On Minimal Sturmian Partial Words*

Francine Blanchet-Sadri[1] and John Lensmire[2]

1 Department of Computer Science, University of North Carolina
P.O. Box 26170, Greensboro, NC 27402–6170, USA
blanchet@uncg.edu

2 Department of Mathematics, UCLA
Box 951555, Los Angeles, CA 90095, USA
jlensmire@gmail.com

Abstract

Partial words, which are sequences that may have some undefined positions called holes, can be viewed as sequences over an extended alphabet $A_\diamond = A \cup \{\diamond\}$, where $\diamond$ stands for a hole and matches (or is *compatible* with) every letter in A. The *subword complexity* of a partial word w, denoted by $p_w(n)$, is the number of distinct full words (those without holes) over the alphabet that are compatible with factors of length n of w. A function $f : \mathbb{N} \to \mathbb{N}$ is (k, h)-*feasible* if for each integer $N \geq 1$, there exists a k-ary partial word w with h holes such that $p_w(n) = f(n)$ for all n, $1 \leq n \leq N$. We show that when dealing with feasibility in the context of finite binary partial words, the only linear functions that need investigation are $f(n) = n + 1$ and $f(n) = 2n$. It turns out that both are $(2, h)$-feasible for all non-negative integers h. We classify all minimal partial words with h holes of order N with respect to $f(n) = n + 1$, called Sturmian, computing their lengths as well as their numbers, except when $h = 0$ in which case we describe an algorithm that generates all minimal Sturmian full words. We show that up to reversal and complement, any minimal Sturmian partial word with one hole is of the form $a^i \diamond a^j b a^l$, where i, j, l are integers satisfying some restrictions, that all minimal Sturmian partial words with two holes are one-periodic, and that up to complement, $\diamond(a^{N-1}\diamond)^{h-1}$ is the only minimal Sturmian partial word with $h \geq 3$ holes. Finally, we give upper bounds on the lengths of minimal partial words with respect to $f(n) = 2n$, which are tight for $h = 0, 1$ or 2.

1998 ACM Subject Classification F.2.2; F.4.3

Keywords and phrases Automata and formal languages; Combinatorics on words; Graph theory; Subword complexity; Feasible functions; Partial words; Sturmian words.

Digital Object Identifier 10.4230/LIPIcs.STACS.2011.225

1 Introduction

Let A be a k-letter alphabet and w be a finite or infinite word over A. A *subword* or *factor* of w is a block of consecutive letters of w. The *subword complexity* of w is the function which assigns to each integer n, the number, $p_w(n)$, of distinct subwords of length n of w. The subword complexity of finite and infinite words has become an important topic in combinatorics on words. Application areas include dynamical systems, ergodic theory, and theoretical computer science. Infinite words achieving various subword complexities have been widely studied: $p_w(n) = n + 1$ [13, 11], $p_w(n) = 2n$ [14], $p_w(n) = 2n + 1$ [4], to name a few (see Allouche [2] and Ferenczi [9] for some surveys). Chapter 10 of Allouche and Shallit's

* This material is based upon work supported by the National Science Foundation under Grant No. DMS–0754154. The Department of Defense is also gratefully acknowledged.

28th Symposium on Theoretical Aspects of Computer Science (STACS'11).
Editors: Thomas Schwentick, Christoph Dürr; pp. 225–236
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

book [3] provides a good overview for subword complexity of finite and infinite words. Our focus in this paper is on finite words.

Motivated by molecular biology of nucleic acids, Berstel and Boasson introduced partial words which are finite sequences that may have some undefined positions called holes (a (full) word is just a partial word without holes) [5]. Partial words can be viewed as sequences over an extended alphabet $A_\diamond = A \cup \{\diamond\}$, where $\diamond \notin A$ stands for a hole. Here $\diamond$ matches (or is *compatible* with) every letter in the alphabet. In this context, $p_w(n)$ is the number of distinct full words over the alphabet that are compatible with factors of length n of the partial word w (if $A = \{a, b\}$ and $w = a\diamond abaa$, then $p_w(3) = 5$ since aaa, aab, aba, baa and bab match factors of length 3 of w). Manea and Tiseanu showed that computing the subword complexity of partial words is a “hard” problem [12].

In this paper, we investigate minimal partial words with given subword complexity. This was done for a particular case of full words in [16]. There, it was shown that the minimal length of a word w such that $p_w(n) = F_{n+2}$ for all n, $1 \leq n \leq N$ is $F_N + F_{N+2}$, where $(F_n)_{n\geq 1}$ is the Fibonacci sequence and N is a positive integer, and an algorithm was given for generating such minimal words for each $N \geq 1$.

A function $f : \mathbb{N} \to \mathbb{N}$ is called (k, h)-*feasible* if for each integer $N \geq 1$, there exists a k-ary partial word w with h holes such that $p_w(n) = f(n)$ for all n, $1 \leq n \leq N$. In this case, w is an f-complex k-ary partial word with h holes of order N. Note that this is equivalent to saying there exists an integer N_0 such that for each $N \geq N_0$, there exists a k-ary partial word w with h holes such that $p_w(n) = f(n)$ for all n, $1 \leq n \leq N$. If f is a feasible function, it is immediate that f is non-decreasing and let us denote the length of a shortest f-complex k-ary partial word w with h holes of order N (called *minimal*) by $L_k(f, N, h)$. Similarly, denote the number of such minimal partial words by $N_k(f, N, h)$.

First, let us consider functions of the form $f(n) = k^n$, where k is the alphabet size. When we restrict our attention to the case of $h = 0$, a k-ary de Bruijn sequence of order N is a full word over a k-letter alphabet A where each of the k^n full words of length n over A appears as a factor exactly once. It is well known that $L_k(k^n, N, 0) = k^n + n - 1$. Moreover, $N_k(k^n, N, 0) = k!^{k^{n-1}}$, and these sequences can be efficiently generated by constructing Eulerian cycles in corresponding de Bruijn directed graphs. The technical report of de Bruijn provides a history on the existence of these sequences [8]. De Bruijn graphs find applications, in particular, in genome rearrangements [1], etc.

In [7], the case of $h > 0$ was initiated. For positive integers N, h and k, Blanchet-Sadri et al. introduced the concept of a de Bruijn partial word of order N with h holes over an alphabet A of size k, as being a partial word w with h holes over A of minimal length with the property that $p_w(n) = k^n$. There, the authors gave lower and upper bounds on $L_k(k^n, N, h)$, and showed that their bounds are tight when $h = 1$ and $k \in \{2, 3\}$. They provided an algorithm to construct 2-ary de Bruijn partial words with one hole of order N. Finally, they showed how to compute $N_2(2^n, N, 1)$ by adapting the so called BEST theorem that counts the number of Eulerian cycles in directed graphs [15].

Now, let us look at constant functions over the binary alphabet $\{a, b\}$. Note that $f \equiv 1$ is $(2, 0)$-feasible, and that a^N and b^N are the only minimal f-complex full words of order N (so that $L_2(1, N, 0) = N$ and $N_2(1, N, 0) = 2$). Furthermore, $f \equiv 1$ is not $(2, h)$-feasible for any $h \geq 1$, as any $\diamond$ in a partial word w implies that $2 = p_w(1) = f(1)$. Note also that $f \equiv 2$ is $(2, 0)$- and $(2, 1)$-feasible, but not $(2, h)$-feasible for $h \geq 2$. To see this, words of the form ab^N and $\diamond a^{N-1}$ show that f is $(2, 0)$- and $(2, 1)$-feasible respectively. Furthermore, these words are minimal and unique up to reverse and complement. Thus, $L_2(2, N, 0) = N + 1$, $L_2(2, N, 1) = N$, and $N_2(2, N, 0) = N_2(2, N, 1) = 4$. Now suppose that a word w has at least two holes. If w has two consecutive holes, note that $p_w(2) = 4$. If the holes are spread out, e.g. both $\diamond c$ and $d\diamond$ are factors of w for some letters $c, d \in \{a, b\}$, then $p_w(2) \geq 3$.

In this paper, let us investigate linear functions for binary partial words. It is obvious that if $f(1) = 1$, then $f \equiv 1$. Thus, when characterizing linear functions f, we only need to look at the case when $f(1) = 2$, that is, $f(n) = pn + q$ for integers p, q such that $p + q = 2$ and $p > 0$. Note that if $p > 2$, then $f(2) > 4$. Thus, the only linear options are $f(n) = n + 1$ or $f(n) = 2n$. The contents of our paper is as follows: In Section 2, we review some basics on partial words. We also give a bound on the subword complexity of any binary partial word with h holes. In Section 3, we show that the linear function $f(n) = n + 1$ is $(2, h)$-feasible for all non-negative integers h, and we consider $(n + 1)$-complex partial words referred to as *Sturmian*. We classify all minimal Sturmian partial words with h holes of order N, computing the exact length $L_2(n + 1, N, h)$ as well as the exact number $N_2(n + 1, N, h)$, except for $N_2(n + 1, N, 0)$. Instead of computing the latter, we describe an algorithm that generates all Sturmian full words of order N. We show that any minimal Sturmian partial word with one hole is of the form $a^i \diamond a^j b a^l$ (up to reversal and complement), where i, j, l are integers satisfying some restrictions, that all minimal Sturmian partial words with two holes are one-periodic, and that up to complement, $\diamond(a^{N-1}\diamond)^{h-1}$ is the only minimal Sturmian partial word with $h \geq 3$ holes. Finally in Section 4, we prove that the linear function $f(n) = 2n$ is also $(2, h)$-feasible for all non-negative integers h, and we conclude with some results on $2n$-complex partial words.

2 Preliminaries

We recall some basic terminology and notation on partial words that are useful throughout the paper. For more background, we refer the reader to [6].

Let A be a nonempty finite set of symbols called an *alphabet*. Each element $a \in A$ is called a *letter*. A *partial word* over A is a finite sequence of symbols from the alphabet enlarged with the *hole* symbol, $A_\diamond = A \cup \{\diamond\}$, where a *(full) word* is a partial word which does not contain any $\diamond$'s. The *length* of a partial word u is denoted by $|u|$ and represents the number of symbols in u. The *empty word* has length zero and is denoted by ε. If $\mathcal{S}$ is a set of partial words, $\|\mathcal{S}\|$ denotes its cardinality.

We denote by $u(i)$ the symbol at position i of the partial word u, the labelling of the positions starting at 0. Position i in u is in the *domain* of u, denoted by $D(u)$, if $u(i) \in A$. Otherwise if $u(i) = \diamond$, position i belongs to the *set of holes* of u. A positive integer p is called a *period* of a partial word u if $u(i) = u(j)$ whenever $i, j \in D(u)$ and $i \equiv j \bmod p$. In such a case, we call u *p-periodic*. The powers of u are defined recursively by $u^0 = \varepsilon$ and for $n \geq 1$, $u^n = uu^{n-1}$.

A *completion* of a partial word w over A is a full word $\hat{w}$ constructed by filling in the holes of w with letters from A. If u and v are two partial words of equal length, then u and v are *compatible*, denoted by $u \uparrow v$, if $u(i) = v(i)$ whenever $i \in D(u) \cap D(v)$, that is there exist completions $\hat{u}, \hat{v}$ such that $\hat{u} = \hat{v}$.

A partial word u is a *factor* of a partial word v if there exist partial words x, y such that $v = xuy$. We adopt the notation $v[i..j)$ to denote the factor $v(i) \cdots v(j - 1)$ of v. Here u is a *prefix* of v if $x = \varepsilon$ and a *suffix* of v if $y = \varepsilon$. A full word u is a *subword* of a partial word w if $u \uparrow v$ for some factor v of w. Informally, u is a subword of w if there is some completion $\hat{w}$ such that u is a factor of $\hat{w}$. Note that subwords in this paper are always full. We let $\text{Sub}_w(n)$ denote the set of all subwords of w of length n, and we let $\text{Sub}(w) = \bigcup_{0 \leq n \leq |w|} \text{Sub}_w(n)$, the set of all subwords of w. Note that if $\hat{w}$ is a completion of w, then $p_{\hat{w}}(n) \leq p_w(n)$, since

$\text{Sub}_{\hat{w}}(n) \subset \text{Sub}_w(n)$.

We end this section by giving a bound on the subword complexity of any binary partial word w. Let $n \leq |w|$ be a positive integer. A factor u of length n of w is repeated, if there exist integers $i \neq j$ such that $u = w[i..i+n) = w[j..j+n)$. Similarly, a subword u of length n of w is repeated, if there exist integers $i \neq j$ such that $u \uparrow w[i..i+n)$ and $u \uparrow w[j..j+n)$. Note that repeated factors imply repeated subwords, but the converse does not hold in general.

▶ Proposition 1. Let w be a partial word with h holes over a binary alphabet. For index $i = 0, \dots, h$ and positive integer $n \leq |w|$, let $\mathcal{F}_i(w, n)$ denote the multiset containing the factors of w of length n with exactly i holes. Then

$$\sum_{i=0}^{h} \|\mathcal{F}_i(w, n)\| = |w| - n + 1 \tag{1}$$

$$p_w(n) \leq \sum_{i=0}^{h} 2^i \|\mathcal{F}_i(w, n)\| \tag{2}$$

with equality holding in (2) if and only if w has no repeated subwords of length n. The following zero-hole and one-hole bounds hold:

1. Let $h = 0$. For $n \leq |w|$, we have $p_w(n) \leq |w| - n + 1$, with equality holding if and only if w has no repeated subwords of length n.
2. Let $h = 1$ and $n \leq |w|$. If $|w| \leq 2n - 1$, then $p_w(n) \leq 2(|w| - n + 1)$. Else, $p_w(n) \leq |w| + 1$. In both cases, equality holds if and only if w has no repeated subwords of length n.

Proof. For Statement (2), Inequality (2) implies that $p_w(n) \leq \|\mathcal{F}_0(w, n)\| + 2\|\mathcal{F}_1(w, n)\|$ with equality holding if and only if w contains no repeated subwords of length n. First suppose that $|w| \leq 2n - 1$. In this case, it is possible that w satisfies $\mathcal{F}_0(w, n) = \emptyset$. Note that since Equality (1) holds, this situation maximizes the subword complexity. Therefore, $p_w(n) \leq 2\|\mathcal{F}_1(w, n)\| = 2(|w| - n + 1)$. Now suppose that $|w| > 2n - 1$. We have $\|\mathcal{F}_1(w, n)\| \leq n$. If $\|\mathcal{F}_1(w, n)\| = n$, then $\|\mathcal{F}_0(w, n)\| = |w| - 2n + 1$. Thus, $p_w(n) \leq |w| - 2n + 1 + 2n = |w| + 1$ as desired. ◀

3 Sturmian partial words

In this section, we investigate Sturmian partial words. Recall that a finite partial word w is called Sturmian of order N if $p_w(n) = n + 1$ for all n, $1 \leq n \leq N$. We will fill out Table 1, whose first three columns show that for $h \geq 0$, $f(n) = n + 1$ is $(2, h)$-feasible.

▶ Remark. Note that the lengths of the words in the third column of Table 1 give upper bounds on $L_2(n + 1, N, h)$, listed in the fourth column. For $N \geq 1$, let $w = a^N b^N$. By Proposition 1(1), a word z must have length $l \geq 2N$ to satisfy $p_z(N) \geq N + 1$. Thus, w is a minimal $(n + 1)$-complex partial word of order N, and so $L_2(n + 1, N, 0) = 2N$.

Now for $N \geq 6$, let $w = a^{\lfloor N/2 \rfloor} \diamond a^{\lfloor N/2 \rfloor} b a^{\lceil (N-4)/2 \rceil}$. By Table 1, w is an $(n + 1)$-complex partial word of order N with $|w| = \frac{3N}{2}$ when N is even, and $|w| = \frac{3N}{2} - \frac{1}{2}$ when N is odd. By Proposition 1(2), a word z with one hole must have length $l \geq \frac{3N}{2} - \frac{1}{2}$ to satisfy $p_z(N) \geq N + 1$, implying that w is minimal, and so $L_2(n + 1, N, 1)$ is as shown in the table.

As is proved later, the other upper bounds also turn out to be lower bounds.

Table 1 Sturmian partial words with h holes of order N

h	N	partial word	$L_2(n+1,N,h)$	$N_2(n+1,N,h)$
0	≥ 1	$a^N b^N$	$2N$	
1	≥ 6	$a^{\lfloor N/2\rfloor}\diamond a^{\lfloor N/2\rfloor}ba^{\lceil (N-4)/2\rceil}$	$\frac{3N}{2}$ if N is even	12 if N is even
			$\frac{3N}{2}-\frac{1}{2}$ if N is odd	4 if N is odd
2	≥ 12	$a^{\lfloor (N-6)/2\rfloor}\diamond a^{N-5}\diamond a^{\lceil (N-6)/2\rceil}$	$2N-9$	$2N-22$
≥ 3	$\geq h+1$	$\diamond(a^{N-1}\diamond)^{h-1}$	$N(h-1)+1$	2

3.1 The case of $h = 0$

The aim of this section is to provide an algorithm that generates all minimal Sturmian full words. In constructing them, some graph theory is useful (the reader is refereed to [10] for more information).

Let $G = (V, E)$ be a directed graph. The *line digraph* of G, denoted by $L(G)$, is the graph $G' = (V', E')$ where $V' = E$, and for all $v'_1, v'_2 \in V'$, $(v'_1, v'_2) \in E'$ if $v'_1 = (v_1, v_2)$ and $v'_2 = (v_2, v_3)$ for some $v_1, v_2, v_3 \in V$. Combining ideas from de Bruijn and Rauzy graphs, we define a labelled directed graph $G_{\mathcal{S}} = (V, E)$ on a set $\mathcal{S}$ of words of length n as follows: V consists of the set of factors of length $n-1$ of words in $\mathcal{S}$ and E consists of the set of edges (x, x') for which there exists $y \in \mathcal{S}$ such that x is a prefix of y and x' is a suffix of y (such edges are labelled by y). This definition provides us with a natural correspondence between graphs and words.

▶ **Lemma 3.1.** *Given a set $\mathcal{S}$ consisting of words of length n, there exists a word w such that $Sub_w(n) = \mathcal{S}$ if and only if $G_{\mathcal{S}} = (V, E)$ has a path that includes all of the edges of $G_{\mathcal{S}}$. If such a path p exists, then there exists a word w of length $|p| + n - 1$ with $Sub_w(n) = \mathcal{S}$. Furthermore, $Sub_w(n-1) \supset V$.*

The following properties of a directed graph $G = (V, E)$ are well known and are useful throughout this section. The notation $\texttt{ideg}(v)$ refers to the in-degree of vertex v, $\texttt{odeg}(v)$ to its out-degree, and $(\texttt{ideg}(v), \texttt{odeg}(v))$ to its degree.

1. The size of the line digraph $L(G) = (V', E')$ of G is $|V'| = |E|$ and $|E'| = \sum_{v\in V} \texttt{ideg}(v) \times \texttt{odeg}(v)$.
2. The graph G has an Eulerian circuit if and only if G is strongly connected and for every vertex $v \in V$, $\texttt{ideg}(v) = \texttt{odeg}(v)$.
3. If $x, y \in V$ are such that $\texttt{odeg}(x) = \texttt{ideg}(x) + 1$ and $\texttt{ideg}(y) = \texttt{odeg}(y) + 1$, then G has an (x, y)-Eulerian path (or an Eulerian path from x to y) if and only if G is weakly connected and for every vertex $v \in V \setminus \{x, y\}$, $\texttt{ideg}(v) = \texttt{odeg}(v)$.

We call a directed graph G *Sturmian of order* n if G has n vertices, $n + 1$ edges, and contains an Eulerian path. The graph G is *Sturmian Type I or II* if G has degree sequence $(2, 2), (1, 1), \ldots, (1, 1)$ or $(2, 1), (1, 2), (1, 1), \ldots, (1, 1)$ respectively.

▶ Proposition 2. **1.** Suppose that $G = (V, E)$ is Sturmian Type II of order n. Then $L(G)$ is Sturmian of order $n + 1$.

2. Suppose that $G = (V, E)$ is Sturmian Type I of order n. Then it is possible to remove one edge from $L(G)$ to get G', where G' is Sturmian of order $n + 1$. Furthermore, it is impossible to remove an edge from $L(G)$ to get a graph G' such that G' contains a path that contains all of the edges of G' and G' is not Sturmian.

Proof. For Statement (2), note that $L(G)$ has $n+1$ vertices and $n+3$ edges. Since G contains an Eulerian path, $L(G)$ has a Hamiltonian path, and thus is weakly connected. Thus, we are left to show that we can remove one edge from $L(G)$ to get a graph G' that is still weakly connected and contains an Eulerian path. The graph G being Sturmian Type I, there is a distinct vertex v that has degree $(2,2)$. Label the edges in and out of v as i_1, i_2 and o_1, o_2 respectively. Note that all the vertices in $L(G)$ not in $S = \{i_1, i_2, o_1, o_2\}$ have degree $(1,1)$. Two cases remain which are illustrated in Figure 1: Case (i) where each member in S is distinct, and Case (ii) where $i_2 = o_2$.

For Case (i), i_1, i_2 have degree $(1,2)$ while o_1, o_2 have degree $(2,1)$. Note that there are edges from i_j to o_l for each $j, l \in \{1,2\}$. Remove the edge (i_2, o_2) to get a graph G'. Note that G' is still weakly connected. Furthermore, in G', i_1 has degree $(1,2)$, o_1 has degree $(2,1)$, and all other vertices have degree $(1,1)$. Thus, G' has an Eulerian path and is Sturmian Type II. Note that removing any of the edges (i_j, o_l) can be handled similarly. Further, note that removing any other edge from $L(G)$ results in a graph that no longer has a path containing all the edges. ◀

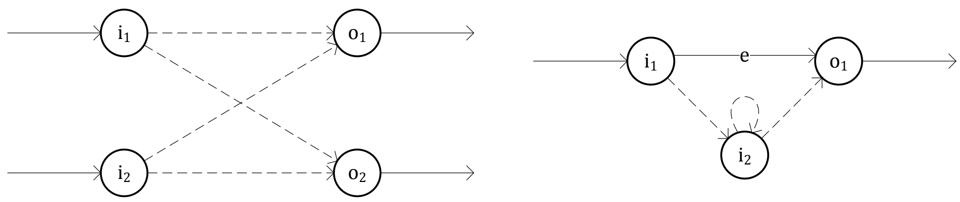

Figure 1 Part of $L(G)$ in Proposition 2(2): Left: Case (i); Right: Case (ii).

We are now ready to present an algorithm (similar to one used by Rote in [14]) to generate minimal Sturmian full words. Note that Proposition 2 implies that the graph G' created in line 2, 6, or 8 is always Sturmian. Since G'_N has $N+1$ edges, Algorithm 1 generates a minimal Sturmian word.

Algorithm 1 Constructing a minimal Sturmian full word of order $N \geq 3$.

1: Create $G_2 = G_{\mathcal{S}}$, where $\mathcal{S} = \{aa, ab, ba, bb\}$
2: Create G'_2 by deleting an edge from G_2
3: **for** $i = 3$ to N **do**
4: Build $G_i = L(G'_{i-1})$
5: **if** G_i has $i+2$ edges **then**
6: Create G'_i by deleting an edge from G_i (so that G'_i has $i+1$ edges), but ensure that G'_i still contains an Eulerian path
7: **else**
8: Set $G'_i = G_i$
9: Find an Eulerian path p in G'_N
10: **return** p

▶ **Theorem 3.2.** *Algorithm 1 generates all minimal Sturmian full words.*

Proof. Suppose that w is a minimal Sturmian full word of order N. Thus, Lemma 3.1 implies that there is a sequence of graphs $G_2, \ldots, G_N$ such that G_i has i vertices and $i+1$ edges.

Furthermore, G'_2 is a subgraph of $G_{\mathcal{S}}$, where $\mathcal{S} = \{aa, ab, ba, bb\}$, for $i = 2, \ldots, N-1$, G_{i+1} is a subgraph of $L(G_i)$, and for each G_i there exists a path containing all its edges. Thus, w can be generated by Algorithm 1 unless there exists some G_i that does not contain an Eulerian path. However, since G_{i+1} is a subgraph of $L(G_i)$, Proposition 2 ensures that G_{i+1} contains an Eulerian path. ◀

3.2 The case of $h = 1$

Recall the minimal Sturmian partial word $a^{\lfloor N/2 \rfloor} \diamond a^{\lfloor N/2 \rfloor} b a^{\lceil (N-4)/2 \rceil}$ of order $N \geq 6$ in Table 1, which has the form $a^i \diamond a^j b a^l$ for some i, j, l. We show that any minimal Sturmian partial word with one hole has a similar form. Note that since $N \geq 6$, any Sturmian partial word w of order N with one hole satisfies $N < |w|$ (otherwise, $N = |w|$, and we get $N + 1 = p_w(N) = 2$, a contradiction).

▶ **Lemma 3.3.** *1. For $N \geq 6$, any Sturmian partial word w of order N of the form $w = \diamond z$, where z is a full word, is not minimal.*
2. Any Sturmian partial word w of order N of the form $w = a^i \diamond (a^j b)^m y$, where $i, j \geq 1$, $m \geq 2$, and y is a prefix of $a^j b$, is not minimal.

Proof. For Statement (2), we first prove that $N \leq \min(s, t)$, where $s = i + j + 1$ and $t = (j+1)m + |y| + 1$. First suppose that $s \leq t$ and $N \geq s + 1$. Note that $\mathrm{Sub}_w(s+1) = \{a^s b, \ldots, a^{i+1} b a^j, a^i b a^j b, \ldots, b(a^j b)^{s/(j+1)}\}$. This implies that $p_w(s+1) = s - (i+1) + 1 + i + 1 = s + 1 < s + 2$, a contradiction.

Next suppose that $t < s$ (so that $t + 1 \leq i + j + 1$). If $N \geq t + 1$, then $\mathrm{Sub}_w(t+1) = \{a^{t+1}, a^t b, \ldots, ab(a^j b)^{(t-1)/(j+1)}\}$, so $p_w(t+1) = 1 + t - 1 + 1 = t + 1 < t + 2$, a contradiction. Hence, $N \leq \min(s, t)$ as claimed. Therefore, if w has order N, then $s, t \geq N$ or $i + j + 1, (j+1)m + |y| + 1 \geq N$. Thus, $|w| = i + 1 + (j+1)m + |y| = s - j - 1 + t$. For a fixed t, j takes on a maximum value when $m = 2$ and $|y| = 0$. Hence, $2(j+1) + 1 \leq t$ so that $j \leq \frac{t-3}{2}$ and $|w| = s - j - 1 + t \geq s - \frac{t-3}{2} - 1 + t = s + \frac{t}{2} + \frac{1}{2} \geq \frac{3N}{2} + \frac{1}{2}$. However, $L_2(n+1, N, 1) \leq \frac{3N}{2}$ from Remark 3, so w is not minimal. ◀

▶ **Theorem 3.4.** *Suppose w is a Sturmian partial word with one hole of order $N \geq 6$ with a factor $z = \diamond a^i b$, where $i \geq 1$. Then w contains no other b's or w is not minimal.*

Proof. Similarly to the above lemma, we use the fact (from Remark 3) that if $|w| > \frac{3N}{2}$ then w is not minimal. If w contains no other b's we are done. Otherwise, w contains a factor of the form $b a^j z$ or $z a^j b$, for some $j \geq 1$. Note that if $j = 0$, w would contain all the four subwords of length 2, contrary to our assumption that w is Sturmian. First assume that $u = b a^j \diamond a^i b$ is a factor of w. Let $t = \min(i, j)$. Note that

$$\mathrm{Sub}_u(t+2) = \{a^{t+2}, b a^{t+1}, a^t b a, \ldots, a b a^t, a^{t+1} b, b a^t b\}$$

has size $t + 4$, implying that $N \leq t + 1$. Thus, $|w| \geq |u| = i + j + 3 \geq 2t + 3 > 2t + 2 \geq 2N$, so w is not minimal. Next assume that $u = \diamond a^i b a^j b$ is a factor of w.

First, suppose that $i > j$. Thus,

$$\mathrm{Sub}_u(j+2) = \{a^{j+2}, b a^{j+1}, a^{j+1} b, a b a^j, \ldots, a^j b a, b a^j b\}$$

has size $j + 4$ implying that $N \geq j + 1$ Similarly to the above, this implies that $|w| \geq 2N$ and w is not minimal. The case where $j > i + 1$ is handled similarly since $\mathrm{Sub}_u(i+2)$ is too large. Now, suppose that $i = j$. So $w = xuy = x \diamond a^i b a^i b y$ for some full words x, y. Note that if x contains a b, it has already been shown that w is not minimal. Furthermore, if $x = \varepsilon$, then w is not minimal by Lemma 3.3(1). Therefore, $w = a^l \diamond (a^i b)^2 y$ for some

$l \geq 1$. Note that if $N < i+2$, then $|w| \geq 2N$ and w is not minimal. So suppose $N \geq i+2$. We have that $\mathrm{Sub}_w(i+2) \supset \{a^{i+2}, a^{i+1}b, a^iba, \ldots, aba^i, ba^ib\} = \mathcal{S}$. Since the latter set is of size $i+3$, w must avoid $\{a,b\}^{i+2} \setminus \mathcal{S}$. Thus, $w = a^l\diamond(a^ib)^m y$ for some $m \geq 2$ and some prefix y of a^ib, and by Lemma 3.3(2), w is not minimal. Finally, suppose that $j = i+1$. So $w = xuy = x\diamond a^iba^{i+1}by$ for some full words x, y. Similarly to the above, we only need to consider the case when $w = a^l\diamond a^iba^{i+1}by$ for some $l \geq 1$. Note that $\mathrm{Sub}_w(i+2) \supset \{a^{i+2}, ba^{i+1}, a^iba, \ldots, aba^i, a^{i+1}b, ba^ib\}$, so $\|\mathrm{Sub}_w(i+2)\| \geq i+4$ and $N \leq i+1$. However, this implies that $|w| \geq 2N$, and w is not minimal. ◀

▶ **Corollary 3.5.** *For $N \geq 6$, any minimal Sturmian partial word with one hole is of the form $a^i\diamond a^jba^l$ for some i, j, l (up to reversal and complement).*

The next lemma gives some restrictions on the integers i, j, l.

▶ **Lemma 3.6.** *Let $w = a^i\diamond a^jba^l$ be a minimal Sturmian partial word with one hole of order $N \geq 6$. If N is odd, w has no repeated subwords of length N, and $i, j+l+1 < N$ (e.g. all factors of w of length N contain a hole). If N is even, exactly one of the following holds:*

- *w has exactly one subword of length N repeated exactly once, and $i, j+l+1 < N$.*
- *w has no repeated subwords of length N, and $i < N$, $j+l+1 = N$.*

Proof. Assume that N is odd. Thus, $|w| = \frac{3N}{2} - \frac{1}{2} \leq 2N-1$. From Proposition 1(2), $p_w(N) \leq 2(|w| - N + 1) = N+1$, and we have equality if and only if w has no repeated subwords of length N. Furthermore, the proof of Proposition 1(2) shows that each factor of w of length N contains a hole, and so $i, j+l+1 < N$.

Assume that N is even. Thus, $|w| = \frac{3N}{2} \leq 2N-1$ and $p_w(N) \leq 2(|w| - N + 1) = N+2$ from Proposition 1(2). More details follow. If $\|\mathcal{F}_0(w,N)\| \geq 2$, then $\|\mathcal{F}_1(w,N)\| \leq |w| - N - 1$ and $p_w(N) \leq \|\mathcal{F}_0(w,N)\| + 2\|\mathcal{F}_1(w,N)\| \leq N$, and so w is not Sturmian. If $\|\mathcal{F}_0(w,N)\| = 1$, then $\|\mathcal{F}_1(w,N)\| = |w| - N$ and $p_w(N) \leq \|\mathcal{F}_0(w,N)\| + 2\|\mathcal{F}_1(w,N)\| = N+1$, and equality holds if and only if no subword of length N repeats. This can only be the case when $i < N, j+l+1 = N$ (note that w has a^N as a repeated subword of length N when $i = N$, $j+l+1 < N$). If $\|\mathcal{F}_0(w,N)\| = 0$, then $\|\mathcal{F}_1(w,N)\| = |w| - N + 1$ and $p_w(N) \leq \|\mathcal{F}_0(w,N)\| + 2\|\mathcal{F}_1(w,N)\| \leq N+2$, and so $p_w(N) = N+1$ implies that exactly one subword must repeat exactly once. This can only be the case when $i, j+l+1 < N$. Again, the proof of Proposition 1(2) makes it evident that the two cases listed above are the only ones that lead to $p_w(N) = N+1$. ◀

The next lemma gives upper and lower bounds on j.

▶ **Lemma 3.7.** *Let $w = a^i\diamond a^jba^l$ be a minimal Sturmian partial word with one hole of order $N \geq 6$. Then $\lfloor\frac{N-1}{2}\rfloor \leq j \leq \lfloor\frac{N}{2}\rfloor$.*

Proof. To show the lower bound $j \geq \lfloor\frac{N-1}{2}\rfloor$, suppose that $j < \lfloor\frac{N-1}{2}\rfloor$. First suppose that $l \geq j+1$. Here $i, j \geq 1$. Since $N \geq 6$, we have that $N \geq j+2$. However, $\mathrm{Sub}_w(j+2) \supset \{a^{j+2}, a^{j+1}b, a^jba, \ldots, aba^j, ba^{j+1}, ba^jb\}$ so that $p_w(j+2) \geq j+4$. Thus, $l \leq j$.

Assume that N is even. Thus, $j = \frac{N}{2} - m$ for some $m \geq 2$. Noting that $|w| = \frac{3N}{2} = i+j+l+2$ we have that $i \geq \frac{N}{2} - 2 + 2m$. Thus, $i+j+1 \geq \frac{N}{2} - 2 + 2m + \frac{N}{2} - m + 1 \geq N + m - 1 \geq N+1$, with equality holding if and only if $l = j$. If $l = j$, both a^N and $a^{N-l-1}ba^l$ are repeated subwords of length N of w, contradicting Lemma 3.6. Similarly, $l < j$ implies that $i+j+1 \geq N+2$, meaning that a^N appears as a subword at least three times, again contradicting Lemma 3.6. ◀

The next theorem gives the classification of the one-hole minimal Sturmian words.

▶ **Theorem 3.8.** *Let $N \geq 6$.*

1. *If N is odd, then the only minimal Sturmian partial word with one hole of order N (up to reversal and complement) is $a^{N/2-1/2} \diamond a^{N/2-1/2} b a^{N/2-3/2}$, or equivalently $a^{\lfloor N/2 \rfloor} \diamond a^{\lfloor N/2 \rfloor} b a^{\lceil (N-4)/2 \rceil}$, and so $N_2(n+1, N, 1) = 4$.*
2. *If N is even, then the only minimal Sturmian partial words with one hole of order N (up to reversal and complement) are $a^{N/2} \diamond a^{N/2-1} b a^{N/2-1}$, $a^{N/2-1} \diamond a^{N/2} b a^{N/2-1}$, and $a^{N/2} \diamond a^{N/2} b a^{N/2-2} = a^{\lfloor N/2 \rfloor} \diamond a^{\lfloor N/2 \rfloor} b a^{\lceil (N-4)/2 \rceil}$, and so $N_2(n+1, N, 1) = 12$.*

Proof. Let w be a minimal Sturmian partial word with one hole. By Corollary 3.5, $w = a^i \diamond a^j b a^l$ for some i, j, l. For Statement (2), when N is even, $j = \frac{N}{2} - 1$ or $j = \frac{N}{2}$. Assume that $j = \frac{N}{2} - 1$. From Lemma 3.6 we have two cases to consider. Suppose $j + l + 1 = N$, so that $l = \frac{N}{2}$ and $i = \frac{N}{2} - 1$. Setting $t = \frac{N}{2} + 1$, we have that $t \leq N$ and that $\text{Sub}_w(t) = \{a^t, a^{t-1}b, a^{t-2}ba, \ldots, aba^{t-2}, ba^{t-1}, ba^{t-2}b\}$ is of size $t+2$, a contradiction. Thus, $i, j+l+1 < N$, and w can have at most one repeated subword of length N. Set $l = \frac{N}{2} - m$ for some $m \geq 1$, so that $i = \frac{N}{2} - 1 + m$. Further note that $a^i b a^l$ is a repeated subword of length N of w. We also have that $i + j + 1 = N - 1 + m$, so that if $m > 1$, a^N is also a repeated subword of length N, a contradiction. Therefore, $m = 1$ and $w = a^{N/2} \diamond a^{N/2-1} b a^{N/2-1}$. ◀

3.3 The case of $h = 2$

Recall from Table 1 that $a^{\lfloor (N-6)/2 \rfloor} \diamond a^{N-5} \diamond a^{\lceil (N-6)/2 \rceil}$ is a Sturmian partial word of order $N \geq 12$ of length $2N - 9$. We show that this form is minimal, and in fact all minimal Sturmian partial words with two holes are similar. The next proposition describes behavior between the holes.

▶ Proposition 3. Suppose that w is a Sturmian partial word of order N. Let z be a factor of w of the form $z = \diamond a_0 \cdots a_{l-1} \diamond$, where $a_0, \ldots, a_{l-1} \in \{a, b\}$. Then, $N < \frac{l}{2} + \frac{3}{2}$, or z is one-periodic, or $z = w = \diamond a^j b a^{n_1} b a^{n_2} b \cdots b a^{n_i} b a^j \diamond$ for some $i, j \geq 0$ and some $n_1, n_2, \ldots, n_i \in \{j, j+1\}$.

Proof. If $N < \frac{l}{2} + \frac{3}{2}$ we are done. Thus, assume $N \geq \frac{l}{2} + \frac{3}{2}$ throughout the rest of the proof. If $l < 2$ the statement is immediate. So assume that $l \geq 2$. Without loss of generality assume that $a_0 = a$. For j, $0 \leq j < \frac{l}{2}$, we show that either z avoids $ba^j b$ or $z = w = \diamond a^j b a^{n_1} b a^{n_2} b \cdots b a^{n_i} b a^j \diamond$ for some $i \geq 0$ and some $n_1, n_2, \ldots, n_i \in \{j, j+1\}$.

Assume first that $j = 0$. Suppose that $a_{l-1} = b$. Thus, $\diamond a$ and $b \diamond$ are factors of z, and $aa, ba, bb \in \text{Sub}_z(2)$. Since $p_z(2) = 3$, z must avoid ab. However, since $a_0 = a$ we have that $a_{l-1} = a$, a contradiction. Thus, $a_{l-1} = a$, and $aa, ab, ba \in \text{Sub}_z(2)$ implying that z avoids bb. Inductively, suppose that z avoids $bb, bab, \ldots, ba^{j-1}b$. This implies that $a_0 = \cdots = a_{j-1} = a_{l-1-j+1} = \cdots = a_{l-1} = a$. If z is one-periodic we are done, so suppose otherwise. Note that this also implies that $j < \frac{l}{2}$, else z would be one-periodic. Thus, $j + 2 \leq \frac{l}{2} + \frac{3}{2} \leq N$. Since z avoids $bb, \ldots, ba^{j-1}b$, we have that

$$\text{Sub}_z(j+2) \subset \{a^{j+2}, a^{j+1}b, a^j ba, \ldots, ba^{j+1}, ba^j b\} = \mathcal{S}$$

Note that $\|\mathcal{S}\| = j + 4$, so exactly one element of $\mathcal{S}$ must be avoided. Further, note that since z is not one-periodic, $\{a^{j+1}b, a^j ba, \ldots, ba^{j+1}\} \subset \text{Sub}_w(j+2)$. If z avoids $ba^j b$ we are done. Thus suppose that z avoids a^{j+2}. Thus, $z = \diamond a^j b a^{n_1} b a^{n_2} b \cdots b a^{n_i} b a^j \diamond$ for some $i \geq 0$ and $n_1, n_2, \ldots, n_i \in \{j, j+1\}$. Suppose that $z \neq w$, so we can write $w = xzy$ for some partial words x, y where at least one of $x, y \neq \varepsilon$. Without loss of generality assume that $y \neq \varepsilon$. Note that since $p_z(2) = p_w(2) = 3$, we have that w avoids bb. Therefore, $\diamond \neq y_0 \neq b$ so $y_0 = a$. However, this implies that a^{j+2} is a subword of w that is avoided by z, so that $p_w(j+2) > p_z(j+2) = j + 3$, a contradiction. Thus, both $x, y = \varepsilon$ and $w = z$. ◀

▶ **Corollary 3.9.** *Minimal Sturmian partial words of order $N \geq 12$ with two holes are one-periodic.*

It remains to restrict the placement of the holes.

▶ Proposition 4. Let $N \geq 12$. Any minimal Sturmian partial word w of order N with two holes having a factor of the form $z = \diamond a^j \diamond$, where $j \geq 1$, satisfies $|w| = 2j + 1 = 2N - 9$.

Proof. We first show that any minimal one-periodic Sturmian partial word w with two holes of order N having a factor of the form $z = \diamond a^j \diamond$, where $j \geq 1$, satisfies $N \geq j + 2$. Suppose not, that is $N < j+2$, so that $N-2 < j$. Set $j = N-2+m$ for some $m \geq 1$. It is easy to note that w avoids $bb, bab, \ldots, ba^{N-2}b$. Setting $\mathcal{S} = \{a^N, a^{N-1}b, a^{N-2}ba, \ldots, aba^{N-2}, ba^{N-1}\}$, we have that $\mathrm{Sub}_z(N) \subset \mathcal{S}$. If w is Sturmian of order N, then $\mathrm{Sub}_w(N) = \mathcal{S}$. Since w is one-periodic, $w = a^i \diamond a^j \diamond a^l$ for some $i, l \geq 0$. If $i + l < N - 2$, then $\mathrm{Sub}_w(N) \neq \mathcal{S}$, so $i + l \geq N - 2$. However, this implies that $|w| \geq N - 2 + N - 2 + m + 2 \geq 2N - 1$. Thus, by Remark 3, w is not minimal, a contradiction.

Now, note that $\mathrm{Sub}_z(j + 2) = \{a^{j+2}, a^{j+1}b, ba^{j+1}, ba^j b\}$ so $p_z(j + 2) = 4$. Suppose $|w| \geq 2j + 2$. Then w has a factor $v = a^i \diamond a^j \diamond a^{j-i}$, for some i, j, $0 \leq i \leq j$. However, we have

$$\mathrm{Sub}_v(j + 2) = \mathrm{Sub}_z(j + 2) \cup \{a^i ba^{j-i+1}, \ldots, aba^j, a^j ba, \ldots, a^{i+1}ba^{j-i}\}$$

Thus, $p_w(j+2) \geq p_v(j+2) = j+4$, a contradiction. Suppose $|w| \leq 2j$. Then $w = a^i \diamond a^j \diamond a^{m-i}$ for some i, j, m, $0 \leq i \leq m < j - 1$. Thus,

$$\mathrm{Sub}_w(j + 2) = \{a^{j+2}, a^{j+1}b, ba^{j+1}, ba^j b, a^i ba^{j-i+1}, \ldots, aba^j, a^j ba, \ldots, a^{j+i-m+1}ba^{m-i}\}$$

so $p_w(j + 2) < 4 + j - 1 = j + 3$, a contradiction. Therefore, $|w| = 2j + 1$.

Note also that $p_w(j+6) \leq 1+\|\mathcal{F}_1(w, j+6)\|+3\|\mathcal{F}_2(w, j+6)\| \leq 1+(|w|-(j+6)-5)+3\times 5 = j + 6 < j + 7$ (there is 1 subword with no b, at most $\|\mathcal{F}_1(w, j + 6)\|$ subwords with one b (fill the hole with b), and at most $3\|\mathcal{F}_2(w, j + 6)\|$ other subwords (fill the holes with ab, ba, bb)). Thus, $j + 2 \leq N < j + 6$. So $N - 5 \leq j \leq N - 2$. The only option is $j = N - 5$ in order to achieve $|w| = 2j + 1 \leq 2N - 9$. Finally, $w = a^{\lfloor (j-1)/2 \rfloor} \diamond a^j \diamond a^{\lceil (j-1)/2 \rceil}$ is of length $2j + 1$ and is Sturmian of order $N = j + 5$ when $j \geq 7$. ◀

Our two-hole description of minimal Sturmian partial words follows.

▶ **Theorem 3.10.** *The only minimal Sturmian partial words with two holes of order $N \geq 12$ are those of the form $a^i \diamond a^j \diamond a^l$, where $j = N - 5$, $i, l \geq 3$, and $i + l = N - 6$, and so $N_2(n + 1, N, 2) = 2N - 22$.*

Proof. Let w be a minimal Sturmian partial word of order N with two holes. The fact that $j = N-5$ and $i+l = N-6$ is evident from Proposition 4. We are left to show that $i, l \geq 3$. Since a^N is trivially a subword of length N of w, we have that $p_w(N) \leq 1 + \|\mathcal{F}_1(w, N)\| + 3\|\mathcal{F}_2(w, N)\|$. Note that since $|w| = 2N - 9$, we have that $\|\mathcal{F}_1(w, N)\| + \|\mathcal{F}_2(w, N)\| \leq |w| - N + 1 = N - 8$. Thus, $\|\mathcal{F}_1(w, N)\| \leq N-8-\|\mathcal{F}_2(w, N)\|$. Therefore, $p_z(N) = N+1 \leq 1+N-8+2\|\mathcal{F}_2(w, N)\|$, implying that $\|\mathcal{F}_2(w, N)\| \geq 4$. Note that if $i < 3$ (the case where $l < 3$ is similar), there are $i + 1 < 4$ factors containing two holes, a contradiction. Thus, there are $N - 11$ hole placements that are valid for a minimal Sturmian partial word of order N with two holes. Since the partial word is one-periodic, we have $N_2(n + 1, N, 2) = 2(N - 11) = 2N - 22$ as desired. ◀

3.4 The case of $h \geq 3$

Recall from Table 1 that $w = \diamond(a^{N-1}\diamond)^{h-1}$ is a Sturmian partial word with h holes of order N of length $N(h+1)+1$, when $h \geq 3$ and $N \geq h+1$. By Remark 3, $L_2(n+1, N, h) \leq N(h-1)+1$ in that case. We show that w is in fact minimal, and that (up to complement) it is the unique such word.

▶ **Lemma 3.11.** *Any Sturmian partial word w having a factor $z = \diamond a^i \diamond a^j \diamond, ba^i \diamond a^j \diamond$, or $\diamond a^i \diamond a^j b$ is of order $N < \min(i, j) + 2$. Furthermore, if w has another factor u compatible with $ba^l b$ where $l < \min(i, j)$ then $N < l + 2$.*

Proof. Set $t = \min(i, j)$. Assume that $N \geq t + 2$. We immediately note that $\text{Sub}_w(t + 2) \supset \text{Sub}_z(t + 2) = \{a^{t+2}, ba^t b, a^{t+1}b, \dots, ba^{t+1}\}$, so $\|\text{Sub}_w(t + 2)\|$ is at least $t + 4$, contradicting the fact that w is Sturmian. Now assume such a factor u exists. Assume that $N \geq l + 2$. Trivially, $ba^l b \in \text{Sub}_w(l + 2)$. Furthermore, $\{a^{l+2}, a^{l+1}b, \dots, ba^{l+1}\} \subset \text{Sub}_z(l + 2)$. Thus, $p_w(l + 2)$ is at least $l + 4$, a contradiction. ◀

▶ **Theorem 3.12.** *For $h \geq 3$ and $N \geq h + 1$, $L_2(n + 1, N, h) = N(h - 1) + 1$. Furthermore, any minimal Sturmian partial word with h holes of order N is of the form $\diamond(a^{N-1}\diamond)^{h-1}$, and so $N_2(n + 1, N, h) = 2$.*

Proof. Any minimal Sturmian partial word w with $h \geq 3$ holes of order N must have a factor of the form $\diamond a^i \diamond$, where $i \geq 1$. By Lemma 3.11, w must be of form $a^{n_0}c_0a^{n_1}c_1 \cdots a^{n_j}c_ja^{n_{j+1}}$, where each $c_i \in \{\diamond, b\}$ and each $n_i \geq N - 1$. It is thus evident that $w = \diamond(a^{N-1}\diamond)^{h-1}$, which was shown in Table 1 to be Sturmian of order N for $N \geq h + 1$, is the only form possible for a minimal Sturmian partial word with h holes. ◀

4 Conclusion

We have thus classified all the $(n + 1)$-complex partial words with any number of holes. The number of minimal Sturmian full words of order N, $N_2(n + 1, N, 0)$, remains to be computed, but an algorithm has been presented that can generate all such words. It would be interesting to complete the classification of the minimal $2n$-complex partial words as well. In this section, we give some preliminary results by filling out Table 2.

Table 2 $2n$-complex partial words with h holes of order N

h	N	partial word	$L_2(2n, N, h)$
0	≥ 3	$a^N ba^{N-2}bba^{N-2}$	$3N - 1$
1	≥ 3	$a^{N-2}b\diamond a^{N-2}b$	$2N - 1$
2	≥ 5	$a^{\lfloor (N-4)/2 \rfloor}b(\diamond a^{\lceil (N-4)/2 \rceil}b)^2$	$\frac{3N}{2} - 1$ if N is even $\frac{3N}{2} - \frac{1}{2}$ if N is odd
≥ 3	≥ 5	$a^{\lfloor (N-4)/2 \rfloor}b(\diamond a^{\lceil (N-4)/2 \rceil}b)^h$	

▶ Proposition 5. For $h \geq 0$, $f(n) = 2n$ is $(2, h)$-feasible. For $N \geq 3$, $L_2(2n, N, 0) = 3N - 1$ and $L_2(2n, N, 1) = 2N - 1$.

Follows is our h-hole bound.

▶ Proposition 6. Let w be a word with $h \geq 2$ holes, and $n \leq |w|$ be a positive integer.

- If $|w| \geq 2n - 2 + h$, then $p_w(n) \leq 2^{h+1} + (n - h + 1)2^h + |w| - 2n - h - 2$.

- If $|w| \leq 2n - h$, then $p_w(n) \leq 2^h(|w| - n + 1)$.
- Else $2n - h < |w| < 2n - 2 + h$, and set $d = 2n - 2 + h - |w| > 0$. If d is even, then
  $$p_w(n) \leq 2^{h+1} + (n - h + 1)2^h - 4 - 2\sum_{i=1}^{d/2} 2^i = 2^{h+1} + (n - h + 1)2^h - 4 \times 2^{d/2}$$
  If d is odd, then
  $$p_w(n) \leq 2^{h+1} + (n-h+1)2^h - 4 - 2\sum_{i=1}^{(d-1)/2} 2^i - 2^{(d+1)/2} = 2^{h+1} + (n-h+1)2^h - 3 \times 2^{(d+1)/2}$$

▶ **Corollary 4.1.** *For $N \geq 5$, $L_2(2n, N, 2) = \frac{3N}{2} - 1$ if N is even, and $\frac{3N}{2} - \frac{1}{2}$ if N is odd.*

References

1 M. A. Alekseyev and P. A. Pevzner. Colored de Bruijn graphs and the genome halving problem. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(1):98–107, 2007.

2 J.-P. Allouche. Sur la complexité des suites infinies. *Bulletin of the Belgium Mathematical Society*, 1:133–143, 1994.

3 J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003.

4 P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexité $2n + 1$. *Bulletin de la Société Mathématique de France*, 119:199–215, 1991.

5 J. Berstel and L. Boasson. Partial words and a theorem of Fine and Wilf. *Theoretical Computer Science*, 218:135–141, 1999.

6 F. Blanchet-Sadri. *Algorithmic Combinatorics on Partial Words.* Chapman & Hall/CRC Press, Boca Raton, FL, 2008.

7 F. Blanchet-Sadri, J. Schwartz, S. Stich, and B. J. Wyatt. Binary de Bruijn partial words with one hole. In J. Kratochvil et al., editor, *TAMC 2010, 7th Annual Conference on Theory and Applications of Models of Computation, Prague, Czech Republic*, volume 6108 of *Lecture Notes in Computer Science*, pages 128–138, Berlin, Heidelberg, 2010. Springer-Verlag.

8 N. G. De Bruijn. Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of $2n$ zeros and ones that show each n-letter word exactly once. Technical Report 75–WSK–06, Department of Mathematics and Computing Science, Eindhoven University of Technology, The Netherlands, 1975.

9 S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Mathematics*, 206:145–154, 1999.

10 J. L. Gross and J. Yellen. *Handbook of Graph Theory.* CRC Press, 2004.

11 M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge University Press, Cambridge, 2002.

12 F. Manea and C. Tiseanu. Hard counting problems for partial words. In A.-H. Dediu, H. Fernau, and C. Martín-Vide, editors, *LATA 2010, 4th International Conference on Language and Automata Theory and Applications, Trier, Germany*, volume 6031 of *Lecture Notes in Computer Science*, pages 426–438, Berlin, Heidelberg, 2010. Springer-Verlag.

13 M. Morse and G. A Hedlund. Symbolic dynamics II. Sturmian trajectories. *American Journal of Mathematics*, 62(1):1–42, 1940.

14 G. Rote. Sequences with subword complexity $2n$. *Journal of Number Theory*, 46:196–213, 1993.

15 R. P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge University Press, Cambridge, 2001.

16 M.-w. Wang and J. Shallit. On minimal words with given subword complexity. *The Electronic Journal of Combinatorics*, 5:1–16, 1998.