Number of Holes in Unavoidable Sets of Partial Words II

By: F. Blanchet-Sadri, Steven Ji, Elizabeth Reiland

Blanchet-Sadri, F., Ji, S., Reiland, E. (2012). Number of Holes in Unavoidable Sets of Partial Words II. *Journal of Discrete Algorithms, 14*, 65-73.doi: 10.1016/j.jda.2011.12.002

Abstract:

We are concerned with the complexity of deciding the avoidability of sets of partial words over an arbitrary alphabet. Towards this, we investigate the minimum size of unavoidable sets of partial words with a fixed number of holes. Additionally, we analyze the complexity of variations on the decision problem when placing restrictions on the number of holes and length of the words.

Keywords: Automata and formal languages | Computational complexity | Combinatorics on words | Partial words | Unavoidable sets | NP-hard problems

Article:

1. Introduction

An *unavoidable set of* (*full*) *words X* over an alphabet *A* is one such that any two-sided infinite word over *A*has a factor in *X*. Partial words, a generalization of full words, may contain "hole symbols", denoted by "⋄'s", which are not considered part of the alphabet *A*. The ⋄ symbol is *compatible* with, or matches, each letter of*A*. An *unavoidable set of partial words X* over *A* is then defined as a set such that any two-sided infinite full word over *A* has a factor compatible with some element of *X*. This concept of unavoidable sets of partial words was introduced in [3].

Efficient algorithms to decide if a finite set X of full words over an alphabet A is unavoidable are well known[10]. For example, this check can be done by finding whether or not there is a loop in the automaton that recognizes $A^* \setminus A^* X A^*$, which must be finite for a set of words to be unavoidable [1]. This algorithm can be adapted to decide if a finite set X of partial words is

unavoidable by determining the avoidability of $\hat{X}$, the completion of partial words in X. However, the computation is also much less efficient as a word with h holes can be completed in as many as $|A|^h$ ways. AVOIDABILITY, or the problem of deciding the avoidability of a finite set of partial words over a k-letter alphabet, where $k \geqslant 2$, turns out to be NP-hard [5] and [2], which is in contrast with the well-known feasibility results for a set of full words [7] and [10]. This can be proved by using a reduction from the 3SAT problem, known to be NP-complete. AVOIDABILITY also turns out to be in PSPACE [2].

In this paper, we prove several new results related to the complexity of deciding the avoidability of sets of partial words. More specifically, we calculate the minimum cardinality of unavoidable sets of partial words of a given length m with a fixed number of holes over a k-letter alphabet. Previous work has been done in the context of full words. Mykkeltveit, in particular, showed that the minimum number of elements in an unavoidable set of full words of length m over an alphabet of size k is equal to $c(m,k)$, the number of conjugacy classes of words of that length over the given alphabet [12]. We also analyze the complexity of variations on the avoidability problem building on previous work by Blakeley et al. [2]. In particular, we study the complexity of deciding *aperiodic* (non-ultimately periodic) unavoidable sets of partial words. This notion, which is a natural extension of unavoidable sets, was introduced by Higgins and Saker in the context of full words [9]. In addition, we provide a new hard counting problem on partial words adding to previous work by Manea and Tiseanu [11].

The contents of our paper is as follows: In Section 2, we present the basic definitions and terminology regarding the major problem on unavoidable sets we are concerned with, that is, the *complexity problem* or the complexity of the problem of deciding the avoidability of a finite set of partial words over a k-letter alphabet. In Section 3, we provide some bounds on the minimum cardinality of unavoidable sets containing partial words of length m with h holes over a k-letter alphabet. In Section 4, we analyze the complexity of variations on the avoidability problem with restrictions put on the number of holes and length of the words. Additionally, we generalize the concept of aperiodic avoidability to sets of partial words and prove that the problem of deciding if a finite set of partial words over a k-letter alphabet is avoided by a one-sided aperiodic word is NP-hard. In Section 5, we present a hard counting problem on partial words. Finally in Section 6, we conclude with some remarks.

2. Preliminaries

Let A be a non-empty finite set called an *alphabet* whose elements we call *letters*. A *finite full word* (or simply finite word) w over A is a finite sequence of letters from A. We denote the length of w by $|w|$ and the $(i+1)$st letter of w by $w(i)$ (by convention, we index positions of w from zero). By ε we denote the empty word and by A^* the set of all finite words over A.

A *two-sided infinite full word* (or simply infinite word) w over A can be viewed as a function $w: Z \to A$. We say that w has period p for some positive integer p, and call it p-periodic, if $w(i) = w(i+p)$ for all $i \in Z$. If w has a period, we call it *periodic*. If v is a non-

empty finite word, then we denote the unique infinite word $w = \cdots vvvvv\cdots$ such that $v = w(0)\cdots w(|v|-1)$ by $v^{\mathbb{Z}}$. Similarly, a *one-sided infinite full word* w can be viewed as a function $w : \mathbb{N} \to A$. We call w *ultimately periodic* if there exist finite words u and v $(v \neq \varepsilon)$ such that $w = uvvvv\cdots$. We call a finite word v a *factor* of a word w if there exists some integer index i such that $v = w(i)\cdots w(i+|v|-1)$.

A *partial word* w of length m over A is a function $w : \{0, \ldots, m-1\} \to A_\diamond$ where $A_\diamond = A \cup \{\diamond\}$ with $\diamond \notin A$. The $\diamond$ symbol is referred to as a "hole". For the indices $0 \leqslant i \leqslant m-1$ such that $w(i) \in A$, we say that i is in the domain of w, denoted by $D(w)$. Otherwise, i is in the set of holes of w, denoted by $H(w)$. The set denoted by $A_\diamond^*$ represents the set of all finite words over $A_\diamond$ (i.e. the set of all finite partial words over A, including the empty word, ε). If a partial word can be written as $u_1 \diamond u_2 \diamond \cdots \diamond u_{n-1} \diamond u_n$, then the set $\{u_1 a_1 u_2 a_2 \cdots u_{n-1} a_{n-1} u_n \mid a_i \in A\}$ is a *partial expansion* on u. Note that the u_i's are not necessarily full words. In this paper, it is assumed, without loss of generality, that the first and last positions of every partial word in a set be defined (i.e. that these positions not be holes).

We say a finite partial word v is a *factor* of a partial word w if there exist x and y such that $w = xvy$. Two partial words u and v of equal length are said to be *compatible*, denoted as $u \uparrow v$, if $u(i) = v(i)$ for all $i \in D(u) \cap D(v)$. A word w is said to *meet* a set of partial words X if some element of X is compatible with a factor of w. A two-sided infinite word w *avoids* X if no factor of w is compatible with any element of X. If no two-sided infinite word avoids X, we say that X is *unavoidable*. Otherwise, we call X *avoidable*. In [3], an algorithm is given for deciding avoidability on the basis of four reductions that maintain avoidability: factoring, prefix–suffix, hole truncation and expansion. This reduction method will be used in some of our proofs.

Two full words u and v are said to be *conjugate* if there exist x and y such that $u = xy$ and $v = yx$. Conjugacy is an equivalence relation, which we can use to form equivalence classes of words of a given length m over a fixed alphabet of size k. The number of conjugacy classes is denoted by $c(m,k)$.
In the next sections, we examine some complexity problems on partial words related to AVOIDABILITY and some variations of it.

3. Minimum size of unavoidable sets of constant length

In [12], Mykkeltveit proved that for the case of full words, the minimal cardinality of an unavoidable set of words of constant length m over a k-letter alphabet, $\alpha(m,k)$, is precisely $c(m,k)$, the number of conjugacy classes of words of length m over a k-letter alphabet. The inequality $\alpha(m,k) \geqslant c(m,k)$ holds since an unavoidable set needs to contain at least one word from each conjugacy class. For example, if $m = 2$ and $k = 2$, there are three conjugacy classes $\{aa\}$, $\{bb\}$ and $\{ab, ba\}$ of words of length two over the binary alphabet $\{a, b\}$, and so $\{aa, bb, ab\}$ is an unavoidable set.

In this section, we are interested in the problem of calculating the cardinality of minimal unavoidable sets of partial words of length m with h holes over a k -letter alphabet, which we denote by $\alpha(m,h,k)$. Results, in the case of $h=0$, have been obtained (for instance, see [13]). Using the algorithm for testing avoidability described in [3], Table 1 was obtained that gives $\alpha(m,h,k)$ for $2\leqslant m\leqslant 10$, $0\leqslant h\leqslant 8$, and $k=2$. Note that an empty entry in the table indicates an impossible case (i.e. too many holes) or an entry that has not yet been discovered due to extensive computation time.

Table 1. Some values of $\alpha(m,h,2)$.

h \ m	2	3	4	5	6	7	8	9	10
0	3	4	6	8	14	20	36	60	108
1		3	4	5					
2			3	3	5				
3				3	4	5			
4					3	3	5		
5						3	4	5	
6							3	3	5
7								3	3
8									3

In the following case, we can determine the exact value of $\alpha(m,h,k)$.

Proposition 1.
For $m\geqslant 2$, $\alpha(m,m-2,k)=c(2,k)$.

Proof.
We first show that $\alpha(m,m-2,k)\leqslant c(2,k)$. Take a minimal size unavoidable set X with full words of length 2. Then $|X|=c(2,k)$ by definition. Create a new set $X'=\{x(0)\diamond^{m-2}x(1)\,|\,x\in X\}$, so X' contains the elements of X with $m-2$ holes inserted into the middle. Note that the only configuration of $m-2$ holes in a length m word is having all the holes in the middle, since we specify that a partial word cannot begin or end with holes. Also, $|X|=|X'|$. We claim that X' is unavoidable, and prove it through contradiction.

Assume that X' is avoidable. Then there is some two-sided infinite word v' that avoids X. We define the words v_i for $0\leqslant i<m-1$ as containing every $(m-1)$st symbol from v', starting with the i th symbol. That is, $v_i=\cdots v'(i-j(m-1))\cdots v'(i-m+1)v'(i)v'(i+m-1)\cdots v'(i+jm-j)\cdots$. Then some v_i must avoid X, since if any factor $v'(i+jm-j)v'(i+j(m-1)+1)\cdots v'(i+(j+1)(m-1))$ is compatible with $x(0)\diamond^{m-2}x(1)\in X'$ then $x=x(0)x(1)\in X$ must be compatible with the factor $v_i(j)v_i(j+1)=v'(i+j(m-1))v'(i+(j+1)(m-1))$ of v_i. Since X' being avoidable implies that X is avoidable, we have a contradiction, so X' must be unavoidable. Since there exists an unavoidable set of partial words of length m with $m-2$ holes over a k -letter alphabet, we have $\alpha(m,m-2,k)\leqslant c(2,k)$.

Next we show that $\alpha(m, m-2, k) \geqslant c(2, k) = \alpha(2, k)$. Consider an unavoidable set X' of minimal cardinality containing partial words of length m with $m-2$ holes over a k-letter alphabet. Define a set of full words $X = \{x(0)x(m-1) \mid x(0) \diamond^{m-2} x(m-1) \in X'\}$. We claim that X must be unavoidable as well.
Suppose that X is avoided by a two-sided infinite word v. Let

$$v' = v(-i)^{m-1} \cdots v(-1)^{m-1} v(0)^{m-1} v(1)^{m-1} \cdots v(i)^{m-1} \cdots$$

be such that every symbol in v is repeated $m-1$ times. Then v' avoids X', since if it does not avoid X', some factor $v'(i) \cdots v'(i+m-1)$ is compatible with $x = x(0) \diamond^{m-2} x(m-1) \in X'$, and therefore $v(\lfloor i/(m-1) \rfloor) v(\lfloor i/(m-1) \rfloor + 1)$ is compatible with $x(0)x(m-1) \in X$.
Since X being avoidable implies that X' is avoidable, we have a contradiction. Thus X is unavoidable. Since $|X| \geqslant \alpha(2, k) = c(2, k)$ and $|X| = |X'|$, we have
that $\alpha(m, m-2, k) \geqslant c(2, k)$.
Since we have shown both directions of the inequality, $\alpha(m, m-2, k) = c(2, k)$. □

Additionally, from observation of the data along other diagonals in the table, we propose the following conjecture, which is a generalization of the previous proposition.

Conjecture 1.

For $m \geqslant n \geqslant 2$, $\alpha(m, m-n, k) \leqslant c(n, k)$.
Clearly this conjecture holds for all cases we were able to check using our computer program. Bounds determined for some cases we were unable to get exact results for are also consistent with those suggested above. We can show the following.

Proposition 2.
For $m \geqslant n \geqslant 2$:

$$\frac{1}{k^{m-n}} c(m, k) \leqslant \alpha(m, m-n, k) \leqslant k c(n-1, k)$$

Proof.
For the lower bound, we first note that any unavoidable set of partial words of length m over a k-letter alphabet with $m-n+1$ holes can be made into an unavoidable set of partial words of the same length over the same alphabet with each word having $m-n$ holes by performing a partial expansion on one hole of each word. This gives us the inequality $\alpha(m, m-n, k) \leqslant k\alpha(m, m-n+1, k)$. By repeatedly applying this inequality, we find that $\alpha(m, m-n, k) \geqslant \frac{1}{k}\alpha(m, m-n-1, k) \geqslant \cdots \geqslant \frac{1}{k^{m-n}} c(m, k)$ as desired.
Consider an unavoidable set X of full words of length $n-1$ of minimal cardinality. We construct an unavoidable set of partial words of length m with $m-n$ holes by first

appending $m-n+1$ holes to the end of each $w \in X$, to get words of length m. We then partially expand the last position of every word, so the cardinality of the set is multiplied by k. Since a partial expansion of elements of an unavoidable set also gives an unavoidable set, our new set is unavoidable and we have $\alpha(m, m-n, k) \leqslant k\, c(n-1, k)$. □

Though we believe that these bounds can be improved upon, they do still offer some insight into values of $\alpha(m, m-n, k)$ for large values of m and n, which represent cases for which our computer program is unable to generate exact results.

Corollary 1.
For large values of m and n, $\alpha(m, m-n, k) \sim c(n, k)$.

Proof.

Previously, it has been demonstrated that $c(n, k)$ is asymptotically equivalent to $\frac{k^n}{n}$ [14] and [6]. This implies that as n grows large, the ratio of $c(n, k)$ to $c(n-1, k)$ approaches k, or $c(n, k) \sim k\, c(n-1, k)$ for sufficiently large n. Thus, though the absolute error between this proven upper bound and our conjectured upper bound increases as m and n grow, the relative error goes to zero.
By similar logic, we note that for large values of m and n,

$$\frac{1}{k^{m-n}} c(m, k) \sim \frac{1}{k^{m-n-1}} c(m-1, k) \sim \cdots \sim c(n, k)$$

Thus both our upper bound and lower bound on $\alpha(m, m-n, k)$ are asymptotically equivalent to $c(n, k)$, implying that for large m and n, $\alpha(m, m-n, k) \sim c(n, k)$. □

4. Complexity of avoidability problems

In [5] and [2], AVOIDABILITY was shown to be both NP-hard and in PSPACE. In this section, we first present an alternative, NFA-based approach to the AVOIDABILITY problem that also results in a polynomial space algorithm. In this approach we avoid having to transform the input into a set of equal length partial words.

Proposition 3.
AVOIDABILITY is in PSPACE.

Proof.

Let $X = \{x_1, \ldots, x_n\}$ be a set of partial words over an alphabet A of size k. For each $x_i \in X$ we can define in a natural way a regular expression R_i such that $L(R_i)$ is the set of all full words over A compatible with x_i. For example, let $x_1 = b \diamond\diamond a \diamond b$ over $\{a, b\}$; then $R_1 = b(a+b)(a+b)a(a+b)b$. Note that the representation of R_i has size $O(k|x_i|)$. We can then define a regular expression $R = A^{\square}(R_1 + R_2 + \cdots + R_n)A^{\square}$ so that $L(R)$ is the set of all words over A that contain at least one occurrence of a factor compatible with a partial word

in X . Note that the size of R is linear in the size of the representation of X . Finally, an NFA M accepting $L(R)$ can be constructed in linear time from R in a natural way so that M has at most $N = 1 + \sum_{x \in X} |x|$ states and $O(kN)$ transitions. It follows that there is a two-sided infinite word avoiding X if and only if there are infinitely many words not accepted by M . In other words, X is avoidable if and only if the complement of $L(M)$ is infinite. However, the problem of determining if the complement of a language accepted by an NFA is infinite is in PSPACE (in fact, this problem is PSPACE-complete) [15, Exercise 16, p. 199]. We can therefore decide if X is avoidable by constructing the NFA M (in linear time), and then applying a polynomial space algorithm to decide if the complement of $L(M)$ is infinite. □

In [2], the complexity of natural variations of AVOIDABILITY were analyzed. While some of them were shown to be NP-hard, others were shown to be efficiently decidable. We next build on this work by considering variations of AVOIDABILITY when restrictions are put on the number of holes and length of the words. On the one hand, we prove the following two propositions.

Proposition 4.
The problem of deciding the avoidability of a finite set of partial words with an equal number of holes over an alphabet of size $k \geqslant 2$ *is NP-hard.*

Proof.
We provide a reduction from the unrestricted AVOIDABILITY problem. Consider an instance of this problem: a finite set X of partial words over a k -size alphabet A . We construct a set X' of partial words with an equal number of holes. Let $h(w)$ denote the number of holes in a partial word w , and let h' denote the maximal number of holes in any word in X . Then we set

$$X' = \left\{ w \diamond^{h'-h(w)} a \mid w \in X,\ h(w) < h',\ a \in A \right\} \cup \left\{ w \mid w \in X,\ h(w) = h' \right\}$$

Note that the first part of X' has the same avoidability as
$\left\{ w \diamond^{h'-h(w)+1} \mid w \in X,\ h(w) < h' \right\}$

by an expansion operation, and this set in turn has the same avoidability as $\{w \mid w \in X,\ h(w) < h'\}$ by hole truncation. Thus, X' has the same avoidability as X ; that is, X' is avoidable if and only if X is avoidable. Finally, the length of the description of X being $\|X\| = \sum_{u \in X} |u|$, since the maximum number of holes in a word in X is upper bounded by the length of the longest word in X and the length of the longest word in X is in turn upper bounded by $\|X\|$, we get that $\|X'\| < \|X\|^2 k$, so this reduction runs in polynomial time. Thus, since AVOIDABILITY has been shown to be NP-hard, we have that our problem is also NP-hard. □

Proposition 5.
The problem of deciding the avoidability of a finite set of partial words where each word has equal length m and an equal number of holes $h < m - 2$ *over an alphabet of size* $k \geqslant 2$ *is NP-hard.*

Proof.
We proceed by reduction from the DIRECTED HAMILTONIAN CIRCUIT problem, known to be NP-complete [8]. Consider an instance of the problem: a digraph $G = (V, E)$. We want to determine whether G contains a Hamiltonian circuit. We construct a set X of partial words of equal length with an equal number of holes such that X is avoidable if and only if G has a Hamiltonian circuit. Let our alphabet be $V = \{v_1, v_2, \ldots, v_n\}$. Then our set X consists of the three parts $\{v_i v_j \diamond^{n-1} a \mid (v_i, v_j) \notin E,\ v_l \in V\}$, $\{v_i \diamond^{n-1} v_j a \mid v_i \neq v_j,\ v_l \in V\}$, and $\{v_i \diamond^{j} v_i \diamond^{n-j-1} a \mid 0 \leqslant j < n-1,\ v_l \in V\}$.
Suppose there exists some Hamiltonian circuit $(u_1, u_2, \ldots, u_n, u_1)$ in G. Then the word $(u_1 u_2 \cdots u_n)^{\mathbb{Z}}$ avoids X. Notice that since $(u_i, u_{i+1}) \in E$, for all $1 \leqslant i < n$, and $(u_n, u_1) \in E$, this word avoids the first part of *X*. Since it is *n*-periodic, it avoids the second part of *X*, and since no vertex can appear twice in a Hamiltonian circuit, instances of a particular letter are spaced *n* apart, thus avoiding the third part of *X*.

Next suppose there exists a two-sided infinite word w which avoids X. To avoid the second part of X, it must be the case that every n th letter is the same, so w is n-periodic, say $w = (u_1 u_2 \cdots u_n)^{\mathbb{Z}}$. By the third part of X, each letter can appear only once per period, and by the first part, each set of consecutive letters in X must represent an edge in G. Thus, $(u_1, \ldots, u_n, u_1)$ must be a Hamiltonian circuit in *G*.
While every word here has length $n+2$ with precisely three defined positions, we can extend the result to all sets such that each word has equal length and an equal number of holes, since if we could solve that problem efficiently, we could solve the specific case with three defined positions efficiently. □

On the other hand, since k^h is a constant when we fix the number of holes *h* and the alphabet size *k*, the following proposition shows that for an avoidable set *X* of partial words with a fixed number of holes, the minimal period of an avoiding word is polynomial with respect to the length of the words, *m*, and the cardinality of *X*, *n*.

Proposition 6.
Given an avoidable set X of n words of length m with h holes over a k-letter alphabet, X is avoided by a word of period at most $k^h mn$.

Proof.
Let Y be the expansion of all the words in X into full words. Then there are at most $k^h n$ elements in Y. By Proposition 3 given in Blakeley et al. [2], there is some avoiding word with period at most $k^h mn$. □

We define AVOID-H as the problem of deciding the avoidability of a set *X* of *n* partial words of length *m* with a fixed number of holes *h* over a fixed *k*-letter alphabet.

Proposition 7.

AVOID-H is in P.

Proof.
First, we prove that AVOID-H is in NP. If X is avoidable, by Proposition 6 there is an avoiding word u with period polynomial with respect to m and n. Choose non-deterministically a finite subword v of length $m + p$ of u, where p is the minimal period of u. Then v has length that is polynomial with respect to m, n, and $v$avoids X if and only if u avoids X. We can check in polynomial time if v avoids X, and thus if u avoids X.

Moreover, we can modify an automaton given in [10, p. 31] to decide AVOID-H in polynomial time. The automaton is as follows: Given a set of full words X, create a graph G such that the vertices are the prefixes of the elements of X, and there is an edge between two vertices u and v if there is some letter a in the alphabet such that u is the longest suffix of va. There, it is proved that a set X is unavoidable if and only if every cycle in G contains a vertex in X.

Now, we prove that AVOID-H is in P. Given a set X of n partial words of length m with a fixed number of holes h over a fixed k-letter alphabet, we can use the graph G associated with the set Y, the full expansion of X. The graph G has at most $k^h mn$ nodes, which is polynomial with respect to m and n. We can compute all the strongly connected components in polynomial time by using Tarjan's strongly connected components algorithm [16]. We can also check if each component has a node in Y in polynomial time. Thus, we can check if X is avoidable in polynomial time. □

We define AVOID-MAX-H to be the problem of deciding the avoidability of a set X of n partial words of length$m$ over a fixed k-letter alphabet, with each word having at most h holes. Since h is an upper bound on the number of holes per word, $k^h n$ is an upper bound on the number of elements in the expansion of X, giving us the following corollary.

Corollary 2.
AVOID-MAX-H is in P.

Remark 1.
We can generalize Proposition 7 and Corollary 2 to the case when the set X consists of words of length less than or equal to m. Indeed, since the graph defined in [10, p. 31] does not require the members of X to be of the same length, we can use m as an upper bound on the length and keep the same bounds as proven above. That is, deciding the avoidability of a set X of n words of length at most m with at most h holes can be done in polynomial time, with exactly the same proof as in Proposition 7.
Shifting focus somewhat, we define l-AVOIDABILITY to be the problem of deciding whether a finite set of partial words is avoided by a two-sided infinite word with period l. In addition, we define l-CIRCUIT to be the decision problem of determining whether a graph has a simple circuit of length l. The l-CIRCUIT problem can easily be shown to be NP-complete by a reduction from the well-known HAMILTONIAN CIRCUIT problem.

Proposition 8.
l-AVOIDABILITY is NP-complete.

Proof.
In [2, Lemma 1], it was shown that given a finite word w and a finite set Y of partial words, it can be determined in polynomial time whether the infinite periodic word w^Z avoids Y. So we can decide a set X by non-deterministically selecting a word w of length l and verifying that w^Z avoids X. Thus, l-AVOIDABILITY is in NP.

Next we show the problem is NP-hard by reducing from the l-CIRCUIT problem. Given an instance $G=(V,E),l$ of the l-CIRCUIT problem, we construct a set X of partial words such that G contains a simple circuit of length l if and only if X is avoided by some two-sided infinite word of period l. Let the alphabet be $V=\{v_1,v_2,\ldots,v_n\}$ and set $X=\{v_iv_j|(v_i,v_j)\notin E\}\cup\{v_i\diamond^j v_i|0\leqslant j\leqslant l-2\}$.

Suppose there exists a simple circuit $(u_1,u_2,\ldots,u_l,u_1)$ in G (where $u_i\in V$). Then the word $(u_1u_2\cdots u_l)^Z$ avoids X, since $(u_l,u_1)\in E$, $(u_i,u_{i+1})\in E$ and $u_i\neq u_j$ for every $1\leqslant i,j\leqslant l-1$, $i\neq j$. Now suppose there exists a word $w=(u_1u_2\cdots u_l)^Z$ which avoids X. Then any factor of w of length l contains distinct letters. Additionally, each pair of adjacent letters in w must be an element in E. Thus $(u_1,\ldots,u_l,u_1)$ must be a simple circuit in G of length l. □

Extending the concept of *aperiodic* (*non-ultimately periodic*) *avoidability* of Higgins and Saker [9], we now investigate aperiodic unavoidable sets of partial words. We call a set of partial words X over a finite alphabet A *aperiodic unavoidable* or *a-unavoidable* if every one-sided infinite aperiodic word over A has a factor compatible with some element of X, and *a-avoidable* otherwise. Note that all unavoidable sets are a-unavoidable, but the converse does not hold.

We define a-AVOIDABILITY to be the problem of deciding whether a finite set X of partial words over an alphabet of size $k\geqslant 2$ is a-avoidable. In [2], Blakeley et al. provided a polynomial space algorithm that decides whether a finite set of partial words is a-unavoidable, so a-AVOIDABILITY is in PSPACE (the same algorithm also decides if the number of words of length n avoiding a given finite set of partial words grows polynomially or exponentially with n).

Proposition 9.
a-AVOIDABILITY is NP-hard.

Proof.
We model our proof after the one that AVOIDABILITY is NP-hard [5]. We proceed by reduction from the 3SAT problem, which is well known to be NP-complete [8]. Consider an instance of 3SAT: a set of binary variables $x_1,x_2,\ldots,x_n$ and m clauses each containing three literals (i.e.

either a variable or its negation). We want to determine whether there exists a truth assignment for the variables such that each clause has at least one literal that is true.

We construct a set X of partial words over the alphabet $A = \{0, T, F\}$ such that X is a-avoidable if and only if there exists an appropriate truth assignment for our 3SAT instance. The elements of X are divided in three parts. First, X contains $0T0, 0T\diamond 0, \ldots, 0T\diamond^{n-2}0, 0F0, 0F\diamond 0, \ldots, 0F\diamond^{n-2}0$. Second, X contains $T\diamond^{n-1}T, T\diamond^{n-1}F, F\diamond^{n-1}T, F\diamond^{n-1}F$. Third, for each clause in our 3SAT instance, we add a word of length $n+2$. This word begins and ends with zeros, and each character in between represents the correspondingly indexed variable x_i. If x_i does not appear in a clause, the i th index of the corresponding word will be a hole. Otherwise, if x_i appears in the clause, F appears in the i th index or if the negation of x_i appears in the clause, T appears in the i th index. As an example, suppose we have $n = 4$ and the clause $x_1, x_2, \ldots, x_n \in \{T, F\}$. Then the word $0F\diamond TF0$ represents this clause in X.

Suppose there exists an assignment $x_1, x_2, \ldots, x_n \in \{T, F\}$ satisfying our 3SAT instance. Let $w = x_1 x_2 \cdots x_n$. Then we claim the one-sided infinite aperiodic word $v = 0^n w 0^{2n} w 0^{3n} w \cdots$ avoids X.
Notice that v always has precisely n truth values between blocks of zeros, so v avoids the first part of X. Additionally, since v never has more than n consecutive truth values and blocks of truth values are always separated by at least n zeros, v avoids the second part of X. Finally, the third part of X is avoided because any factor of length $n+2$ beginning and ending with zeros is precisely $0x_1 \cdots x_n 0$. This factor is not compatible with any element in the third part of X, as this would imply that the corresponding clause in the 3SAT instance is not satisfied. Thus, v avoids X.

Now suppose X is a-avoidable. Then there exists some one-sided infinite aperiodic word v avoiding X. The second part of X tells us that v must contain some zero; on the other hand, v cannot be all zeros past some point since this would be a periodic word, so v must also contain some truth value character. In particular, v must contain a zero followed by some truth value, which must be eventually followed by another zero by the second part of X. In fact, the first part of X forces precisely n truth values between these two zeros. This gives us a factor of the form $0x_1x_2 \cdots x_n 0$. This factor must avoid all the clause patterns in the third part, implying the assignment of truth values $x_1, x_2, \ldots, x_n$ satisfies all clauses of our 3SAT instance.

To extend this proof to larger size alphabets, simply include the additional letters in X. We can extend this proof to binary alphabets by using binary triples to represent each of $0, T, F$. For details on this extension, we refer the reader to the proof of Theorem 2 in [5]. □
This result has direct implication on variations of the a-AVOIDABILITY problem.

Corollary 3.
The problem of deciding the a-avoidability of a finite set of partial words where each word has equal length m over an alphabet of size $k \geqslant 2$ *is NP-hard.*

Corollary 4.
The problem of deciding the a-avoidability of a finite set of partial words where each word has an equal number of holes over an alphabet of size $k \geqslant 2$ *is NP-hard.*
Both of these corollaries can be proved by reduction from a-AVOIDABILITY by making use of the above proposition. The reduction is similar to the one of AVOIDABILITY to the corresponding variations of fixing length or number of holes.

We end this section with a generalization of the avoidability problem, AVOID-MEET, which we define to be the problem of deciding, given two finite sets of partial words X and Y over an alphabet of size $k \geqslant 2$, whether every word that avoids X meets Y. That is, if any word u avoids X, must some factor of u be compatible with a word in Y?
Proposition 10.
AVOID-MEET is co-NP-hard.

Proof.
A typical co-NP-complete problem is CO-3SAT, or UN3SAT, which is the problem of deciding given a logical formula, if all assignments of truth values renders the formula false. We reduce CO-3SAT to AVOID-MEET, using a technique similar to the one used in the proof given in [5] that AVOIDABILITY is NP-hard. Given a formula ϕ over variables $x_1, \ldots, x_n$, we want to determine if every assignment of truth values makes ϕ evaluate to FALSE. We build a set X of partial words over the alphabet $\{0, T, F\}$ such that the only possible avoiding words are periodic with each period being a concatenation of words of the form $0^n v$, where $v \in \{T, F\}^n$. The exact way to construct X is given in [5], and the size of X is polynomial with respect to n. We then build a set of partial words Y such that for every clause in ϕ, we add a word of length $2n$ to Y with 0 for the last n positions, an F in the i th position if x_i appears in the clause, and a T in the i th position if $\neg x_i$ appears in the clause.

Assume that there exists some assignment $x_1, \ldots, x_n$ such that ϕ evaluates to TRUE. Then the word $(0^n x_1 \cdots x_n)^Z$ avoids X and does not meet Y, since the assignment cannot let any of the clauses evaluate to FALSE. Conversely, if there is some word u that avoids X and does not meet Y, there is some factor of u of the form $0 u_1 \cdots u_n 0$, for $u_i \in \{T, F\}$. If we let u_i be an assignment of x_i, then this assignment does not evaluate to FALSE in any clause and so ϕ evaluates to TRUE. Thus we have reduced CO-3SAT to AVOID-MEET for alphabets of size three, and we can generalize the result to alphabets of any size of at least two by using techniques similar to the ones in [5]. □

5. A hard counting problem

In [11], Manea and Tiseanu presented a number of hard counting problems for partial words and showed them #P-complete. One of these problems is the following, which deals with counting full words, over a restricted alphabet, that are compatible with factors of a given partial word.

Problem 1.
Given a partial word w over an alphabet A with $|A| \geqslant 3$, and a symbol $\$ \in A$, count the full words $v \in (A \setminus \{\$\})^{\square}$, with $0 < |v| \leqslant |w|$, that are compatible with some factor of w.

We examine the problem of counting the full words compatible with some factor of any element in a set of partial words.

Problem 2.
Given a list of partial words $X = \{w_1, w_2, \ldots, w_n\}$ over an alphabet A with $|A| \geqslant 2$, count the words $v \in A^{\square}$ that are compatible with some factor of some w_i.
Note that we do not make restrictions on the lengths of elements of X or the lengths of factors, differentiating this problem from others presented in [11]. We show that Problem 2 is a hard counting problem by giving a Turing reduction from Problem 1.

Proposition 11.
Problem 2 *is #P-complete.*

Proof.
We first show that the problem is in $\#P$. Note that any $v \in A^{\square}$ that is compatible with a factor of an element of X must have length less than or equal to the length of the longest element of X. Then there are only finitely many possible words that can be compatible with a factor of an element of X. We can create a non-deterministic Turing machine that non-deterministically guesses a word v that satisfies the length bound and checks if v is compatible with a factor of an element of X. This check can be done in polynomial time, and this Turing machine has exactly as many accepting paths as the number of words v that are compatible with a factor of X, so the problem is in $\#P$. We next must show that it is complete for the class.

Assume that there exists a function solve(X) that can compute a solution to Problem 2 efficiently, taking as input the set X. Consider an instance of Problem 1: w is a partial word over some alphabet A with $|A| \geqslant 3$ and \$ is a symbol in A. We construct our set X by stepping through w. Every time we encounter the symbol \$, we end an element. For example, the word $w_1\$\$\$w_2\$w_3\$$ where w_1, w_2, w_3 are partial words over $A \setminus \{\$\}$ becomes the set $\{w_1, w_2, w_3\}$. Clearly we can manage this operation in polynomial time.

Since we count factors excluding the \$ symbol, by construction no factor counted for Problem 1 can cross between elements of X. Additionally, since each $w_i \in X$ is a factor of our original word w, no extra factors have been introduced. Thus, by running solve on our constructed set, we obtain the answer to Problem 1. Then if there exists an efficient solution to Problem 2, we must also have an efficient solution to Problem 1. However, since Problem 1 has been shown to be $\#P$-complete, Problem 2 must also be $\#P$-complete. □

6. Conclusion

In this paper, we have given bounds on the minimum cardinality of an unavoidable set of partial words of constant length m with h holes over an alphabet of size k. We have given an alternative proof that the problem of deciding the avoidability of a finite set of partial words over an arbitrary alphabet can be solved in polynomial space. We have analyzed the complexity of variations on this problem with restrictions on the number of holes and length of the words. We have extended the concept of aperiodic avoidability to sets of partial words and have analyzed the complexity of deciding if a finite set of partial words over a k-letter alphabet is avoided by an aperiodic one-sided infinite word. We have also proved that counting the full words that are compatible with some factor of some element in a given set of partial words is #P-complete.
Another problem related to the complexity of deciding avoidability is that of computing bounds on the minimum period of words avoiding sets of partial words. Blakeley et al. [2] found a polynomial bound for sets of full words. We gave a polynomial bound for sets with a fixed number of holes. Proving a polynomial bound for the minimum period, without the hole restriction, would give a proof of the membership of AVOIDABILITY in NP.

Acknowledgments

The authors would like to acknowledge Narad Rampersad from the Department of Mathematics of the University of Winnipeg for providing us with another proof of the membership of AVOIDABILITY in PSPACE (see Proposition 3). We thank him for his contribution. We also thank the referees for their very valuable comments and suggestions.

A World Wide Web server interface has been established at www.uncg.edu/cmp/research/complexity for automated use of a program that takes as input the length, number of holes, and alphabet size for partial words in a set and returns bounds on the minimum cardinality of an unavoidable set with those parameters.

References

[1] A.V. Aho, M.J. Corasick, Efficient string machines, an aid to bibliographic research, Communications of the ACM 18 (1975) 333–340.

[2] B. Blakeley, F. Blanchet-Sadri, J. Gunter, N. Rampersad, On the complexity of deciding avoidability of sets of partial words, Theoretical Computer Science 411 (2010) 4263–4271.

[3] F. Blanchet-Sadri, N.C. Brownstein, A. Kalcic, J. Palumbo, T. Weyand, Unavoidable sets of partial words, Theory of Computing Systems 45 (2) (2009) 381–406.

[4] F. Blanchet-Sadri, B. Chen, A. Chakarov, Minimum number of holes in unavoidable sets of partial words of size three, in: C.S. Iliopoulos, W.F. Smyth (Eds.), IWOCA 2010, 21st International Workshop on Combinatorial Algorithms, London, United Kingdom, in: Lecture Notes in Computer Science, vol. 6460, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 43–55.

[5] F. Blanchet-Sadri, R.M. Jungers, J. Palumbo, Testing avoidability on sets of partial words is hard, Theoretical Computer Science 410 (2009) 968–972.
[6] J.M. Champarnaud, G. Hansel, D. Perrin, Unavoidable sets of constant length, International Journal of Algebra and Computation 14 (2004) 241–251.

[7] C. Choffrut, J. Karhumäki, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, vol. 1, Springer-Verlag, Berlin,
1997, pp. 329–438 (Chapter 6).

[8] M.R. Garey, D.S. Johnson, Computers and Intractability—A Guide to the Theory of NP-Completeness, Freeman, 1979.

[9] P.M. Higgins, C.J. Saker, Unavoidable sets, Theoretical Computer Science 359 (2006) 231–238.

[10] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.

[11] F. Manea, C. Tiseanu, Hard counting problems for partial words, in: A.-H. Dediu, H. Fernau, C. Martín-Vide (Eds.), LATA 2010, 4th International Conference on Language and Automata Theory and Applications, Trier, Germany, in: Lecture Notes in Computer Science, vol. 6031, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 426–438.

[12] J. Mykkeltveit, A proof of Golomb's conjecture for the de Bruijn graph, Journal of Combinatorial Theory, Series B 13 (1972) 40–45.

[13] C.J. Saker, P.M. Higgins, Unavoidable sets of words of uniform length, Information and Computation 173 (2002) 222–226.

[14] M.P. Schützenberger, On the synchronizing properties of certain prefix codes, Information and Control 7 (1964) 23–36.

[15] J. Shallit, A Second Course in Formal Languages and Automata Theory, Cambridge University Press, 2009.

[16] R. Tarjan, Depth-first search and linear graph algorithms, SIAM Journal on Computing 1 (2) (1972) 146–160.