function is a composition of a left sequential and a right sequential function (Eilenberg [9]) and vice-versa. It is easily seen that a (left or right) sequential transduction cannot realize normalization, but very interesting and natural left and right sequential transducers have been given by J. Sakarovitch [18]. The idea is very natural : proceeding from left to right or from right to left, normalize as much as you can do sequentially. The amazing point is that it works. The right sequential transducer reads the word from right to left and outputs an intermediate word which is not yet completely normalized.

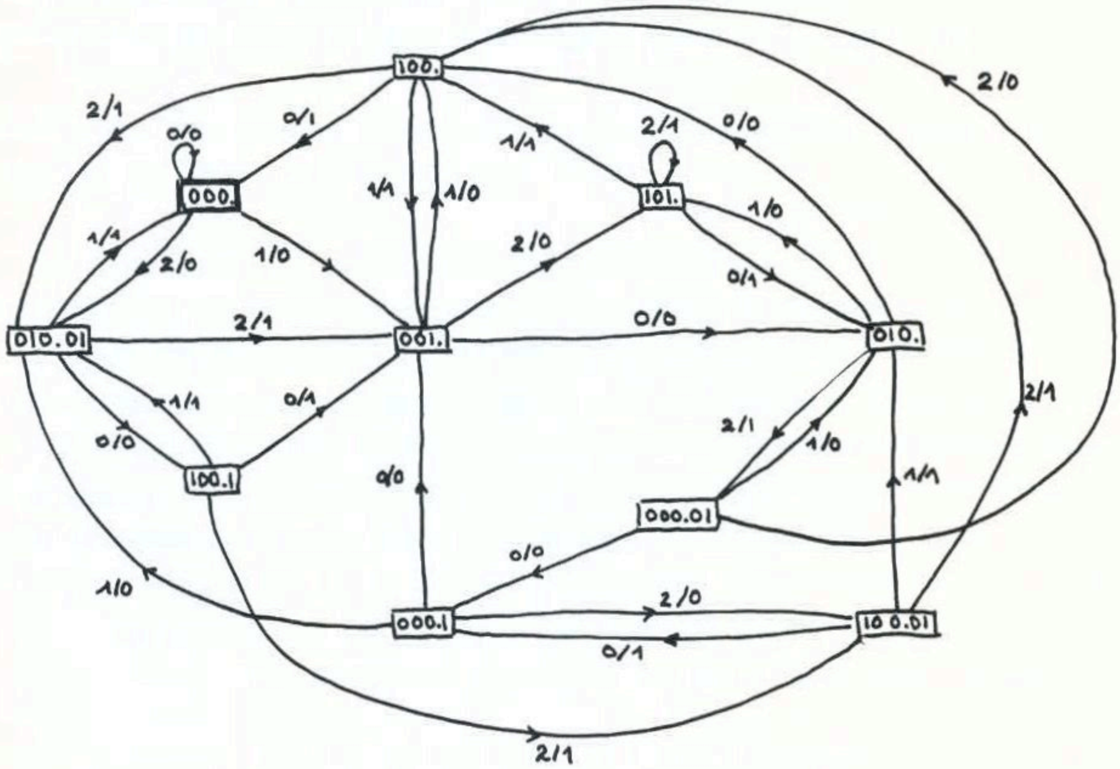


The left sequential transducer takes the word, reads it from left to right, and its output is normalized.



An wonderful property of these sequential transducers is that they can be applied in an arbitrary order and still produce the normalized representation !.

Addition is almost like normalization. Given two numbers represented in the Fibonacci number system, the first step for addition is to add the digits at the corresponding positions. This gives a sequence of 0, 1, and 2. This sequence is fed into the adder, which gives as output the corresponding sequence, written only with 0 and 1.



The automaton works as follows. The input word is read from left to right, starting in the initial state 000. For each input symbol, the corresponding output letter is written. When the end of the input is reached, the part of the state on the left of the dot is written on the output tape. There may be several useless leading zeros in the output word.

<u>Example</u> : In order to add the words 10110 and 11111, which are non normalized representations of 13 and 19, we first form their bitwise sum, namely 21221. This gives the following computation :

$$\boxed{000} \xrightarrow{2/0} \boxed{010.01} \xrightarrow{1/1} \boxed{000} \xrightarrow{2/0} \boxed{010.01} \xrightarrow{2/1} \boxed{001.} \xrightarrow{1/0} \boxed{100.}$$

Thus the resulting word is

01010100

which is indeed a representation of $13 + 19 = 32$.

For a proof of a more general case and for a systematic exposition, see Frougny [11].

WEAK ADDITION - The following method for "easy" addition of numbers is known in folklore as weak addition. Write integers n, m in base 10, but allow two additional digits to $\{0, 1, \ldots, 9\}$, namely 10 and 11. Then n, m may be represented as

$$n = n_t 10^t + n_{t-1} 10^{t-1} + \ldots + n_o$$

$$m = m_t 10^t + m_{t-1} 10^{t-1} + \ldots + m_o$$

with m_i, $n_i \in \{0, 1, \ldots, 10, 11\}$. Of course, this representation, called the weak representation, is by no means unique. For each index i, one has $0 \leqslant n_i + m_i \leqslant 22$, thus

$$n_i + m_i = p_{i+1} 10 + q_i \qquad (*)$$

with $p_{i+1} = 0, 1, 2$ and $0 \leq q_i \leq 9$. The sum $s = n + m$ then admits the expression

$$s = s_{t+1}10^{t+1} + s_t 10^t + \ldots + s_o$$

with

$$s_i = p_i + q_i \qquad (i = 0, \ldots, t+1)$$

Observe that there is no "carry" through several places. The ith "digit" s_i depends only on n_i, m_i, and on n_{i-1}, m_{i-1}.

Example

n =		10	11	3	10	11	11
m =		9	11	7	8	9	11
q =		9	2	0	8	0	2
p =	1	2	1	1	2	2	0
s =	1	11	3	1	10	2	2

The practical interest is in the fact that the computation of q, p and s can be performed in 2 cycles on a parallel computer, so addition can be very fast. The same method holds for any base k instead of 10, in the binary case, one needs one more row to compute the intermediate result.

A similar property also exists for Fibonacci addition. Consider a weak representation of an integer n to be

$$n = n_t F_t + n_{t-1} F_{t-1} + \ldots + n_o$$

with $n_o, \ldots, n_t$ no longer restricted to $\{0, 1\}$, but taken in a set $\{0, 1, \ldots, N\}$ for some $N \geq 1$. It is not too hard to see that these representations also can be normalized. In order to define weak addition, we need an analog of formula (*) in Fibonacci base. This formula will be an extension of the formula

$$2 F_n = F_{n+1} + F_{n-2}$$

In fact, a tedious but easy computation shows that for $n \geqslant 6$, and for any $d \in \{0, 1, \dots, 24\}$, one has an expression

$$dF_n = a_5^{(d)} F_{n+1} + \dots + a_{-6}^{(d)} F_{n-6}$$

with

$$a_5^{(d)}, \dots, a_{-6}^{(d)} \in \{0, 1\}$$

Consequently, weak addition in Fibonacci base exists provided the additional digits $\{2, \dots, 12\}$ are allowed.

ACKNOWLEDGEMENT - I thank Christiane Frougny for helpful discussions.

REFERENCES

[1] S. BEATTY, Problem 3173, Amer. Math. Monthly 33 (1926), p. 159 solutions ibid. 34 (1927), p. 159.

[2] L. CARLITZ, Fibonacci representations, Fib. Quart. (1968), p. 193-220

[3] L. CARLITZ, V.L. HOGGATT Jr, R. SCOVILLE, Fibonacci representations, Fib. Quart. 10 (1972), p. 1-28.

[4] E. COVEN, G. HEDLUND, Sequences with minimal block growth, Math. Syst. Theory 7 (1973), p. 138-153.

[5] M. CROCHEMORE, An optimal algorithm for computing repetitions of a word, Inf. Process. Letters 12 (1981), p. 244-250.

[6] K. CULIK II, A. SALOMAA, Ambiguity and decision problems concerning number systems, Information and Control 56, 1984, 139-153.

[7] A. de LUCA, A combinatorial property of the Fibonacci words, Inf. Process. Letters 12 (1981), p. 193-195.

[8] A. de LUCA, A. RESTIVO, Representations of integers and Language Theory, MFCS, 1984, Lecture Notes Comp. Sciences 176, 407-415.

[9] S. EILENBERG, "Automata, Languages and Machines", Academic Press 1974.

[10] A.S. FRAENKEL, Systems of Numeration, American Math. Monthly, to appear.

[11] C. FROUGNY, On linear Numeration Systems, in preparation.

[12] J. HONKALA, Bases and ambiguity of number systems, Theoret. Comput. Sci., 31, 1984, 61-71.

[13] J. KARHUMAKI, On cube-free ω-words generated by binary morphisms, Discr. Appl. Math. 5 (1983), p. 279-297.

[14] D. KNUTH, "The Art of Computer Programming", Vol. I, Addison Wesley 1968.

[15] D. KNUTH, J. MORRIS, V. PRATT, Fast pattern matching in strings, SIAM J. Comput. 6 (1977), p. 323-350.

[16] G. RAUZY, Mots infinis en arithmétique, in M. Nivat, D. Perrin (eds), "Automata on infinite words", Lecture Notes in Computer Science 192, 1985, Springer, p. 165-171.

[17] G. ROZENBERG, A. SALOMAA, "The Mathematical Theory of L Systems", Academic Press, 1980.

[18] J. SAKAROVITCH, Easy multiplication, Rapport LITP 85-41, to appear.

[19] P. SEEBOLD, Propriétés combinatoires des mots infinis engendrés par certains morphismes, Thèse de 3ème cycle, Rapport LITP 85-16.

[20] K. B. STOLARSKY, Beatty sequences, continued fractions, and certain shift operators, Canad. Math. Bull 19 (1976), p. 473-482.

[21] E. ZECKENDORF, Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas, Bull. Soc. Royale des Sciences de Liège, 1972, p. 179-182.