Borders Of Fibonacci Strings

L.J. Cummings*
University of Waterloo

D. Moore
Curtin University of Technology

J. Karhumäki†
Turku University

ABSTRACT. We determine all borders of the $n\underline{th}$ Fibonacci string, $\mathbf{f_n}$, $n \geq 3$. In particular, we give two proofs that the longest border of $\mathbf{f_n}$ is $\mathbf{f_{n-2}}$. One proof is independent of the Defect Theorem.

1 Introduction

A **Fibonacci string** $\mathbf{f_n}$ is a string on abinary alphabet $A = \{a, b\}$ defined by

$$\mathbf{f_0} = b; \mathbf{f_1} = a; \mathbf{f_n} = \mathbf{f_{n-1}}\mathbf{f_{n-2}}, \forall\, n \geq 2. \tag{1}$$

Equivalently, the strings $\mathbf{f_n}$ may be defined as the **orbit** of a under the morphism $h : A^* \to A^*$ defined by

$$h : a \to ab \qquad b \to a. \tag{2}$$

An interesting facet of the structure of Fibonacci strings was shown by de Luca [8].

*Supported by the Natural Sciences and Engineering Research Council of Canada
†Supported by grant no. 11181 of the Academy of Finland

Proposition 1.1 (de Luca). *For all $n \geq 3$,*

$$\mathbf{f_n} = \alpha_n \mathbf{d_n}$$

where for each n, α_n is a palindrome and

$$\mathbf{d_n} = \begin{cases} ab & \text{if } n \text{ is even} \\ ba & \text{if } n \text{ is odd} \end{cases}.$$

Since $\mathbf{f_7}$ contains the cube $(aba)^3$, for each $n \geq 7$, $\mathbf{f_n}$ contains at least this cube. But one of us has shown that no Fibonacci string contains a fourth power [6].

The well-known Fibonacci numbers can be defined as the lengths of the Fibonacci strings: $F_n = |\mathbf{f_n}|$. The usual definition of these numbers is then immediate:

$$F_n = F_{n-1} + F_{n-2}, \forall\, n \geq 2. \tag{3}$$

It is well-known that $F_n \in \Omega(\phi^n)$ where ϕ is the "golden ratio" satisfying the quadratic equation $\phi^2 = \phi + 1$. It is less widely known that the Fibonacci numbers are the basis of a number system [2, 9].

Theorem 1.1 (Zeckendorf). *Every positive integer n can be represented as a sum of distinct Fibonacci numbers*

$$F_n = F_{k_r} + F_{k_{r-1}} + \cdots + F_1, \tag{4}$$

where $k_r \geq k_{r-1} \geq \cdots \geq 1$.

The representation (4) is unique if $k_{i+1} \geq k_i + 2$, for all i.

A **border** of a string **x** is a substring which is both a prefix and a suffix of **x**. The longest border of a string **x** is often called *the* border of **x**. Arguably the empty string and **x** itself are borders of **x** but we exclude these trivial cases although it can be convenient to include the empty string as a border when initializing variables in an algorithm.

More formally,

Definition 1.1: If for some integer b, $0 < b < n$, a given string **x** of length n has a proper prefix $\mathbf{b} = \mathbf{x}[1..b]$ that is equal to a suffix $\mathbf{x}[n-b+1..n]$, then **b** is said to be a **border** of **x**. If **x** has a border then its longest border is denoted by $\mathbf{b}^*(\mathbf{x})$.

According to Definition 1 a string is never its own border. Clearly any border **b** of a string **x** is both a prefix and a suffix of $\mathbf{b}^*(\mathbf{x})$. In notation: $\mathbf{b} \preceq \mathbf{b}^*(\mathbf{x})$ and $\mathbf{b}^*(\mathbf{x}) \succeq \mathbf{b}$. We let $|\mathbf{x}|$ denotes the length of the string **x**. We will show that if $n \geq 3$ then $\mathbf{b}^*(\mathbf{f_n}) = \mathbf{f_{n-2}}$.

For example the string $f_6 = abaababaabaab$ has two borders: ab and $abaab$. Similarly, the string a^n has borders $a, a^2, \ldots, a^{n-1}$, of which, for $i = \lceil (n+1)/2 \rceil$, the borders $a^i, a^{i+1}, \ldots, a^{n-1}$ overlap.

It has been shown that the longest border of an arbitrary string can be computed in $\Theta(n)$ time[4]. The notion of a border is central to the pattern matching algorithms of Knuth-Morris-Pratt and Boyer-Moore. For an excellent survey see [1].

From Definition 1 we note the relationship between the periods of arbitrary strings and their borders. Namely, the integer p is a period of a string $\mathbf{x}$ if $p = n - |\mathbf{b}|$ where $\mathbf{b}$ is a border of $\mathbf{x}$. If $\mathbf{b}$ is the longest border of $\mathbf{x}$, then p is traditionally called *the period* of $\mathbf{x}$. Accordingly, by determining all borders of $\mathbf{f_n}$ we determine all periods of $\mathbf{f_n}$.

2 The Defect Theorem and $\mathbf{b}^*(\mathbf{f_n})$

The set of all strings with entries in an alphabet A is a free monoid or *code* denoted by A^*. Any *subset* X of A^* has an associated code, $\mathcal{F}(X)$, which is the intersection of all codes containing X. The Defect Theorem says that if X is not a code itself, then $|\mathcal{F}(X)| < |X|$. For a proof and discussion of the history of the Defect Theorem see [5]. Two strings $\mathbf{x}, \mathbf{y} \in A^*$ are *conjugate* if there exist $\alpha, \beta \in A^*$ such that

$$\mathbf{x} = \alpha\beta \tag{5}$$
$$\mathbf{y} = \beta\alpha. \tag{6}$$

One consequence of the Defect Theorem is that two strings $\mathbf{x}$, $\mathbf{y}$ are conjugate if and only if there is a third string $\mathbf{z}$ such that $\mathbf{xz} = \mathbf{zy}$. (See Proposition 1.3.4 of [7].)

Theorem 2.1. *If* $n \geq 3$ *then* $\mathbf{b}^*(\mathbf{f_n}) = \mathbf{f_{n-2}}$.

First Proof: Clearly, $\mathbf{f_{n-2}}$ is a border of $\mathbf{f_n}$ since $\mathbf{f_n} = \mathbf{f_{n-2}f_{n-3}f_{n-2}}$. Therefore, $\mathbf{f_{n-2}} \preceq \mathbf{b}^*(\mathbf{f_n})$ and $\mathbf{b}^*(\mathbf{f_n}) \succeq \mathbf{f_{n-2}}$. Equivalently, there are binary strings $\mathbf{x}, \mathbf{y}$ with $0 < |\mathbf{x}| = |\mathbf{y}| < |\mathbf{f_{n-1}}|$ such that

$$\mathbf{f_{n-2}x} = \mathbf{yf_{n-2}}.$$

The Defect theorem implies that $\mathbf{x}$ and $\mathbf{y}$ are conjugate. (See Proposition 1.3.4 of [7].) Let α and β be elements of A^* such that

$$\mathbf{x} = \alpha\beta \tag{7}$$
$$\mathbf{y} = \beta\alpha \tag{8}$$
$$\mathbf{f_{n-2}} = \beta(\alpha\beta)^t \tag{9}$$

for some $t \geq 0$.

The proof falls into two cases:

Case 1. $|\mathbf{x}| \leq |\mathbf{f_{n-3}}|$.

In this case, $\mathbf{x}$ is a prefix of $\mathbf{f_{n-3}}$ and $\mathbf{f_{n-3}} \preceq \mathbf{f_{n-2}}$, which implies $\mathbf{x} = \mathbf{y}$ since $\mathbf{f_{n-2}x} = \mathbf{yf_{n-2}}$. Therefore, $\alpha\beta = \beta\alpha$. It now follows by Proposition 1.3.2 of [7] that α and β and hence $\mathbf{f_{n-2}}$ are powers of a single primitive string, say $\mathbf{p}$. Since $\mathbf{f_{n-3}} \preceq \mathbf{f_{n-2}}$ and $\mathbf{y}^r \preceq \mathbf{f_{n-3}}^r$, (7) implies that $\mathbf{f_{n-3}}$ is also a power of $\mathbf{p}$. But then we have $\mathbf{f_k}$ is a power of $\mathbf{p}$ for all $k \geq n-3$, a contradiction.

Case 2. $|x| > |\mathbf{f_{n-3}}|$.

If as in (7) $\mathbf{f_{n-2}}$ is of the form $\beta(\alpha\beta)^t$ for some $t \geq 1$ then we can proceed as in Case 1 since $\mathbf{f_{n-2}} \preceq \mathbf{f_{n-3}f_{n-2}}$. (Note that $\mathbf{f_{n-2}} = \mathbf{f_{n-3}f_{n-4}}$ and $\mathbf{f_{n-4}} \preceq \mathbf{f_{n-2}}$.) Therefore, the only remaining case is when $\mathbf{f_{n-2}} = \beta$. In this case, $\mathbf{f_{n-1}} \preceq \beta\alpha\beta$ since $\mathbf{x} = \alpha\beta$ and $(\beta\alpha)^r \preceq \mathbf{f_{n-1}}^r$ where $|\beta\alpha| < |\mathbf{f_{n-1}}|$. It now follows that there exists a string $\mathbf{q}$ and $t \geq 1$ such that $\mathbf{f_{n-1}} = \mathbf{q}^t$. As in Case 1 this implies that every $\mathbf{f_k}$ is a power of $\mathbf{q}$ which is impossible.

3 Expansions and Borders

As noted before, for each $n \geq 3$ it is clear that $\mathbf{f_{n-2}}$ is a border of $\mathbf{f_n}$ since

$$\mathbf{f_n} = \mathbf{f_{n-2}f_{n-3}f_{n-2}}. \tag{10}$$

Proposition 1.1 can be applied immediately to equation (10) to conclude that, since $n-3$ and $n-2$ are of differing parity, the prefix $\mathbf{f_{n-2}f_{n-3}}$ cannot be a border of $\mathbf{f_n}$. We apply a similiar argument after each uniform expansion of (10) by the rules

$$\begin{array}{lcl} \mathbf{f_k} & \rightarrow & \mathbf{f_{k-2}f_{k-3}f_{k-2}} \\ \mathbf{f_2} & \rightarrow & \mathbf{f_1f_0} \\ \mathbf{f_1} & \rightarrow & \mathbf{f_1} \\ \mathbf{f_0} & \rightarrow & \mathbf{f_0}. \end{array} \tag{11}$$

For example, at level 2 of the expansion $\mathbf{f_{n-2}f_{n-3}f_{n-2}}$ becomes

$$\mathbf{f_{n-4}f_{n-5}f_{n-4}f_{n-5}f_{n-6}f_{n-5}f_{n-4}f_{n-5}f_{n-4}}. \tag{12}$$

The expansion terminates when only $\mathbf{f_0}$'s and $\mathbf{f_1}$'s remain. At level k of the expansion with (11) we have

$$\mathbf{f_{n-2k}f_{n-2k-1}f_{n-2k}} \cdots \mathbf{f_{n-2k}f_{n-2k-1}f_{n-2k}}. \tag{13}$$

By symmetry, each such expansion is a palindrome over an alphabet of appropriate $\mathbf{f_n}$. Therefore, each prefix consisting of, say, k consecutive $\mathbf{f_j}$'s is of the same length as the corresponding suffix of k consecutive f_j's. In the level 2 expansion (12), no prefix of f_j's ending in f_{n-5} can be a border of f_n. By induction, we see that at each step of the expansion, each f_k is adjacent only to f_{k+1} or to f_{k-1} on either the left or the right. When, say, $f_{k-1}f_k$ occurs in the expansion, it must be preceded by either f_k or f_{k-2}. In the second case, necessarily $f_{k-1}f_{k-2}f_{k-1}f_k$ must occur. This follows because the definition (11) precludes four consecutive f_i's with strictly increasing (or strictly decreasing) indices.

For each f_{n-j} in some level of the expansion (11) there is a potential border

$$f_{n-2k}f_{n-2k-1}f_{n-2k}\cdots f_{n-j} \quad = \quad f_{n-j}\cdots f_{n-2k}f_{n-2k-1}f_{n-2k}. \qquad (14)$$

Since f_{n-2} is a border of f_n, to show f_{n-2} is the longest border we need only rule out potential borders (14) which end to the right of the f_k which make up f_{n-2}.

If $n-j$ is even and n is odd then $n-2k$ is odd and (14) cannot be a border since f_{n-j} must end in *ab* while f_{n-2k} ends in *ba* by 1.1. Similiarly, if $n-j$ is odd and n is even then $n-2k$ is even and (14) cannot be a border since f_{n-j} must end in *ba* while f_{n-2} ends in *ab*. Consequently, at any level of expansion we need only consider those potential borders (14) for which $n-j$ and $n-2k$ have the same parity.

Lemma 3.1. *If a prefix* $f_{n-2k}f_{n-2k-1}f_{n-2k}\cdots f_{n-j}$ *of* f_n *has the property that it ends in a square of the form* $f_i^2 = f_if_i$ *for some* i *then it is not a border of* f_n.

Proof: If n and i are not both odd or both even then the last element of f_i cannot match the last element of f_n by (1.1). Accordingly, we need only consider the cases when i or n are both odd or both even.

Consider the successive expansions (13). The list of final f_i's from each level of the expansion is:

$$f_{n-2}, f_{n-4}, f_{n-6}, \cdots, f_{n-j}. \qquad (15)$$

If n is odd (even) then every odd (even) integer less than n appears in the list of indices in (15). Accordingly, given a prefix ending in a square $f_i^2 = f_if_i$, some level of expansion will also end in f_i. But by (11) f_i must be preceded by f_{i-1} and by (1.1), f_{i-1} and f_i have differing final entries.

Theorem 3.1. *For all* $n \geq 3$,

$$b^*(f_n) = f_{n-2}.$$

Proof: We argue that any potential border of the form

$$f_{n-2k}f_{n-2k-1}f_{n-2k}\cdots f_{n-j}$$

ends in exactly one of the following:

i) $f_{n-j+1}f_{n-j}$

ii) $f_{n-j-1}f_{n-j-2}f_{n-j-1}f_{n-j}$

iii) $f_{n-j}f_{n-j-1}f_{n-j}$.

For, as previously noted, each f_k is preceded only by f_{k+1} or f_{k-1}. If f_{n-j} is preceded by $f_{n-j+1} = f_{n-2k+1}$ as in i) then (1.1) implies (14) is not a border because the expansion at level k always terminates in $f_{n-2k}f_{n-2k-1}f_{n-2k}$. As a consequence, we may always suppose that f_{n-j} is preceded by f_{n-j-1}. In the same way, f_{n-j-1} is always preceded by f_{n-j} or f_{n-j-2}. If f_{n-j-1} is preceded by f_{n-j-2} as in ii) then (14) is not a border because the expansion at level k always terminates in $f_{n-2k}f_{n-2k-1}f_{n-2k}$.

Consider iii). Either f_{n-j-1} or f_{n-j+1} is the immediate predecessor of this substring. But, if the potential border (14) is

$$f_{n-j-1}f_{n-j}f_{n-j-1}f_{n-j} = (f_{n-j-1}f_{n-j})^2, \tag{16}$$

then Lemma 3.1 yields a contradiction.

In the same way, if the immediate predecessor of iii) is f_{n-j+1} then the potential border (14) ends in

$$\begin{aligned} f_{n-j+1}f_{n-j}f_{n-j-1}f_{n-j} &= f_{n-j}f_{n-j-1}f_{n-j}f_{n-j-1}f_{n-j} \\ &= f_{n-j}(f_{n-j-1}f_{n-j})^2, \end{aligned} \tag{17}$$

again contradicting Lemma 3.1.

The above arguments rule out any border of the form $f_{n-2}u$ of f_n, where u is a substring of length not more than $F_n - F_{n-2}$, since iterations of the expansion (11) terminate when every f_j has become $f_1 = a$ or $f_0 = b$.

This completes the proof.

Theorem 3.2. *For all $n \geq 3$, the borders of f_n are*

$$f_{n-2}, f_{n-4}, \cdots, f_k \tag{18}$$

where $k = 2$ if n is even and 1 otherwise.

Proof: By Theorem 3.1 f_{n-2} is the longest border of f_n. In the successive expansions of (11) we see that each of the Fibonacci strings in (18) are borders of f_n. Further, no Fibonacci string of the form $f_{n-(2k+1)}$ can be a border since, although each such is a prefix, it would then need to be a suffix of f_n but f_{n-2k} is also a suffix of f_n for all appropriate k, and $f_{n-(2k+1)}$, f_{n-2k} have different endings by (1.1).

References

[1] A. Aho, *Pattern Matching in Strings*, Handbook of Theoretical Computer Science, Elsevier,1990, 255–300.

[2] J. Berstel, *Fibonacci words — a survey*, Book of L, Springer-Verlag, 1986, 13–27.

[3] A. Ehrenfeucht and D. M. Silberger, *Periodicity and unbordered segments of words*, Discrete Mathematics, 26(1979), 101–109.

[4] A.V. Aho, J.E. Hopcroft and J.D. Ullman, *The Design and Analysis of Computer Algorithms*, Addison-Wesley,1974, 330–334.

[5] J. Berstel, D. Perrin, J.F. Perrot, and A. Restivo, *Sur le théorème du défault* Journal of Algebra, 60(1979), 169–180.

[6] J. Karhumäki, *On cube-free ω-words generated by binary morphisms*, Discrete Applied Mathematics, 5(1983), 279–297.

[7] M. Lothaire,*Combinatorics on Words* Addison-Wesley,1983.

[8] A. de Luca, *A combinatorial property of Fibonacci words*, Information Processing Letters, 12(1981),193–195.

[9] E. Zeckendorf, *Reprèsentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas*, Bull. Soc. Royale des Sciences de Liége, 1972, 179–182.