On the Iterated Hairpin Completion

Steffen Kopecki

Universität Stuttgart,
Institut für Formale Methoden der Informatik
Universitätsstr. 38, D-70569 Stuttgart, Germany
steffen.kopecki@fmi.uni-stuttgart.de

May 5, 2010

Abstract

The hairpin completion is a natural operation on formal languages which has been inspired by biochemistry and DNA-computing. In this paper we solve two problems which were posed first in 2008 and 2009, respectively, and still left open:

1.) It is known that the iterated hairpin completion of a regular language is not context-free in general, but it was open whether the iterated hairpin completion of a singleton or finite language is regular or at least context-free. We will show that it can be non-context-free. (It is of course context-sensitive.)

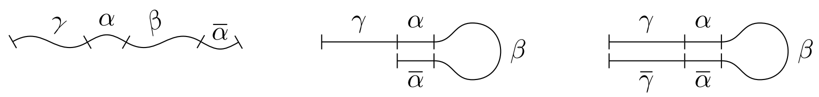
2.) A restricted but also very natural variant of the hairpin completion is the bounded hairpin completion. It was unknown whether the iterated bounded hairpin completion of a regular language remains regular. We prove that this is indeed the case. Actually we derive a more general result. We will present a general representation of the iterated bounded hairpin completion for any language using basic operations. Thus, each language class closed under these basic operations is also closed under iterated bounded hairpin completion.

Keywords: Formal Languages, Hairpin Completion

1 Introduction

The inspiration of the hairpin completion is rooted in DNA-computing and biochemistry, where it appears naturally in chemical reactions. It turned out that the corresponding operation on formal languages gives rise to very interesting and quite subtle decidability and computational problems. The focus of our paper is therefore on these formal language theoretical results. However, let us sketch the biochemical origin of this operation first.

A *DNA strand* (or simply strand) is a polymer composed of nucleotides which differ from each other by their bases A (adenine), C (cytosine), G (guanine) and T (thymine). A strand can be seen as a finite sequence of bases. By *Watson-Crick base pairing* two base sequences can bind to each other if they are pairwise complementary where A is complementary to T and C to G. The hairpin completion is best explained by Figure 1. By the base sequence $\overline{\alpha}$ we mean to read α from right to left and complement its bases. During chemical processes a strand which contains a sequence α and ends on the complementary sequence $\overline{\alpha}$ (a) can form an intramolecular base-pairing, which is known as *hairpin* (b). By complementing the unbound sequence γ, the *hairpin completion* (c) arises.



(a) strand (b) hairpin (c) hairpin completion

Figure 1: Hairpin completion of a DNA-strand.

On an abstract level a strand can be seen as a word and a (possibly infinite) set of strands is a language. The hairpin completion of formal languages has been introduced in [1]. In several papers the hairpin completion and some familiar operations have been studied, see [1, 2, 4–8]. Here we focus on the iterated versions of the hairpin completion and the bounded hairpin completion. For the latter operation we assume the length of the γ-part to be bounded. A formal definition of both operations is given in Section 2.1.

The class of iterated hairpin completions of singletons (HCS) has been investigated in [6]. HCS is included in the class of context-sensitive languages since context-sensitive languages are closed under iterated hairpin completion, see [1]. However, the questions if HCS contains non-regular or non-context-free languages has been unsolved. In Section 3 we answer both questions with "yes".

In Section 4 we state a general representation for the iterated bounded hairpin completion of a formal language using the operations union, intersection with regular sets, and concatenation with regular sets. As a consequence all language classes which are closed under these basic operations are also closed under iterated bounded hairpin completion. This solves the problem whether the iterated bounded hairpin completion of a regular language is always regular which was stated in [4].

An extended abstract of this work will be presented at the 14th International Conference on Developments in Language Theory (DLT 2010).

2 Notation

We assume the reader to be familiar with the fundamental concepts of formal language theory, see [3]. Our focus is on regular and context-free languages.

An *alphabet* is a finite set of *letters*. In this paper the alphabet is always Σ. The set of words over Σ is denoted by Σ^* as usual, and the *empty word* is denoted by 1. We consider Σ with an *involution*, this is a bijection $\overline{} : \Sigma \to \Sigma$ such that $\overline{\overline{a}} = a$ for all $a \in \Sigma$. In DNA-biochemistry $\Sigma = \{A, C, G, T\}$ with $\overline{A} = T$ and $\overline{C} = G$. We extend the involution to words $w = a_1 \cdots a_n$ by $\overline{w} = \overline{a_n} \cdots \overline{a_1}$. (Just like taking inverses in groups.) Let L be a formal language, by $\overline{L}$ we denote the language $\{\overline{w} \mid w \in L\}$.

Given a word w, we denote by $|w|$ its length. If $w = xyz$ for some $x, y, z \in \Sigma^*$, then x, y, and z are called *prefix*, *infix*, and *suffix*, respectively. For the prefix relation we also use the notation $x \leq w$. Note that, if z is a suffix of w, then $\overline{z}$ is a prefix of $\overline{w}$ or $\overline{z} \leq \overline{w}$.

2.1 The hairpin completion

If a word w has a factorization $w = \gamma\alpha\beta\overline{\alpha}$, then the suffix $\overline{\alpha}$ can bind to the infix α to form a hairpin and the new suffix $\overline{\gamma}$ can be created, again see Fig. 1. The word $\gamma\alpha\beta\overline{\alpha}\overline{\gamma}$ is called a *right hairpin completion* of w. Since in biochemistry a hairpin is stable only if the complementary parts are long enough, we fix a small constant $k \geq 1$ and ask $|\alpha| = k$.

Symmetrically, if w has a factorization $w = \alpha\beta\overline{\alpha}\overline{\gamma}$ then $\gamma\alpha\beta\overline{\alpha}\overline{\gamma}$ is a *left hairpin completion* of w. If we simply speak of a hairpin completion we mean either a left or a right hairpin completion. The hairpin completion of a formal language L is the union of all hairpin completions of all words in L.

Instead of defining the bounded hairpin completion as it was defined in former works, we will use a more general operation. This will come in handy in Section 4. The *parameterized hairpin completion* is a variant of the hairpin completion where we allow length-bounds for the γ-parts. For $\ell \geq 0$ let $\Gamma_\ell = \{\gamma \in \Sigma^* \mid |\gamma| \leq \ell\}$ be the language containing all words of at most length ℓ and let α be some word of length k. The left- and right-sided parameterized hairpin completion of a language L are defined as

$$\mathcal{H}_\alpha(L, \ell, 0) = \bigcup_{\gamma \in \Gamma_\ell} \gamma(L \cap \alpha\Sigma^*\overline{\alpha}\overline{\gamma}) \quad \text{and} \quad \mathcal{H}_\alpha(L, 0, r) = \bigcup_{\gamma \in \Gamma_r} (L \cap \gamma\alpha\Sigma^*\overline{\alpha})\overline{\gamma}$$

for $\ell, r \geq 0$, respectively. The *(two-sided) parameterized hairpin completion* is their union

$$\mathcal{H}_\alpha(L, \ell, r) = \mathcal{H}_\alpha(L, \ell, 0) \cup \mathcal{H}_\alpha(L, 0, r).$$

For the constant k let

$$\mathcal{H}_k(L, \ell, r) = \bigcup_{\alpha \in \Sigma^k} \mathcal{H}_\alpha(L, \ell, r).$$

In this paper we are interested in three variants of the hairpin completion:

1. The one-sided (unbounded) k-hairpin completion

$$os\text{-}\mathcal{H}_k(L) = \bigcup_{r \geq 0} \mathcal{H}_k(L, 0, r) = \{\gamma\alpha\beta\overline{\alpha}\overline{\gamma} \mid \gamma\alpha\beta\overline{\alpha} \in L \wedge |\alpha| = k\}.$$

2. The (two-sided unbounded) k-hairpin completion

$$\mathcal{H}_k(L) = \bigcup_{\ell, r \geq 0} \mathcal{H}_k(L, \ell, r) = \{\gamma\alpha\beta\overline{\alpha}\overline{\gamma} \mid (\gamma\alpha\beta\overline{\alpha} \in L \vee \alpha\beta\overline{\alpha}\overline{\gamma} \in L) \wedge |\alpha| = k\}.$$

3. The ℓ-bounded k-hairpin completion

$$\mathcal{H}_k(L, \ell, \ell) = \{\gamma\alpha\beta\overline{\alpha}\overline{\gamma} \mid (\gamma\alpha\beta\overline{\alpha} \in L \vee \alpha\beta\overline{\alpha}\overline{\gamma} \in L) \wedge |\gamma| \leq \ell \wedge |\alpha| = k\}$$

for a bound $\ell \geq 0$.

By applying an arbitrary number of parameterized hairpin completions to a language we obtain the iterated parameterized hairpin completions

$$\mathcal{H}_\alpha^*(L,\ell,r) = \bigcup_{i\geq 0} \mathcal{H}_\alpha^i(L,\ell,r) \qquad \text{and}$$
$$\mathcal{H}_k^*(L,\ell,r) = \bigcup_{i\geq 0} \mathcal{H}_k^i(L,\ell,r)$$

where

$$\mathcal{H}_\alpha^0(L,\ell,r) = L, \qquad \mathcal{H}_\alpha^i(L,\ell,r) = \mathcal{H}_\alpha(\mathcal{H}_\alpha^{i-1}(L,\ell,r),\ell,r),$$
$$\mathcal{H}_k^0(L,\ell,r) = L, \quad \text{and} \quad \mathcal{H}_k^i(L,\ell,r) = \mathcal{H}_k(\mathcal{H}_k^{i-1}(L,\ell,r),\ell,r) \qquad \text{for } i \geq 1.$$

If a word z is included in $\mathcal{H}_k^i(\{w\},\ell,r)$ we say that z can be created with i hairpin completions from w. Fig. 2 shows an example how a word can be created with three hairpin completions from $\alpha u\overline{\alpha}v\alpha$ where $|\alpha| = k$. In each step the dotted part is the newly created prefix or suffix.

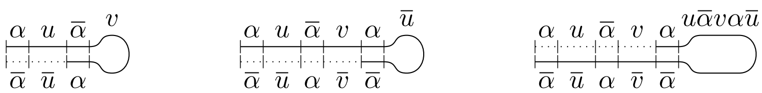


Figure 2: Example for the iterated hairpin completion.

Again, the three cases

1. iterated one-sided k-hairpin completion

$$os\text{-}\mathcal{H}_k^*(L) = \bigcup_{r\geq 0} \mathcal{H}_k^*(L,0,r),$$

2. iterated k-hairpin completion

$$\mathcal{H}_k^*(L) = \bigcup_{\ell,r\geq 0} \mathcal{H}_k^*(L,\ell,r),$$

3. and iterated ℓ-bounded k-hairpin completion $\mathcal{H}_k^*(L,\ell,\ell)$

are of main interest.

3 The iterated hairpin completion of singletons

In this section we will state a singleton whose iterated hairpin completion is not context-free. This solves the problem whether the class HCS contains non-regular or non-context-free languages which was stated in [6]. Moreover, we will show that the result also holds if we consider the iterated one-sided hairpin completion.

Theorem 3.1. *The iterated one- and two-sided hairpin completion of a singleton (or finite language) is context-sensitive and is not context-free in general.*

Proof. It is simple to construct a linear bounded automaton (LBA) which accepts the iterated one- or two-sided hairpin completion of a finite language. This yields the membership to the class of context-sensitive languages. (Or see [1] where the closure of context-sensitive languages under iterated hairpin completion is proved.)

Now we present a witness that the iterated hairpin completion of a singleton may be non-context-free. Let $\Sigma = \{a, \bar{a}, b, \bar{b}, c, \bar{c}\}$, $\alpha = a^k$ and

$$w = a^k b a^k \bar{a}^k a^k c \bar{a}^k = \alpha b \alpha \bar{\alpha} \alpha c \bar{\alpha}.$$

We will prove that the languages $\mathcal{H}_k^*(\{w\})$ and $os\text{-}\mathcal{H}_k^*(\{w\})$ are non-context-free. Since context-free languages are closed under intersection with regular sets, it is enough to show that for some regular set $\mathcal{R}$ the intersections $\mathcal{R} \cap \mathcal{H}_k^*(\{w\})$ and $\mathcal{R} \cap os\text{-}\mathcal{H}_k^*(\{w\})$ are non-context-free.

Let $u = \alpha\bar{\alpha}\bar{b}\bar{\alpha}$ and $v = \bar{b}\bar{\alpha}$. Note that $\bar{v} \leq \bar{u} \leq w$. Let

$$\mathcal{R} = wv^+ u \bar{v}^+ \overline{w} \bar{v}^+ \overline{w}.$$

and consider a word $z \in \mathcal{R}$:

$$z = \underbrace{\alpha b \alpha \bar{\alpha} \alpha c \bar{\alpha}}_{w} \underbrace{(\bar{b}\bar{\alpha})^r}_{v^r} \underbrace{\alpha \bar{\alpha} \bar{b} \bar{\alpha}}_{u} \underbrace{(\alpha b)^s}_{\bar{v}^s} \underbrace{\alpha \bar{c} \bar{\alpha} \alpha \bar{\alpha} \bar{b} \bar{\alpha}}_{\overline{w}} \underbrace{(\alpha b)^t}_{\bar{v}^t} \underbrace{\alpha \bar{c} \bar{\alpha} \alpha \bar{\alpha} \bar{b} \bar{\alpha}}_{\overline{w}}$$

with $r, s, t \geq 1$. We will show that z can be created with hairpin completions from w if and only if $r = s = t$.

The word w is a prefix of z and it occurs at no other position in z (there is only one c in z). The only way to create z from w is by using right hairpin completions. Hence, $z \in \mathcal{H}_k^*(\{w\})$ if and only if $z \in os\text{-}\mathcal{H}_k^*(\{w\})$ and

$$\mathcal{R} \cap \mathcal{H}_k^*(\{w\}) = \mathcal{R} \cap os\text{-}\mathcal{H}_k^*(\{w\}).$$

In every step we have to bind the suffix $\bar{\alpha}$ to some infix α. This leads to two possible hairpin completions for

$$wv^j = \alpha b \underset{I}{\underline{\alpha}} \bar{\alpha} \underset{II}{\underline{\alpha}} c \bar{\alpha} (\bar{b}\bar{\alpha})^j$$

with $j \geq 0$. (We ignore the useless option of binding the suffix to the prefix.) Option *I* creates the new suffix v and option *II* creates the new suffix u. The prefix $wv^r u$ can be created from w in exactly one way: Use r times option *I* and then one time option *II*.

Now we have a third option for the hairpin, the prefix α of u:

$$wv^r u = \alpha b \underset{I}{\underline{\alpha}} \bar{\alpha} \underset{II}{\underline{\alpha}} c \bar{\alpha} (\bar{b}\bar{\alpha})^r \underset{III}{\underline{\alpha}} \bar{\alpha} \bar{b} \bar{\alpha}.$$

Every right hairpin completion creates a new suffix $x\bar{b}\bar{\alpha}$ for some $x \in \Sigma^*$. Since there are only two occurrences of $\bar{b}\bar{\alpha}$ in

$$\bar{v}^s \overline{w} \bar{v}^t \overline{w} = (\alpha b)^s \alpha \bar{c} \bar{\alpha} \alpha \bar{\alpha} \underline{\bar{b}\bar{\alpha}} (\alpha b)^t \alpha \bar{c} \bar{\alpha} \alpha \bar{\alpha} \underline{\bar{b}\bar{\alpha}},$$

z has to be created with two hairpin completions from $wv^r u$ which is only possible by using option *III* two times and if $r = s = t$. (After the first use of option *III* we get a lot more options for the last hairpin completion but, as you can easily verify, none of the other options will lead to z.)

We conclude that a word $wv^r u\bar{v}^s \overline{w}\bar{v}^t \overline{w}$ with $r, s, t \geq 1$ is in $\mathcal{H}_k^*(\{w\})$ if and only if $r = s = t$, and the intersections

$$\mathcal{R} \cap \mathcal{H}_k^*(\{w\}) = \mathcal{R} \cap os\text{-}\mathcal{H}_k^*(\{w\}) = \{wv^r u\bar{v}^r \overline{w}\bar{v}^r \overline{w} \mid r \geq 1\}$$

belong to a family of context-sensitive languages which are well known to be non-context-free. Hence, the iterated one- and two-sided hairpin completion of the singleton $\{w\}$ is non-context-free, too. □

After having proved that HCS contains non-regular and non-context-free languages two new questions naturally arise:

1. Does a singleton exits whose iterated hairpin completion is context-free but non-regular?
2. Can we decide whether the iterated hairpin completion of singleton is non-regular (or non-context-free)?

4 The iterated parameterized hairpin completion

In this section we will give another representation for the iterated parameterized hairpin completion.

Theorem 4.1. *Let L be a formal language and $\ell, r \geq 0$. The iterated parameterized hairpin completion $\mathcal{H}_k^*(L, \ell, r)$ can be represented by an expression using L and the operations union, intersection with regular sets, and concatenation with regular sets.*

Consequentially, all language classes which are closed under these operations — which includes the classes in the Chomsky Hierarchy — are closed under iterated parameterized hairpin completion. From [4] it has been known that the classes of context-free, context-sensitive, and recursively enumerable languages are closed under iterated bounded hairpin completion, but it has been left open whether this is also true for regular languages. Since the iterated bounded hairpin completion is a special case of the iterated parameterized hairpin completion we proved that this is indeed the case.

Corollary 4.2. *The class of regular languages is closed under iterated bounded hairpin completion.*

Another interesting problem for HCS is the question if one can decide whether the iterated hairpin completions of two singleton languages have a non-empty intersection. Since the results of this paper yield that the iterated bounded hairpin completion of a singleton is regular and an NFA accepting this language can effectively be constructed, we can at least decide this problem for the bounded case.

4.1 α-prefixes

Let us introduce the concept of *α-prefixes* which is essential for the representation of the iterated parameterized hairpin completion. Let α be a word of length k. For $p, w \in \Sigma^*$ we say p is an α-prefix of w if $p\alpha \leq w$. We denote the set of all α-prefixes of a maximal length ℓ by

$$\mathcal{P}_\alpha(w, \ell) = \{p \mid p\alpha \leq w \wedge |p| \leq \ell\}.$$

We can use this notation to represent the (non-iterated) parameterized hairpin completion of a singleton. Let $\ell \geq 0$ and $w \in \alpha\Sigma^*$ with $|w| \geq \ell + 2k$. The left hairpin completion of $\{w\}$ is

$$\mathcal{H}_\alpha(\{w\}, \ell, 0) = \mathcal{P}_\alpha(\overline{w}, \ell)w.$$

Symmetrically, let $r \geq 0$ and $w \in \Sigma^*\overline{\alpha}$ with $|w| \geq r + 2k$. The right hairpin completion of $\{w\}$ is

$$\mathcal{H}_\alpha(\{w\}, 0, r) = w\overline{\mathcal{P}_\alpha(w, r)}.$$

For the proof of Theorem 4.1 we mainly consider α-prefixes of words which have α as a prefix. In this case some useful properties arise.

Lemma 4.3. *Let $\alpha \in \Sigma^k$, $\ell \geq 0$, and $w \in \alpha\Sigma^*$.*

1. *For all $p \in \mathcal{P}_\alpha(w, \ell)$ we have $\alpha \leq p\alpha$.*
2. *For all $p_1, p_2 \in \mathcal{P}_\alpha(w, \ell)$ we have*

$$|p_1| \leq |p_2| \;\Leftrightarrow\; p_1\alpha \leq p_2\alpha \;\Leftrightarrow\; p_1 \in \mathcal{P}_\alpha(p_2\alpha, |p_2|).$$

3. *If $v\alpha$ is a prefix of some word in $\mathcal{P}_\alpha(w, \ell)^*\alpha$ then $v \in \mathcal{P}_\alpha(w, \ell)^*$.*

Proof. If two words x, y are prefixes of w and $|x| \leq |y|$, then $x \leq y$. This yields property 1 and 2.

For property 3 let $v\alpha \leq p_1 \cdots p_m\alpha$ where $p_1, \ldots, p_m \in \mathcal{P}_\alpha(w, \ell)$. We can factorize $v = p_1 \cdots p_{i-1}v'$ such that $v' \leq p_i$ for some $1 \leq i \leq m$. By property 1 and induction, we see that α is a prefix of $p_{i+1} \cdots p_m\alpha$ and therefore $v'\alpha \leq p_i\alpha \leq w$. Moreover, $v' \in \mathcal{P}_\alpha(w, \ell)$ and $v \in \mathcal{P}_\alpha(w, \ell)^*$. □

4.2 Proof of Theorem 4.1

Let the language L and $\ell, r \geq 0$ be fixed for the rest of this section. We will state a representation for $\mathcal{H}_k^*(L, \ell, r)$. Every word w in the (non-iterated) parameterized hairpin completion $\mathcal{H}_k(L, \ell, r)$ has a factorization $w = \gamma\delta\beta\overline{\delta}\overline{\gamma}$ where $|\delta| = k$. Let α be the prefix of length k of w then $\overline{\alpha}$ is a suffix of w. If we apply one parameterized hairpin completion to w this will lead to a new word

$$w' \in \mathcal{H}_k(\{w\}, \ell, r) = \mathcal{H}_\alpha(\{w\}, \ell, r) \subseteq w\overline{\mathcal{P}_\alpha(w, r)} \cup \mathcal{P}_\alpha(\overline{w}, \ell)w \subseteq \alpha\Sigma^*\overline{\alpha}.$$

For the second inclusion we use Lemma 4.3. By induction,

$$\mathcal{H}_k^*(\{w\}, \ell, r) = \mathcal{H}_\alpha^*(\{w\}, \ell, r) \subseteq \alpha\Sigma^*\overline{\alpha}.$$

For $\alpha \in \Sigma^k$ let

$$L_\alpha = \mathcal{H}_k(L, \ell, r) \cap \alpha\Sigma^*\overline{\alpha}.$$

With the observations made above, it is plain that

$$\mathcal{H}_k(L,\ell,r) = \bigcup_{\alpha\in\Sigma^k} L_\alpha,$$
$$\mathcal{H}_k^*(L_\alpha,\ell,r) = \mathcal{H}_\alpha^*(L_\alpha,\ell,r) \subseteq \alpha\Sigma^*\overline{\alpha},$$

and each language L_α has a representation using L and the operations union, intersection with regular sets, and concatenation with regular sets. The iterated parameterized hairpin completion of L is

$$\begin{aligned}\mathcal{H}_k^*(L,\ell,r) &= L\cup\mathcal{H}_k^*(\mathcal{H}_k(L,\ell,r),\ell,r)\\ &= L\cup\mathcal{H}_k^*\Big(\bigcup_{\alpha\in\Sigma^k} L_\alpha,\ell,r\Big)\\ &= L\cup\bigcup_{\alpha\in\Sigma^k}\mathcal{H}_\alpha^*(L_\alpha,\ell,r).\end{aligned}$$

In order to prove Theorem 4.1 we will state a representation for $\mathcal{H}_\alpha^*(L_\alpha,\ell,r)$ for any $\alpha\in\Sigma^k$. From now on let $\alpha\in\Sigma^k$ be fixed. For the rest of the proof we will heavily rely on the fact that every word in $\mathcal{H}_\alpha^*(L_\alpha,\ell,r)$ has the prefix α and the suffix $\overline{\alpha}$. We will define the representation recursively. Note that

$$\mathcal{H}_\alpha^*(L_\alpha,0,0) = L_\alpha.$$

By symmetry we may assume $\ell\geq r$ and $\ell\geq 1$. Let

$$z\in\mathcal{H}_\alpha^*(L_\alpha,\ell,r)\setminus\mathcal{H}_\alpha^*(L_\alpha,\ell-1,r).$$

This word can be created with $s\geq 1$ hairpin completions from some word in $w\in\mathcal{H}_\alpha^*(L_\alpha,\ell-1,r)$ such that the first hairpin completion creates a new prefix x_1 of length ℓ. Hence, z has a factorization

$$z = x_s\cdots x_1 w\overline{y_1}\cdots\overline{y_s}$$

where for all $1\leq i\leq s$:

1. If the i-th hairpin completion is a left hairpin completion, then $y_i=1$ and x_i is the prefix that is created. Therefore $|x_i|\leq\ell$ and
$$x_{i-1}\cdots x_1 w\overline{y_1}\cdots\overline{y_{i-1}}\in\alpha\Sigma^*\overline{\alpha}\overline{x_i}.$$

2. If the i-th hairpin completion is a right hairpin completion, then $x_i=1$ and $\overline{y}_i$ is the suffix that is created. Therefore $|y_i|\leq r$ and
$$x_{i-1}\cdots x_1 w\overline{y_1}\cdots\overline{y_{i-1}}\in y_i\alpha\Sigma^*\overline{\alpha}.$$

The crucial point is that the word x_1wy_1 has the prefix $x_1\alpha$, the suffix $\overline{\alpha}\overline{x_1}$, and $|x_1|=\ell\geq r$. Therefore, the words $x_2,\ldots,x_s$ and $y_2,\ldots,y_s$ are controlled by x_1, ℓ, and r.

Lemma 4.4. *For all $1\leq i\leq s$ we have $x_i\in\mathcal{P}_\alpha(x_1\alpha,\ell)^*$ and $y_i\in\mathcal{P}_\alpha(x_1\alpha,r)^*$.*

Proof. Obviously $x_1 \in \mathcal{P}_\alpha(x_1\alpha, \ell)^*$ and $y_1 = 1 \in \mathcal{P}_\alpha(x_1\alpha, r)^*$.

Let $2 \leq i \leq s$. By induction, we may assume that $x_j \in \mathcal{P}_\alpha(x_1\alpha, \ell)^*$ and $y_j \in \mathcal{P}_\alpha(x_1\alpha, r)^*$ for all $1 \leq j < i$.

1. If the i-th hairpin completion is a left hairpin completion, we have $y_i = 1 \in \mathcal{P}_\alpha(x_1\alpha, r)^*$ and, by induction hypothesis,

$$x_i\alpha \leq y_{i-1} \cdots y_1 x_1 \alpha \in \mathcal{P}_\alpha(x_1\alpha, r)^* x_1\alpha \subseteq \mathcal{P}_\alpha(x_1\alpha, \ell)^*\alpha.$$

   Together with Lemma 4.3 property 3 this leads to $x_i \in \mathcal{P}_\alpha(x_1\alpha, \ell)^*$.

2. Otherwise we have $x_i = 1 \in \mathcal{P}_\alpha(x_1\alpha, \ell)^*$ and

$$y_i\alpha \leq x_{i-1} \cdots x_1\alpha \in \mathcal{P}_\alpha(x_1\alpha, \ell)^*\alpha,$$

   hence $y_i \in \mathcal{P}_\alpha(x_1\alpha, \ell)^*$. Since $|y_i| \leq r$, all the factors of y_i which are included in $\mathcal{P}_\alpha(x_1\alpha, \ell)$ are also at most of length r and therefore $y_i \in \mathcal{P}_\alpha(x_1\alpha, r)^*$.

□

Now let us define the languages

$$\mathcal{L}_\alpha(x, \ell, r) = \mathcal{P}_\alpha(x\alpha, \ell)^* x \left(\mathcal{H}^*_\alpha(L, \ell - 1, r) \cap \alpha\Sigma^*\overline{\alpha}\overline{x}\right) \overline{\mathcal{P}_\alpha(x\alpha, r)}^*.$$

By Lemma 4.4, the word z is included in $\mathcal{L}_\alpha(x_1, \ell, r)$. Since every word

$$z' \in \mathcal{H}^*_\alpha(L_\alpha, \ell, r) \setminus \mathcal{H}^*_\alpha(L_\alpha, \ell - 1, r)$$

has a factorization as above, there exists an $x' \in \Sigma^\ell$ such that $z' \in \mathcal{L}_\alpha(x', \ell, r)$ and, moreover,

$$\mathcal{H}^*_\alpha(L_\alpha, \ell, r) \subseteq \mathcal{H}^*_\alpha(L_\alpha, \ell - 1, r) \cup \bigcup_{x \in \Sigma^\ell} \mathcal{L}_\alpha(x, \ell, r).$$

Of course we intend to replace the inclusion by an equal sing.

Lemma 4.5. *$\mathcal{L}_\alpha(x, \ell, r) \subseteq \mathcal{H}^*_\alpha(L_\alpha, \ell, r)$ for all $x \in \Sigma^\ell$.*

Proof. For some $x \in \Sigma^\ell$ consider a word $z \in \mathcal{L}_\alpha(x, \ell, r)$ with the factorization

$$z = u_s \cdots u_1 w \overline{v_1} \cdots \overline{v_t}$$

where

- $w \in x\left(\mathcal{H}^*_\alpha(L, \ell - 1, r) \cap \alpha\Sigma^*\overline{\alpha}\overline{x}\right) \subseteq \mathcal{H}^*_\alpha(L_\alpha, \ell, r) \cap x\alpha\Sigma^*\overline{\alpha}\overline{x}$,
- $u_1, \ldots, u_s \in \mathcal{P}_\alpha(x\alpha, \ell)$, and
- $v_1, \ldots, v_t \in \mathcal{P}_\alpha(x\alpha, r)$.

We will prove that z can be created with $s + t$ hairpin completions from w.

Let p be the longest α-prefix of $x\alpha$ which is equal to one of the factors $u_1, \ldots, u_s$ or $v_1, \ldots, v_t$ and let m be the largest index such that $u_m = p$ (or 0 if no such index exists). Respectively, let n be the largest index such that

$v_n = p$. The word $w' = u_m \cdots u_1 w \overline{v_1} \cdots \overline{v_n}$ can be created with $m + n$ hairpin completions as follows:

Since $\overline{\alpha}\overline{x}$ is a suffix of w and $u_1, \dots, u_m \in \mathcal{P}_\alpha(x\alpha, \ell)$, we can create the word $u_m \cdots u_1 w$ with m hairpin completions from w. The new word has a prefix $p\alpha$ (even if $m = 0$ since $p\alpha \leq x\alpha$) and since $|v_i| \leq |p|$ for all $1 \leq i \leq n$, we have $v_i \in \mathcal{P}_\alpha(p\alpha, r)$, cf. Lemma 4.3. Hence, w' can be created with n hairpin completions from $u_m \cdots u_1 w$.

Now, $z = u_t \cdots u_{m+1} w' \overline{u_{n+1}} \cdots \overline{u_t}$ where

- $w' \in \mathcal{H}^*_\alpha(L_\alpha, \ell, r) \cap p\alpha\Sigma^*\overline{\alpha}\overline{p}$,
- $u_{m+1}, \dots, u_s \in \mathcal{P}_\alpha(p\alpha, |p| - 1)$, and
- $v_{n+1}, \dots, v_t \in \mathcal{P}_\alpha(p\alpha, \min\{|p| - 1, r\})$.

We can resume to create z from w' inductively. □

By Lemma 4.5, if $\ell \geq r$ the iterated parameterized hairpin completion of L_α can be represented by

$$\mathcal{H}^*_\alpha(L_\alpha, \ell, r) = \mathcal{H}^*_\alpha(L_\alpha, \ell - 1, r) \cup \bigcup_{x \in \Sigma^\ell} \mathcal{L}_\alpha(x, \ell, r).$$

Symmetrically, if $r > \ell$ let us define

$$\mathcal{R}_\alpha(y, \ell, r) = \mathcal{P}_\alpha(y\alpha, \ell)^* \left(\mathcal{H}^*_\alpha(L, \ell, r - 1) \cap y\alpha\Sigma^*\overline{\alpha}\right) \overline{y}\,\overline{\mathcal{P}_\alpha(y\alpha, r)}^*.$$

The iterated parameterized hairpin completion of L_α can be represented by

$$\mathcal{H}^*_\alpha(L_\alpha, \ell, r) = \mathcal{H}^*_\alpha(L_\alpha, \ell, r - 1) \cup \bigcup_{y \in \Sigma^r} \mathcal{R}_\alpha(y, \ell, r).$$

Hence, we can give a representation for the iterated parameterized hairpin completion using L and the operations union, intersection with regular sets, and concatenation with regular sets.

Acknowledgements

I would like to thank Volker Diekert for many suggestions on the presentation of this paper, Florin Manea for numerous helpful discussions, and the anonymous referees of the DLT committee for several useful comments.

References

[1] D. Cheptea, C. Martin-Vide, and V. Mitrana. A new operation on words suggested by DNA biochemistry: Hairpin completion. *Transgressive Computing*, pages 216–228, 2006.

[2] V. Diekert, S. Kopecki, and V. Mitrana. On the hairpin completion of regular languages. In M. Leucker and C. Morgan, editors, *ICTAC*, volume 5684 of *Lecture Notes in Computer Science*, pages 170–184. Springer, 2009.

[3] J. E. Hopcroft and J. D. Ulman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 1979.

[4] M. Ito, P. Leupold, and V. Mitrana. Bounded hairpin completion. In *LATA '09: Proceedings of the 3rd International Conference on Language and Automata Theory and Applications*, pages 434–445, Berlin, Heidelberg, 2009. Springer-Verlag.

[5] F. Manea and V. Mitrana. Hairpin completion versus hairpin reduction. In S. B. Cooper, B. Löwe, and A. Sorbi, editors, *CiE*, volume 4497 of *Lecture Notes in Computer Science*, pages 532–541. Springer, 2007.

[6] F. Manea, V. Mitrana, and T. Yokomori. Some remarks on the hairpin completion. In E. Csuhaj-Varju and Z. Esik, editors, *12th International Conference AFL 2008 Proceedings*, pages 302–312, 2008.

[7] F. Manea, V. Mitrana, and T. Yokomori. Two complementary operations inspired by the dna hairpin formation: Completion and reduction. *Theor. Comput. Sci.*, 410(4-5):417–425, 2009.

[8] V. Mitrana, F. Manea, and C. Martín-Vide. On some algorithmic problems regarding the hairpin completion. *Electronic Notes in Discrete Mathematics*, 27:71–72, 2006.