

Generalizing the Classic Greedy and Necklace Constructions of de Bruijn Sequences and Universal Cycles

Joe Sawada*

Aaron Williams†

Dennis Wong‡

Submitted: Aug 26, 2015; Accepted: Jan 23, 2016; Published: Feb 5, 2016

Mathematics Subject Classifications: 05A05, 05A15

Abstract

We present a class of languages that have an interesting property: For each language \mathbf{L} in the class, both the classic greedy algorithm and the classic Lyndon word (or necklace) concatenation algorithm provide the lexicographically smallest universal cycle for \mathbf{L} . The languages consist of length n strings over $\{1, 2, \dots, k\}$ that are closed under rotation with their subset of necklaces also being closed under replacing any suffix of length i by i copies of k . Examples include all strings (in which case universal cycles are commonly known as de Bruijn sequences), strings that sum to at least s , strings with at most d cyclic descents for a fixed $d > 0$, strings with at most d cyclic decrements for a fixed $d > 0$, and strings avoiding a given period. Our class is also closed under both union and intersection, and our results generalize results of several previous papers.

1 Introduction

1.1 Constructing de Bruijn Sequences

Let $\mathbf{T}(n, k)$ be the set of k -ary strings of length n . For example, $\mathbf{T}(2, 3) = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$. A *de Bruijn sequence* for $\mathbf{T}(n, k)$ is a sequence of length k^n that contains each string in $\mathbf{T}(n, k)$ exactly once as a substring when the sequence is viewed circularly.

Martin showed that a de Bruijn sequence for $\mathbf{T}(n, k)$ can be constructed by a simple greedy algorithm in 1934 [23]. The algorithm starts with sequence k^{n-1} (where exponentiation denotes repetition) and then repeatedly applies the following rule:

Append the smallest symbol in $\{1, 2, \dots, k\}$ so that substrings of length n in the resulting linear sequence are distinct.

As an example, let us illustrate one step of the algorithm when $n = 2$ and $k = 3$. After applying the rule a handful of times, the partial sequence is 31121_. At this point, the algorithm does not append 1 since 11 already appears in the sequence. Similarly, 2 is not appended since 12 already appears. However, 3 can be appended and so the algorithm continues with the sequence 311213_. Martin proved that the algorithm always terminates with a sequence that has length $k^n + n - 1$ and suffix k^n , and that a de Bruijn sequence is obtained by removing the initial k^{n-1} prefix. In particular, when $n = 2$ and $k = 3$, the algorithm terminates with 3112132233 and so 112132233 is a de Bruijn sequence for $\mathbf{T}(2, 3)$. We mention that the choice of the initial sequence is critical to the success of Martin's algorithm. For example, the greedy algorithm for $n = 2$ and $k = 3$ will get stuck after generating 112131 if the initial sequence is the

*School of Computer Science, University of Guelph, Canada. Research supported by NSERC. email: jsawada@uoguelph.ca

†Division of Science, Mathematics, and Computing, Bard College at Simon's Rock, USA. email: haron@uvic.ca

‡Department of Mathematics, Computer Science, and Information Systems, Northwest Missouri State University, USA. email: cwong@uoguelph.ca

empty string ϵ or the single symbol 1. Knuth refers to Martin's sequence as the *grand-daddy de Bruijn sequence* [19].

The general term for these sequences is named after de Bruijn who showed that there are $2^{2^{n-1}-n}$ such sequences for $\mathbf{T}(n, 2)$ in 1946 [6], and $(k!)^{k^{n-1}}k^{-n}$ for $\mathbf{T}(n, k)$ in joint work with Aardenne-Ehrenfest in 1951 [35]. Later detective work by Stanley revealed that Flye Sainte-Marie proved the same formula for $k = 2$ in 1894 [7, 29].

Martin's greedy algorithm is easy to implement, but it is impractical when n is large since it requires $\Omega(k^n)$ -space. Fredricksen and Kessler [10, 11, 12] (for $k = 2$) and later Fredricksen and Maiorana [13] (for $k \geq 2$) provided a beautiful alternative that generates each character in $O(1)$ -amortized time and uses only $O(n)$ space. Their construction is known as the *FKM algorithm* and can be summarized as follows:

Concatenate the aperiodic prefixes of the necklaces in $\mathbf{T}(n, k)$ in lexicographic order.

A *necklace* is the lexicographically smallest string in an equivalence class of strings under rotation, and its *aperiodic prefix* is the shortest prefix that can be repeated to create the string. For example, the necklaces in $\mathbf{T}(2, 3)$ are 11, 12, 13, 23, 33 and thus the FKM algorithm creates the following sequence

$$1 \cdot 12 \cdot 13 \cdot 2 \cdot 23 \cdot 3 = 112132233,$$

where \cdot denotes concatenation. Interestingly, the FKM algorithm always generates the same sequence as Martin's algorithm and this sequence is the lexicographically smallest de Bruijn cycle for $\mathbf{T}(n, k)$ (see Knuth's discussion in [19]). The efficiency of the FKM algorithm was fully analyzed by Ruskey, Savage, and Wang [25].

We note that the FKM algorithm can also be described in a slightly different, but equivalent manner. The *period* of a string is the length of its aperiodic prefix. A string is *aperiodic* if its period equals its length. An aperiodic necklace is a *Lyndon word*. The FKM algorithm can be defined as the lexicographic concatenation of the k -ary Lyndon words whose length divides n . We will use the necklace definition because it is better suited for generalizations, as pointed out by Ruskey, Sawada, Williams [27].

1.2 Constructing Universal Cycles

Given a set of strings $\mathbf{S} \subseteq \mathbf{T}(n, k)$, a *universal cycle* for \mathbf{S} is a sequence of length $|\mathbf{S}|$ that contains each string in \mathbf{S} exactly once as a substring when the sequence is viewed circularly. For example, the subset $\mathbf{S}_1 \subseteq \mathbf{T}(4, 3)$ of strings that have sum at least 10 is

$$\mathbf{S}_1 = \{1333, 2233, 2323, 2332, 2333, 3133, 3223, 3232, 3233, 3313, 3322, 3323, 3331, 3332, 3333\}$$

and the lexicographically smallest universal cycle for \mathbf{S}_1 is

$$133322332323333.$$

This natural generalization of de Bruijn sequences for $\mathbf{T}(n, k)$ to subsets of $\mathbf{T}(n, k)$ was introduced by Chung, Diaconis, and Graham in 1992 [5]. Since that time universal cycles have been proven to exist for subsets of $\mathbf{T}(n, k)$ that represent a variety of combinatorial objects including permutations, partitions, subsets, multisets, labeled graphs, various functions, and more [3, 4, 5, 15, 16, 17, 18, 19, 21, 28, 26, 32].

Given a specific set of strings $\mathbf{S} \subseteq \mathbf{T}(n, k)$, the problem of finding an explicit construction (and generation algorithm) for a universal cycle for \mathbf{S} is generally a more difficult problem than proving that one exists. The approach considered in this article is to generalize the greedy algorithm and the FKM algorithm. Previous results using this approach include the following:

- Moreno proved that a generalized FKM algorithm creates universal cycles for the set of rotations of the lexicographically largest i necklaces [24].
- Au proved that generalized FKM and greedy algorithm create universal cycles for the aperiodic strings in $\mathbf{T}(n, k)$ [2].

- The authors proved that generalized FKM and greedy algorithm create universal cycles for the binary strings with sum at least s [31].

We will see that all three of these results can be explained using the same generalization of the FKM algorithm and the same generalization of the greedy algorithm. To illustrate how these unified generalizations work, let us reconsider the strings \mathbf{S}_1 in $\mathbf{T}(4, 3)$ with sum at most 10. Let the necklaces of a set \mathbf{S} be denoted by $\mathbf{N}(\mathbf{S})$. Then

$$\mathbf{N}(\mathbf{S}_1) = \{1333, 2233, 2323, 2333, 3333\},$$

and the concatenation of their aperiodic prefixes in lexicographic order gives

$$1333 \cdot 2233 \cdot 23 \cdot 2333 \cdot 3.$$

This is identical to the lexicographically smallest universal cycle for \mathbf{S}_1 that we saw earlier. More broadly, the generalized version of the FKM algorithm is stated below:

Concatenate the aperiodic prefixes of the necklaces in \mathbf{S} in lexicographic order.

Notice that the previous FKM results use $\mathbf{S} \subseteq \mathbf{T}(n, k)$ that are *closed under rotation*, meaning $\alpha\beta \in \mathbf{S}$ implies $\beta\alpha \in \mathbf{S}$. To understand this fact, let $\text{Rot}(\mathbf{S})$ denote closure of \mathbf{S} under rotation. Note that the number of distinct rotations of an individual string is equal to its period. For example, $\text{Rot}(\{2233\}) = \{2233, 2332, 3322, 3223\}$ and $\text{Rot}(\{2323\}) = \{2323, 3232\}$. In other words, the length of the aperiodic prefix of $\alpha \in \mathbf{T}(n, k)$ is equal to $|\text{Rot}(\alpha)|$. Thus, the FKM algorithm creates sequences of the correct length for each \mathbf{S} that is closed under rotation.

Generalizing the greedy algorithm is a bit more subtle due to the choice of the initial sequence. In this article we will be focused on $\mathbf{S} \subseteq \mathbf{T}(n, k)$ that are guaranteed to have $k^n \in \mathbf{S}$ (or $xk^{n-1} \in \mathbf{S}$ for some $x < k$). Thus, k^{n-1} remains a reasonable choice for the initial sequence. Martin's algorithm is then generalized as follows

Append the smallest symbol in $\{1, 2, \dots, k\}$ so that substrings of length n in the resulting linear sequence are distinct and in \mathbf{S} .

If we apply this greedy algorithm to \mathbf{S}_1 , then it terminates with sequence

$$333133322332323333$$

and removing initial 333 results in the same universal cycle as generated by the generalized FMK approach.

Universal cycles have also been constructed using alternate approaches. For example, Ruskey and Williams [28] and later Holroyd, Ruskey, and Williams [15] provided an efficient algorithm to construct universal cycles for $\mathbf{T}(n-1, n)$. The authors also constructed universal cycles for binary strings of length n whose sum falls within a given range [30] which extends [27] and Stevens and Williams [34].

1.3 New Results and Their Significance

We prove that the generalized FKM and greedy algorithms generate a wide variety of natural universal cycles. A list of examples appear in Sections 3.3.1–3.3.11 and it includes the aforementioned results by Moreno [24], Au [2], and the authors [31]. Furthermore, each of our examples follows from a single general result.

Theorem 1. *The greedy and FKM algorithms create the lexicographically smallest universal cycle for any $\mathbf{S} \subseteq \mathbf{T}(n, k)$ that satisfies the following closure properties:*

- (C1) The set of strings \mathbf{S} is closed under rotation.
- (C2) Its subset of necklaces is closed under replacing any suffix of length i by k^i .

We note that the first closure property is not sufficient for guaranteeing the existence of universal cycles. For example, $\{11, 22\}$ is closed under rotation and has no universal cycle. Our second closure property is also insufficient for proving the existence. For example, $\{11, 12, 22\} \subseteq \mathbf{T}(2, 2)$ does not have a universal cycle.

To underscore the significance of our results, we note that the greedy algorithm, the FKM algorithm, and the lexicographically smallest universal cycle do not always operate in such harmony. For example, the subset $\mathbf{S}_2 \subset \mathbf{T}(3, 3)$ that does not have 13 as a cyclic substring is

$$\mathbf{S}_2 = \{111, 112, 121, 122, 123, 211, 212, 221, 222, 223, 231, 232, 233, 312, 322, 323, 332, 333\}.$$

In other words, \mathbf{S}_2 does not have strings in the form $13x$, $x13$, or $3x1$. Clearly \mathbf{S}_2 is closed under rotation, so it satisfies our first condition. The necklaces in \mathbf{S}_2 are

$$\mathbf{N}(\mathbf{S}_2) = \{111, 112, 122, 123, 222, 223, 233, 333\}.$$

Notice that $\mathbf{N}(\mathbf{S}_2)$ does not satisfy our second closure property since 111 is included but 113 is not. Thus, our results do not guarantee that the FKM algorithm will create a universal cycle for \mathbf{S}_2 . In fact, the FKM algorithm creates

$$1 \cdot 112 \cdot 122 \cdot 123 \cdot 2 \cdot 223 \cdot 233 \cdot 3.$$

This is not a universal cycle for \mathbf{S}_2 since 211 and 312 do not appear, while 212 appears twice and 311 appears when it should not. On the other hand, the greedy algorithm applied to \mathbf{S}_2 creates

$$221112122231232233,$$

which is a universal cycle for \mathbf{S}_2 . However, it is not the lexicographically smallest universal cycle for \mathbf{S}_2 (when viewed linearly).

The previous example showed that the harmony produced the greedy algorithm, the FKM algorithm, and the lexicographically smallest universal cycle does not hold for the strings in $\mathbf{T}(3, 3)$ that avoid 13 as a cyclic substring. Notice that in this case the avoided substring 13 contains the symbol $k = 3$. We will see in Section 3.3.6 that the three interrelated results do hold when the avoided substring does not contain the symbol k . This illustrates one way in which our result is ‘tight’.

One motivation for constructing universal cycles is so that they can be used in applications. Without an explicit construction it can be computationally infeasible to create these sequences when n is large. Sample applications of de Bruijn sequences include dynamic connections in overlay networks [9], genomics [1], software calculation of the ruler function in computer words [19], and indexing a 1 in a computer word [20]. Our two-pronged results have an additional benefit: Those in need of one of our universal cycles can start by implementing the simple greedy algorithm; if this proves to be infeasible due to memory constraints, then they can implement the FKM algorithm. De Bruijn sequences are also used as education tools (see Graham, Knuth, and Patashnik’s *Concrete Mathematics* [14]) and are involved in many interesting academic papers (for recent examples, see Levine [22] and Ehrenborg, Kitaev, and Steingrímsson [8]).

To conclude our introduction we mention that our second closure property can be slightly relaxed to allow $k^n \notin \mathbf{S}$. For example, the constructions for \mathbf{S}_1 still work when k^n is omitted from \mathbf{S}_1 and the final k is omitted from the universal cycles generated. We use this minor generalization when considering periodic and aperiodic strings.

Corollary 1. *Theorem 1 holds if $\mathbf{S} \cup \{k^n\} \subseteq \mathbf{T}(n, k)$ satisfies (C1) and (C2).*

The main results of this paper are also found in Wong’s PhD thesis [36].

1.4 Outline

The remainder of this paper is organized as follows. Section 2 investigates our second closure property under the name k -suffix languages. Section 3 proves that the universal cycles discussed in Theorem 1 exist, and gives a list of specific universal cycles that are constructed by it. Section 4 proves our result on the FKM algorithm. Section 5 proves our result on the greedy algorithm and proves that our universal cycles are lexicographically smallest. Section 6 concludes the paper with open problems and remarks.

2 k -suffix languages and the k -suffix poset

In this section we investigate our second closure property from Theorem 1 in more detail. First we define the k -suffix property and prove that set of all necklaces satisfies it. Then we formulate k -suffix languages as partially ordered sets, which provides helpful visualizations and closure properties at the end of the section.

Definition 1. A k -suffix language \mathbf{S} is a subset of $\mathbf{T}(n, k)$ that satisfies the following closure property:

$$\text{If } a_1a_2 \cdots a_n \in \mathbf{S} \text{ then } a_1a_2 \cdots a_{n-i}k^i \in \mathbf{S} \text{ for all } 1 \leq i \leq n.$$

As an example, let $\mathbf{N}(n, k)$ denote the set of necklaces in $\mathbf{T}(n, k)$. It is straightforward to observe that necklaces are a k -suffix language.

Lemma 1. The set of necklaces $\mathbf{N}(n, k)$ is a k -suffix language.

Proof. Let $a_1a_2 \cdots a_n = a_1a_2 \cdots a_{n-j}k^j \in \mathbf{N}(n, k)$ where $j \geq 0$. By the definition of a necklace it is easy to see that $a_1a_2 \cdots a_{n-j-1}kk^j \in \mathbf{N}(n, k)$. Thus it follows that $a_1a_2 \cdots a_{n-i}k^i \in \mathbf{N}(n, k)$ for all $1 \leq i \leq n$. \square

Note that this definition of a k -suffix language implies closure under replacing any suffix of length i by k^i . Equivalently, k -suffix languages can be defined as closed under replacing the rightmost value that is less than k by k . We next formulate this idea using a partially ordered set over $\mathbf{T}(n, k)$. Consider the following definition,

$$\tau_k(\alpha) = \begin{cases} k^n & \text{if } \alpha = k^n; \\ \beta kk^j & \text{if } \alpha = \beta x k^j \text{ for some } x < k. \end{cases}$$

In other words, k^n is *terminal*, and otherwise $\tau(\alpha)$ is obtained by replacing the rightmost value in α that is less than k with k . When the context is clear we use τ instead of τ_k . For example, when $k = 3$,

$$\tau(1122) = 1123 \text{ and } \tau(2313) = 2333 \text{ and } \tau(3333) = 3333.$$

This provides our cover relation, and our partial order \prec_k is the non-reflexive transitive closure of τ_k . Again we use \prec instead of \prec_k when the context allows. For example, when $k = 3$,

$$1123 \prec 1122 \text{ and } 1333 \prec 1122 \text{ whereas } 2313 \not\prec 2113 \text{ and } 2113 \not\prec 2113.$$

Our partially ordered set is formally defined below.

Definition 2. The k -suffix poset $\mathbf{Poset}(n, k)$ has ground set $\mathbf{T}(n, k)$ and partial order \prec_k .

Figure 1 illustrates the Hasse diagram of $\mathbf{Poset}(4, 2)$ and $\mathbf{Poset}(2, 3)$, in which each cover relation $\beta = \tau(\alpha)$ is shown as an edge from α down to β . It is easy to see that $\mathbf{Poset}(n, k)$ is a tree poset with unique minimum element k^n .

An *ideal* (also known as a *lower set*) of $\mathbf{Poset}(n, k)$ is a subset $\mathbf{I} \subseteq \mathbf{Poset}(n, k)$ such that $x \in \mathbf{I}$ and $y \prec x$ implies $y \in \mathbf{I}$. Figure 1 illustrates an ideal of necklaces in both of its posets. The following theorem proves that every k -suffix language can be visualized in this way.

Theorem 2. A set $\mathbf{S} \subseteq \mathbf{T}(n, k)$ is a k -suffix language if and only if \mathbf{S} is an ideal of $\mathbf{Poset}(n, k)$.

Proof. Suppose \mathbf{S} is a k -suffix language. If $s \in \mathbf{S}$ and $s \neq k^n$, then $s = \alpha x k^j$ for some $x < k$. By the definition of k -suffix language, $\alpha k k^j \in \mathbf{S}$. Therefore, $\tau_k(s) \in \mathbf{S}$, and thus \mathbf{S} is an ideal of $\mathbf{Poset}(n, k)$. The other direction is similar. \square

Ideals of a given poset are closed under union and intersection, so Theorem 2 immediately gives the following corollary.

Corollary 2. If $\mathbf{S}_A, \mathbf{S}_B \subseteq \mathbf{T}(n, k)$ are k -suffix languages, then $\mathbf{S}_A \cup \mathbf{S}_B$ and $\mathbf{S}_A \cap \mathbf{S}_B$ are also k -suffix languages.

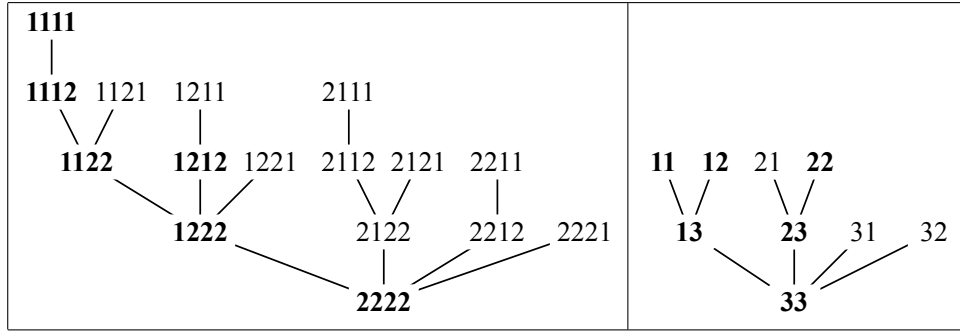


Figure 1: The Hasse diagram of $\text{Poset}(4, 2)$ and $\text{Poset}(2, 3)$ respectively. In both cases the necklaces form the ideal in bold.

3 Existence and Examples

In this section we ground our construction results with a simple existence proof. We also give a list of interesting universal cycles constructed by our results.

3.1 Class of Strings

The class of strings covered by Theorem 1 is defined below. Recall that $\mathbf{N}(\mathbf{S})$ denotes the necklaces in $\mathbf{S} \subseteq \mathbf{T}(n, k)$. That is, $\mathbf{N}(\mathbf{S}) = \mathbf{S} \cap \mathbf{N}(n, k)$.

Definition 3. Let $\mathcal{C}(n, k)$ be the set that contains precisely all $\mathbf{S} \subseteq \mathbf{T}(n, k)$ that are closed under rotation, and whose necklaces $\mathbf{N}(\mathbf{S})$ form a k -suffix language.

To prove that $\mathbf{S} \in \mathcal{C}(n, k)$ we need to prove that \mathbf{S} is closed under rotation and its subset of necklaces $\mathbf{N}(\mathbf{S})$ is a k -suffix language. Sometimes it is more convenient to prove that the entire set \mathbf{S} is a k -suffix language rather than its necklace subset $\mathbf{N}(\mathbf{S})$. The following lemma proves that this approach is also sufficient.

Lemma 2. Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$. If \mathbf{S} is closed under rotation and \mathbf{S} is a k -suffix language, then $\mathbf{S} \in \mathcal{C}(n, k)$.

Proof. Recall that $\mathbf{N}(n, k)$ is a k -suffix language by Lemma 1. Therefore, if \mathbf{S} is a k -suffix language, then so is $\mathbf{N}(\mathbf{S}) = \mathbf{S} \cap \mathbf{N}(n, k)$. Therefore, if \mathbf{S} is also closed under rotation, then \mathbf{S} is a k -suffix language by Definition 3. \square

We note that Lemma 2 does not cover all sets in $\mathcal{C}(n, k)$. For example,

$$\mathbf{S}_3 = \{1111, 1112, 1121, 1211, 2111, 1122, 1221, 2211, 2112, 1222, 2122, 2212, 2221, 2222\},$$

is in $\mathcal{C}(4, 2)$. However, since $1211 \in \mathbf{S}_3$ and $1212 \notin \mathbf{S}_3$, the set \mathbf{S}_3 is not a k -suffix language by itself.

3.2 Existence

Stevens and Williams [33] characterized the existence of universal cycles for subsets of $\mathbf{T}(n, k)$ that are closed under rotation. We now use one of their results to prove that every set in $\mathcal{C}(n, k)$ has a universal cycle. First we recount a definition from [33]. A set $\mathbf{S} \subseteq \mathbf{T}(n, k)$ is *increasable* if the following is true:

If $\alpha \in \mathbf{S}$ and α is not equal to k^n , then there exists a symbol in α that can be increased and the resulting string is also in \mathbf{S} .

More precisely, \mathbf{S} is increasable if $a_1 a_2 \cdots a_n \in \mathbf{S}$ and $a_1 a_2 \cdots a_n \neq k^n$, then there exist i and $b > a_i$ such that $a_1 a_2 \cdots a_{i-1} b a_{i+1} a_{i+2} \cdots a_n \in \mathbf{S}$.

Theorem 3 ([33]). *If $\mathbf{S} \subseteq \mathbf{T}(n, k)$ is increasable and closed under rotation, then \mathbf{S} has a universal cycle.*

Corollary 3. *Every $\mathbf{S} \in \mathcal{C}(n, k)$ has a universal cycle.*

Proof. By Definition 3, each $\mathbf{S} \in \mathcal{C}(n, k)$ is closed under rotation. Consider an arbitrary $\alpha \in \mathbf{S}$ with $\alpha \neq k^n$, and let $\beta \in \mathbf{S}$ be its smallest rotation in lexicographic order. By Definition 3, there is at least one symbol in β that is less than k that can be increased to k so that the resulting string is also in \mathbf{S} . Since \mathbf{S} is closed under rotation, the same statement is true for α . Thus, \mathbf{S} is increasable. Hence, \mathbf{S} has a universal cycle by Theorem 3. \square

3.3 Examples

In Sections 3.3.1–3.3.11 we describe interesting members of $\mathcal{C}(n, k)$, including the sets of strings considered by Moreno [24], Au [2], and the authors [31]. Figure 2 illustrates the lexicographically smallest universal cycles for each example over $\mathbf{T}(4, 4)$. When considering periodic and aperiodic strings we prove that $\mathbf{S} \cup \{k^n\} \in \mathcal{C}(n, k)$; our constructive results from Theorem 1 will still apply in these cases by Corollary 1.

3.3.1 Minimum Sum

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings with sum at least s . This set is closed under rotation since rotation does not change a string's sum. Also, \mathbf{S} is a k -suffix language by Theorem 2 since replacing any symbol $x < k$ with k increases the sum of the string. Therefore, $\mathbf{S} \in \mathcal{C}(n, k)$ by Lemma 2. The authors previously considered sets of this type when $n = 2$ [31].

3.3.2 At most $d > 0$ cyclic descents

A *descent* in a string $a_1 a_2 \cdots a_n$ is a pair of consecutive elements a_i, a_{i+1} such that $a_i > a_{i+1}$. A *cyclic descent* is a descent or the pair a_n, a_1 where $a_n > a_1$. Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings with at most d cyclic descents for some fixed $d > 0$. This set is clearly closed under rotation. Also, $\mathbf{N}(\mathbf{S})$ is a k -suffix language since replacing any suffix of length i by k^i will increase the number of descents only in the special case where a necklace is of the form x^n for $x < k$. For these strings the number of descents increases from 0 to 1, which does not violate the upperbound since $d > 0$. Therefore, $\mathbf{S} \in \mathcal{C}(n, k)$.

3.3.3 At most $d > 0$ cyclic decrements

A *decrement* in a string $a_1 a_2 \cdots a_n$ is a pair of consecutive elements a_i, a_{i+1} such that $a_i = a_{i+1} + 1$. A *cyclic decrement* is a decrement or the pair a_n, a_1 where $a_n = a_1 + 1$. Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings with at most d cyclic decrements for some fixed $d > 0$. This set is clearly closed under rotation. Also, $\mathbf{N}(\mathbf{S})$ is a k -suffix language since replacing any suffix of length i by k^i will only increase the number of decrements in the special case of the necklace $(k-1)^n$. For this string the number of decrements increases from 0 to 1, which does not violate the upperbound since $d > 0$. Therefore, $\mathbf{S} \in \mathcal{C}(n, k)$.

3.3.4 Frequency of k

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings with at least ℓ_k copies of k . This set is closed under rotation since rotation does not change a string's symbol frequencies. Also, \mathbf{S} is a k -suffix language by Theorem 2 since replacing any symbol $x < k$ with k increases the frequency of k in the string. Therefore, $\mathbf{S} \in \mathcal{C}(n, k)$ by Lemma 2.

3.3.5 Frequency of $i < k$

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings with at most u_i copies of $i < k$. This set is closed under rotation since rotation does not change a string's symbol frequencies. Also, \mathbf{S} is a k -suffix language by Theorem 2 since replacing any symbol $x < k$ with k does not increase the frequency of i in the string. Therefore, $\mathbf{S} \in \mathcal{C}(n, k)$ by Lemma 2.

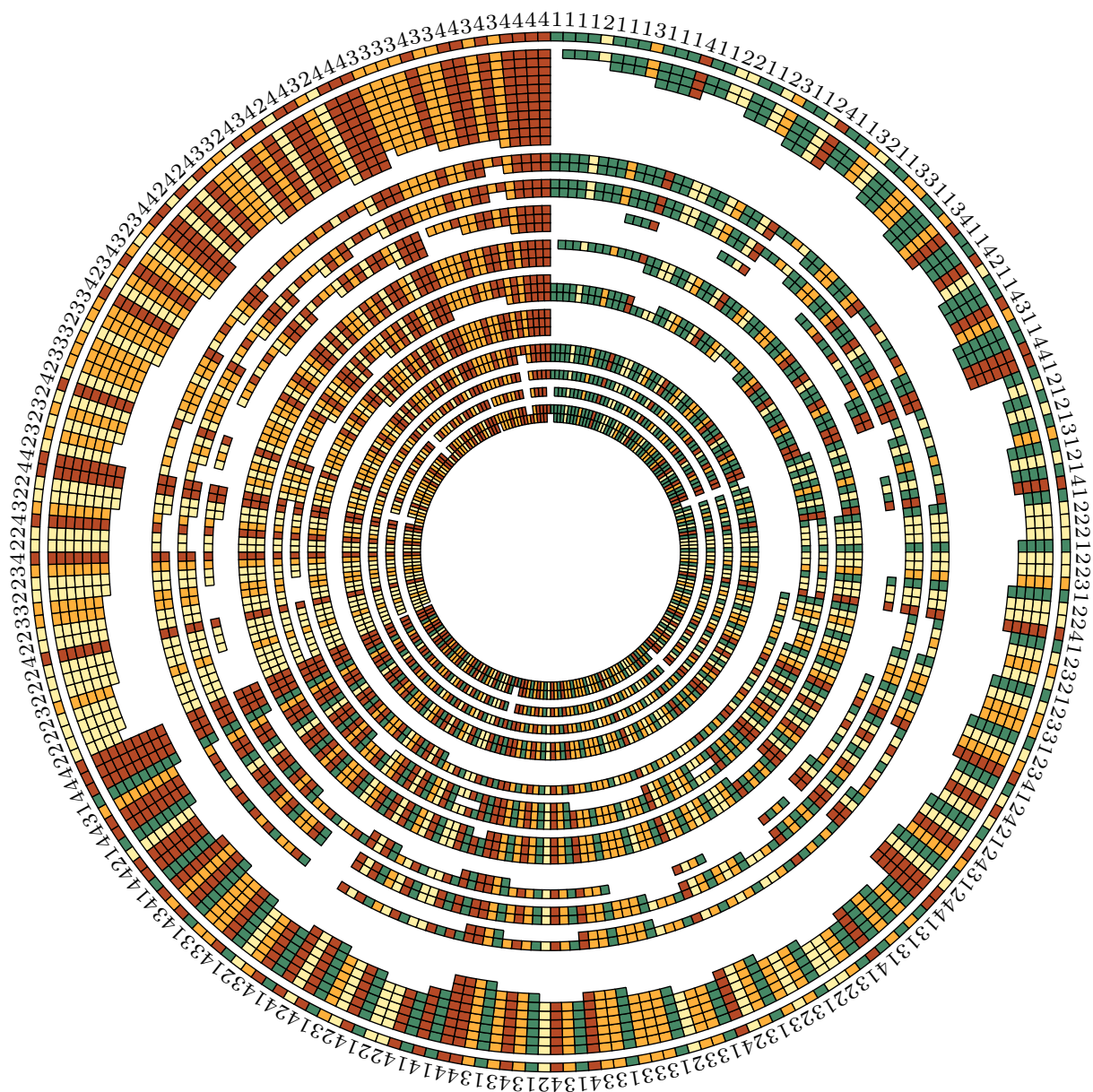


Figure 2: The lexicographically smallest universal cycles for a variety of 4-ary strings of length four. Each set \mathbf{S} is in $\mathcal{C}(4, 4)$ (or $\mathbf{S} \cup \{4444\} \in \mathcal{C}(4, 4)$) and hence can be generated by the generalized FKM and greedy algorithms. The outermost ring is a universal cycle for all strings $\mathbf{T}(n, k)$. The remaining rings from outer-to-inner are for subsets that (i) have sum at least 5, 6, ..., 15; (ii) have at most 2, 1 cyclic descents; (iii) have at most 2, 1 cyclic decrements; (iv) have frequency for symbol $k = 4$ at least 1, 2, 3; (v) have frequency for symbol 1 at most 3, 2, 1; (vi) do not have 11, 22, 33 as a cyclic substring; (vii) are rotations of the largest 50, 30, 10 necklaces; (viii) are not rotations of the periodic necklace 1212, 3434; (ix) are in the intersection of the previous two universal cycles (i.e. not rotations of 1212 or 3434); (x) are aperiodic; (xi) have period in $\{1, 4\}$, $\{1, 2\}$. Each universal cycle starts from 12 o'clock and proceeds in clockwise order, with a gap between the examples of each type. For example, the fourth ring from the outside gives the universal cycle 111411231124 ... 34444 which contains all strings in $\mathbf{T}(4, 4)$ with sum at least 7.

3.3.6 Avoiding a Substring

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings that do not contain $\beta \in \mathbf{T}(m, k-1)$, for some $m \geq 1$, as a cyclic substring. This set is closed under rotation since rotation does not change a string's cyclic substrings. Also, \mathbf{S} is a k -suffix language by Theorem 2 since replacing any symbol $x < k$ with k cannot create a new cyclic substring in $\mathbf{T}(m, k-1)$. Therefore, $\mathbf{S} \in \mathcal{C}(n, k)$ by Lemma 2.

3.3.7 Rotations of Large Necklaces

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings that are rotations of the largest i necklaces in $\mathbf{N}(n, k)$ in lexicographic order. This set is closed under rotation by its definition. If $\alpha \in \mathbf{N}(\mathbf{S})$, then replacing its rightmost symbol $x < k$ with k will create another necklace by Lemma 1, and this necklace is larger in lexicographic order. Therefore, $\mathbf{N}(\mathbf{S})$ is a k -suffix language. Hence, $\mathbf{S} \in \mathcal{C}(n, k)$. Moreno previously considered sets of this type [24].

3.3.8 Avoiding the Rotations of a Periodic Necklace

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the strings that are not rotations of a fixed periodic necklace $\beta \in \mathbf{N}(n, k)$. That is, $\mathbf{S} = \mathbf{T}(n, k) - \text{Rot}(\{\beta\})$. This set is closed under rotation by its definition. If $\beta = k^n$, then $\mathbf{S} \cup \{k^n\} = \mathbf{T}(n, k) \in \mathcal{C}(n, k)$. Otherwise, consider $\beta \neq k^n$. Let $\alpha x k^j \in \mathbf{N}(\mathbf{S})$ with $x < k$ be an arbitrary necklace in \mathbf{S} that is not equal to β (where $j = 0$ is possible). We know that $\alpha k k^j$ is a necklace by Lemma 1. Suppose that $\alpha k k^j = \beta$. Since β is periodic and $\beta \neq k^n$, we can write β as $\beta = (\gamma k^{j+1})^t$ where $\gamma \neq \epsilon$ and $t > 1$. Observe that $\alpha x k^j = (\gamma k^{j+1})^{t-1} \gamma x k^j$ is not a necklace since $x < k$, which contradicts $\alpha x k^j$ being a necklace. Thus $\alpha k k^j \neq \beta$ and $\mathbf{N}(\mathbf{S})$ is a k -suffix language. Therefore $\mathbf{S} \in \mathcal{C}(n, k)$.

3.3.9 Unions and Intersections

If \mathbf{S}_A and \mathbf{S}_B are closed under rotation, then both $\mathbf{S}_A \cup \mathbf{S}_B$ and $\mathbf{S}_A \cap \mathbf{S}_B$ are closed under rotation. Thus, Corollary 2 gives the following.

Lemma 3. *If $\mathbf{S}_A, \mathbf{S}_B \in \mathcal{C}(n, k)$, then $\mathbf{S}_A \cup \mathbf{S}_B \in \mathcal{C}(n, k)$ and $\mathbf{S}_A \cap \mathbf{S}_B \in \mathcal{C}(n, k)$.*

Lemma 3 allows us to combine the previous results in interesting ways, as illustrated in Sections 3.3.10 and 3.3.11.

3.3.10 Aperiodic Strings

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ contain the aperiodic strings. Let the periodic necklaces in $\mathbf{N}(n, k)$ be $\beta_1, \beta_2, \dots, \beta_t$. Notice that

$$\begin{aligned} \mathbf{S} &= \mathbf{T}(n, k) - \text{Rot}(\{\beta_1, \beta_2, \dots, \beta_t\}) \\ &= (\mathbf{T}(n, k) - \text{Rot}(\{\beta_1\})) \cap (\mathbf{T}(n, k) - \text{Rot}(\{\beta_2\})) \cap \dots \cap (\mathbf{T}(n, k) - \text{Rot}(\{\beta_t\})). \end{aligned}$$

Therefore, $\mathbf{S} \cup \{k^n\} \in \mathcal{C}(n, k)$ by Section 3.3.8 and Lemma 3. Au previously considered sets of this type [2].

3.3.11 Strings with Given Periods

Let $\mathbf{S} \subseteq \mathbf{T}(n, k)$ be the strings with period in $P \subseteq \{1, 2, \dots, n\}$ with $n \in P$. We can prove that $\mathbf{S} \cup \{k^n\} \in \mathcal{C}(n, k)$ by using the same approach as in Section 3.3.10. More specifically, replace $\beta_1, \beta_2, \dots, \beta_t$ by the necklaces whose period is not in P . This provides a nice generalization of Au's result [2].

4 The generalized FKM construction

In this section, we prove the FKM portion of Theorem 1 and Corollary 1.

Let $\mathbf{S} \in \mathcal{C}(n, k)$ where $|\mathbf{S}| > 1$ and let $\alpha_1, \alpha_2, \dots, \alpha_m$ be the lexicographic ordering of necklaces in $\mathbf{N}(\mathbf{S})$. Let $\text{ap}(\alpha)$ be the aperiodic prefix of α . Let $\text{FKM}(\mathbf{S})$ be the sequence created by the generalized FKM algorithm applied to \mathbf{S} . That is,

$$\text{FKM}(\mathbf{S}) = \text{ap}(\alpha_1) \cdot \text{ap}(\alpha_2) \cdot \dots \cdot \text{ap}(\alpha_m)$$

where \cdot denotes concatenation. We first prove the following results:

1. $m > 1$ and $k^n \in \mathbf{N}(\mathbf{S})$,
2. there are no consecutive periodic necklaces in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$,
3. if $\alpha_i = a_1 a_2 \dots a_{n-j-1} x k^j$ for some $x < k$ and $1 \leq i < m$, then α_{i+1} has prefix $a_1 a_2 \dots a_{n-j-1}$,
4. α_1 is a prefix of $\text{FKM}(\mathbf{S})$,
5. $x k^n$ is a suffix of $\text{FKM}(\mathbf{S})$ where x is the maximum value less than k such that $x k^{n-1} \in \mathbf{N}(\mathbf{S})$, and
6. if $\alpha_i = a_1 a_2 \dots a_n = a_1 a_2 \dots a_{n-j-1} x k^j$ for some $x < k$ and $1 \leq i < m$, then $\text{FKM}(\mathbf{S})$ contains the substring $a_1 a_2 \dots a_n \cdot a_1 a_2 \dots a_{|\text{ap}(\alpha)|-j-1}$.

Lemma 4. *If $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$, then $|\mathbf{N}(\mathbf{S})| > 1$ and $k^n \in \mathbf{N}(\mathbf{S})$.*

Proof. Since $|\mathbf{S}| > 1$, there exists a string α in \mathbf{S} such that $\alpha \neq k^n$. Since \mathbf{S} is closed under rotation, it also contains a necklace β such that $\beta \neq k^n$ and $\beta \in \text{Rot}(\alpha)$. Also, since $\mathbf{N}(\mathbf{S})$ is a k -suffix language, k^n must also be in $\mathbf{N}(\mathbf{S})$. Thus $|\mathbf{N}(\mathbf{S})| > 1$ and $k^n \in \mathbf{N}(\mathbf{S})$. \square

Lemma 5. *If $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$, then there are no consecutive periodic necklaces in the lexicographic order of $\mathbf{N}(\mathbf{S})$.*

Proof. Let α and β be consecutive necklaces in the lexicographic order of $\mathbf{N}(\mathbf{S})$ with $\alpha < \beta$. Since $\alpha < \beta$ we have $\alpha \neq k^n$, and so α contains a rightmost symbol x that is less than k . That is, $\alpha = a_1 a_2 \dots a_{n-j-1} x k^j$ where $x < k$ and $j \geq 0$.

Let y be the smallest value such that $a_1 a_2 \dots a_{n-j-1} y k^j \in \mathbf{N}(\mathbf{S})$ and $x < y \leq k$; the value y exists because $\mathbf{N}(\mathbf{S})$ is a k -suffix language. Notice that β must have $a_1 a_2 \dots a_{n-j-1} y$ as a prefix.

Now suppose that α is periodic. Therefore, $\alpha = (\gamma x k^j)^t$ for some γ and $t > 1$. Therefore, β has $(\gamma x k^j)^{t-1} \gamma y$ as a prefix. However, $(\gamma x k^j)^{t-1} \gamma y$ is not the prefix of any periodic necklace of length n . Therefore, β is not periodic. \square

Lemma 6. *Suppose $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$. Let α and β denote consecutive necklaces in the lexicographical ordering of $\mathbf{N}(\mathbf{S})$ such that $\alpha < \beta$. If $\alpha = a_1 a_2 \dots a_{n-j-1} x k^j$ for some $x < k$, then β has prefix $a_1 a_2 \dots a_{n-j-1}$.*

Proof. Since α is a necklace, clearly $a_1 a_2 \dots a_{n-j-1} y k^j$ is a necklace for all $x < y \leq k$. There exists some smallest value of y such that $a_1 a_2 \dots a_{n-j-1} y k^j \in \mathbf{N}(\mathbf{S})$ since $\mathbf{N}(\mathbf{S})$ is a k -suffix language. Since β is lexicographically smaller or equal to $a_1 a_2 \dots a_{n-j-1} y k^j$ but lexicographically larger than α , clearly β has prefix $a_1 a_2 \dots a_{n-j-1}$. \square

Lemma 7. *If $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$, then the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$ is a prefix of $\text{FKM}(\mathbf{S})$.*

Proof. Let $\alpha = a_1 a_2 \dots a_{n-j-1} x k^j$ be the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$ for some $x < k$. Such a value of x exists since $\alpha \neq k^n$ by Lemma 4. If α is aperiodic, then clearly $\text{FKM}(\mathbf{S})$ has prefix α . Otherwise if α is periodic, then $\alpha = (\text{ap}(\alpha))^t$ for some $t > 1$. Let β be the necklace that is after α in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$. Such a necklace exists since $|\mathbf{N}(\mathbf{S})| > 1$ by Lemma 4. By Lemma 5, β is aperiodic. Also by Lemma 6, β has prefix $a_1 a_2 \dots a_{n-j-1} = (\text{ap}(\alpha))^{t-1} a_{(t-1)|\text{ap}(\alpha)|+1} a_{(t-1)|\text{ap}(\alpha)|+2} \dots a_{n-j-1}$. Therefore, $\text{ap}(\alpha) \cdot \text{ap}(\beta)$ has prefix $\text{ap}(\alpha) \cdot (\text{ap}(\alpha))^{t-1} = \alpha$. \square

Lemma 8. If $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$, then $FKM(\mathbf{S})$ has suffix xk^n , where x is the maximum value such that $x < k$ and $xk^{n-1} \in \mathbf{N}(\mathbf{S})$.

Proof. Let α and β denote the two lexicographically largest necklace in $\mathbf{N}(\mathbf{S})$ such that $\alpha < \beta$. By Lemma 4, $\beta = k^n$ and α exists since $|\mathbf{N}(\mathbf{S})| > 1$. There exists a maximum value $x < k$ such that $xk^{n-1} \in \mathbf{N}(\mathbf{S})$ since $\mathbf{N}(\mathbf{S})$ is a k -suffix language. Observe that $\alpha = xk^{n-1}$ since there can be no necklace between it and β in the lexicographical ordering of $\mathbf{N}(\mathbf{S})$. Therefore, $FKM(\mathbf{S})$ has suffix $\text{ap}(\alpha) \cdot \text{ap}(\beta) = xk^{n-1} \cdot k = xk^n$. \square

Lemma 9. Suppose $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$. If $\alpha = a_1a_2 \cdots a_n = a_1a_2 \cdots a_{n-j-1}xk^j \in \mathbf{N}(\mathbf{S})$ for some $x < k$, then $FKM(\mathbf{S})$ contains the substring $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{|\text{ap}(\alpha)|-j-1}$.

Proof. Let $p = |\text{ap}(\alpha)|$ and let β denote the necklace that is after α in the lexicographic ordering $\mathbf{N}(\mathbf{S})$. Clearly $\alpha \neq k^n$ and β exists by Lemma 4. By Lemma 6 β has prefix $a_1a_2 \cdots a_{n-j-1}$. By Lemma 5 at most one of these necklaces is periodic and we proceed in three cases.

1. If both α and β are aperiodic, then $\text{ap}(\alpha) \cdot \text{ap}(\beta) = \alpha \cdot \beta$ is a substring of $FKM(\mathbf{S})$ which has prefix $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{n-j-1} = a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{p-j-1}$.
2. If α is periodic and β is aperiodic, then $\text{ap}(\alpha) \cdot \text{ap}(\beta) = \text{ap}(\alpha) \cdot \beta$ is a substring of $FKM(\mathbf{S})$ which has prefix $\text{ap}(\alpha) \cdot a_1a_2 \cdots a_{n-j-1}$. Let $\alpha = (\text{ap}(\alpha))^t$ for some $t > 1$. Observe that

$$\begin{aligned} \text{ap}(\alpha) \cdot a_1a_2 \cdots a_{n-j-1} &= \text{ap}(\alpha) \cdot (\text{ap}(\alpha))^{t-1}a_{(t-1)p+1}a_{(t-1)p+2} \cdots a_{n-j-1} \\ &= \alpha \cdot a_{(t-1)p+1}a_{(t-1)p+2} \cdots a_{n-j-1} \\ &= a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{p-j-1}. \end{aligned}$$

3. If α is aperiodic and β is periodic, then there are two subcases. If $\beta = k^n$, the substring $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{p-j-1}$ is simply equal to $a_1a_2 \cdots a_n$ due to the fact that α has suffix k^{n-1} . The desired substring is found in the length $n+1$ suffix of $FKM(\mathbf{S})$ by Lemma 8. Otherwise if $\beta \neq k^n$, then let γ be the necklace that is after β in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$. Such a necklace exists since $\beta \neq k^n$. Notice that γ is aperiodic by Lemma 5. Therefore, by the arguments in the previous case, $\text{ap}(\beta) \cdot \text{ap}(\gamma)$ has prefix β . Therefore, $\text{ap}(\alpha) \cdot \text{ap}(\beta) \cdot \text{ap}(\gamma)$ is a substring of $FKM(\mathbf{S})$ which has prefix $\alpha \cdot \beta$. The prefix $\alpha \cdot \beta$ contains the prefix $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{p-j-1}$.

Thus $FKM(\mathbf{S})$ contains the substring $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{|\text{ap}(\alpha)|-j-1}$. \square

We now prove the FKM portion of Theorem 1.

Theorem 4. If $\mathbf{S} \in \mathcal{C}(n, k)$, then $FKM(\mathbf{S})$ is a universal cycle for \mathbf{S} .

Proof. Since \mathbf{S} is closed under rotation by the definition of $\mathcal{C}(n, k)$ and its strings all have length n , the definition of $FKM(\mathbf{S})$ implies $|FKM(\mathbf{S})| = |\mathbf{S}|$. Therefore, to prove $FKM(\mathbf{S})$ is a universal cycle for \mathbf{S} , we only need to show that $FKM(\mathbf{S})$ contains each string in \mathbf{S} as a substring when the sequence is considered circularly.

When $|\mathbf{N}(\mathbf{S})| = 1$, $\mathbf{S} = \mathbf{N}(\mathbf{S}) = \{k^n\}$. In this case, $FKM(\mathbf{S})$ is the single character k which is a universal cycle for \mathbf{S} . For the remainder of the proof we assume $|\mathbf{N}(\mathbf{S})| > 1$.

Now consider a rotation $\beta = a_i a_{i+1} \cdots a_n a_1 a_2 \cdots a_{i-1}$ of an arbitrary necklace $\alpha = a_1 a_2 \cdots a_n$ in $\mathbf{N}(\mathbf{S})$. By Lemma 8 we can assume that $\alpha \neq k^n$ since the only rotation of k^n is found at the end of $FKM(\mathbf{S})$. Therefore, without loss of generality, we suppose α has suffix xk^j for some $x < k$. Let $p = |\text{ap}(\alpha)|$. We show that all p distinct rotations of α exist in $FKM(\mathbf{S})$. There are two cases depending on the value of i .

Case 1: $0 < i \leq p - j - 1$ By Lemma 9, $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{p-j-1}$ is a substring of $FKM(\mathbf{S})$. Observe that β is a substring of $a_1a_2 \cdots a_n \cdot a_1a_2 \cdots a_{p-j-1}$ when $0 < i \leq p - j - 1$.

Case 2: $p - j - 1 < i \leq p$ Observe that $\beta = k^{n-i}a_1a_2 \cdots a_{i-1}$ when $p - j - 1 < i \leq p$. Let γ be the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$ such that it has prefix $a_1a_2 \cdots a_{i-1}$. If γ is not the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$, then the previous necklace of γ has the suffix k^{n-i} due to the fact that $\mathbf{N}(\mathbf{S})$ is a k -suffix language. On the other hand, if γ is the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$, then the previous $n - i$ symbols in $FKM(\mathbf{S})$ are k^{n-i} by Lemma 8 when the sequence is considered circularly. Thus, β is a substring of $FKM(\mathbf{S})$.

Therefore, $FKM(\mathbf{S})$ contains each string in \mathbf{S} as a substring and is a universal cycle for \mathbf{S} since $|FKM(\mathbf{S})| = |\mathbf{S}|$. \square

We now extend this result to the FKM portion of Corollary 1.

Corollary 4. *If $\mathbf{S} \cup \{k^n\} \in \mathcal{C}(n, k)$, then $FKM(\mathbf{S})$ is a universal cycle for \mathbf{S} .*

Proof. If $\mathbf{S} \in \mathcal{C}(n, k)$, then by Theorem 4, $FKM(\mathbf{S})$ is a universal cycle for \mathbf{S} . Otherwise if $\mathbf{S} \notin \mathcal{C}(n, k)$ but $\mathbf{S} \cup \{k^n\} \in \mathcal{C}(n, k)$, then observe that $FKM(\mathbf{S}) \cdot k = FKM(\mathbf{S} \cup k^n)$. By Theorem 4 $FKM(\mathbf{S}) \cdot k$ is a universal cycle for $\mathbf{S} \cup k^n$. By Lemma 8, $FKM(\mathbf{S}) \cdot k$ ends with the suffix k^n . Hence, removing the last k in $FKM(\mathbf{S}) \cdot k$ only removes its substring k^n when considered circularly. Thus, all other strings in $\mathbf{S} \cup \{k^n\} \setminus \{k^n\} = \mathbf{S}$ are preserved. Therefore, $FKM(\mathbf{S})$ is a universal cycle for \mathbf{S} since $FKM(\mathbf{S})$ has length $|\mathbf{S}|$. \square

The proof of Theorem 4 explicitly states where the rotations of each necklace are found. To specify the position of a substring in a universal cycle we introduce the following notation. Suppose $FKM(\mathbf{S}) = u_0u_1 \cdots u_{|\mathbf{S}|-1}$ is a universal cycle for $\mathbf{S} \subseteq \mathbf{T}(n, k)$ and $\alpha \in \mathbf{S}$. Let $\text{last}(\alpha)$ be the last position of the substring α in $FKM(\mathbf{S})$. In other words, if $u_{i-n+1}u_{i-n+2} \cdots u_i = \alpha$, then $\text{last}(\alpha) = i$ because i is the last position of the substring α , where $0 \leq i < |\mathbf{S}|$ and the other index expressions are taken modulo $|\mathbf{S}|$.

Corollary 5. *If $\mathbf{S} \in \mathcal{C}(n, k)$, $\alpha \in \mathbf{N}(\mathbf{S})$ has suffix xk^j for some $x < k$, and β is a rotation of α , then $\text{last}(\beta) \leq \text{last}(\alpha) + |\text{ap}(\alpha)| - j - 1$.*

Proof. Let $\alpha = a_1a_2 \cdots a_n$. By Lemma 9, the rotations of α starting from a_i where $1 \leq i \leq |\text{ap}(\alpha)| - j$ are all found in succession starting from α itself. In particular, the last of these rotations β has $\text{last}(\beta) = \text{last}(\alpha) + |\text{ap}(\alpha)| - j - 1$. The remaining rotations of α end within the first necklace in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$ with prefix $a_1a_2 \cdots a_i$ for some i satisfying $1 \leq i \leq |\text{ap}(\alpha)| - j - 1$. None of these necklaces appear after α in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$, thus $\text{last}(\beta) < \text{last}(\alpha)$ for any such rotation of β . \square

This information will be crucial for the greedy algorithm in Section 5.

5 The generalized greedy approach

In this section we prove the greedy portion of Theorem 1 and Corollary 1, and that the generated universal cycles are all lexicographically smallest.

Let $\text{Greedy}(\mathbf{S})$ denote the sequence generated by the greedy algorithm after removing the initial k^{n-1} . We need to prove that $\text{Greedy}(\mathbf{S}) = FKM(\mathbf{S})$ and this is the lexicographically smallest universal cycle when $\mathbf{S} \in \mathcal{C}(n, k)$.

Lemma 10. *If $\mathbf{S} \in \mathcal{C}(n, k)$ and $|\mathbf{S}| > 1$, then the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$ is a prefix of $\text{Greedy}(\mathbf{S})$.*

Proof. By contradiction. Let $\alpha = a_1a_2 \cdots a_n$ be the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$. The string α is also the lexicographically smallest string in \mathbf{S} since \mathbf{S} is closed under rotation. Suppose $\text{Greedy}(\mathbf{S})$ starts with prefix $a_1a_2 \cdots a_iz$ for some $i < n$ and $z \neq a_{i+1}$. Observe that $z < a_{i+1}$ by the definition of the greedy algorithm. Furthermore, since the greedy algorithm starts with the initial seed k^{n-1} , after $i + 1$ iterations of the greedy algorithm the length n suffix of the sequence is $\beta = k^{n-i-1}a_1a_2 \cdots a_iz$. Notice that if $\beta \in \mathbf{S}$, then its rotation $a_1a_2 \cdots a_izk^{n-i-1}$ must also be in \mathbf{S} since \mathbf{S} is closed under rotation. However, $a_1a_2 \cdots a_izk^{n-i-1}$ is strictly less than α , a contradiction to α being the lexicographically smallest string in \mathbf{S} . \square

Theorem 5. *If $\mathbf{S} \in \mathcal{C}(n, k)$, then $\text{Greedy}(\mathbf{S})$ is equivalent to $\text{FKM}(\mathbf{S})$.*

Proof. Let $\alpha_1, \alpha_2, \dots, \alpha_m$ denote the lexicographic ordering of necklaces in $\mathbf{N}(\mathbf{S})$. Let $\mathcal{L}_t = \text{ap}(\alpha_1)\text{ap}(\alpha_2) \cdots \text{ap}(\alpha_t)$ for $1 \leq t \leq m$. When $m = 1$, $\mathbf{S} = \{k^n\}$ and the greedy algorithm terminates with the correct sequence of length one, namely $\text{Greedy}(\mathbf{S}) = \text{FKM}(\mathbf{S}) = k$.

For $m > 1$, we prove that $\text{Greedy}(\mathbf{S}) = \text{FKM}(\mathbf{S}) = \mathcal{L}_m$ by contradiction. Suppose t is the smallest value such that \mathcal{L}_{t+1} is not a prefix of $\text{Greedy}(\mathbf{S})$, where $\alpha_t = a_1 a_2 \cdots a_n = a_1 a_2 \cdots a_{n-j} k^j$ and $a_{n-j} < k$. From Lemma 10 we know that $1 \leq t < m$. Let $\alpha_{t+1} = b_1 b_2 \cdots b_n$ and $p = |\text{ap}(\alpha_{t+1})|$. Let i be the smallest value such that $0 < i \leq p$ and $\mathcal{L}_t \cdot b_1 b_2 \cdots b_i$ is not a prefix of $\text{Greedy}(\mathbf{S})$. Let β denote the length $n-1$ suffix of $\mathcal{L}_t \cdot b_1 b_2 \cdots b_{i-1}$. Such a suffix exists since both $\text{Greedy}(\mathbf{S})$ and $\text{FKM}(\mathbf{S})$ begin with α_1 by Lemma 7 and Lemma 10 when $m > 1$. There are two cases depending on the value of i .

Case 1: $0 < i \leq p - j - 1$ By Lemma 6, $b_1 b_2 \cdots b_{n-j-1} = a_1 a_2 \cdots a_{n-j-1}$. Therefore by Lemma 9, $\beta = a_{i+1} a_{i+2} \cdots a_n \cdot a_1 a_2 \cdots a_{i-1}$ since $\text{ap}(\alpha_t)$ is prior to $b_1 b_2 \cdots b_{i-1}$. Since $\mathcal{L}_t \cdot b_1 b_2 \cdots b_i$ is not a prefix of $\text{Greedy}(\mathbf{S})$, the greedy algorithm appends z to β where $z < b_i$. By Corollary 5, βz is not a rotation of the necklaces $\alpha_1, \alpha_2, \dots, \alpha_{t-1}$. However, a rotation of βz is equal to $a_1 a_2 \cdots a_{i-1} z a_{i+1} a_{i+2} \cdots a_n$ and is strictly less than α_t . Therefore, βz is a rotation of some necklace that is between α_{t-1} and α_t in lexicographic order, a contradiction to α_t being the necklace after α_{t-1} in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$. Thus z must be equal to $b_i = a_i$.

Case 2: $p - j - 1 < i \leq p$ By Lemma 9, $\beta = a_{i+1} a_{i+2} \cdots a_n b_1 b_2 \cdots b_{i-1}$ which is equal to $a_{i+1} a_{i+2} \cdots a_n a_1 a_2 \cdots a_{p-j-1} \cdot b_{p-j} b_{p-j+1} \cdots b_{i-1}$ since $\text{ap}(\alpha_t)$ is prior to $b_1 b_2 \cdots b_{i-1}$. Since $\mathcal{L}_t \cdot b_1 b_2 \cdots b_i$ is not a prefix of $\text{Greedy}(\mathbf{S})$, the greedy algorithm appends z to β where $z < b_i$. By Corollary 5, βz is not a rotation of the necklaces $\alpha_1, \alpha_2, \dots, \alpha_t$. However, a rotation of βz is equal to $a_1 a_2 \cdots a_{p-j-1} \cdot b_{p-j} b_{p-j+1} \cdots b_{i-1} z a_{i+1} a_{i+2} \cdots a_n$ and is strictly less than α_{t+1} . Therefore, βz is a rotation of some necklace that is between α_t and α_{t+1} in lexicographic order, a contradiction to α_{t+1} being the necklace after α_t in the lexicographic ordering of $\mathbf{N}(\mathbf{S})$. Thus z must be equal to b_i .

Thus by proof of contradiction, $\text{Greedy}(\mathbf{S}) = \text{FKM}(\mathbf{S}) = \mathcal{L}_m$ as claimed. \square

Corollary 6. *$\text{FKM}(\mathbf{S})$ and $\text{Greedy}(\mathbf{S})$ produce the lexicographically smallest universal cycle among all universal cycles for $\mathbf{S} \in \mathcal{C}(n, k)$.*

Proof. The greedy algorithm always starts with the lexicographically smallest necklace in $\mathbf{N}(\mathbf{S})$ by Lemma 10, which is also the lexicographically smallest string in \mathbf{S} . It then greedily appends the lexicographically smallest possible symbol such that the length n suffix is unique and in \mathbf{S} . By the definition of the greedy algorithm, appending any smaller symbol results in either a duplicate string, or a string not in \mathbf{S} . Thus $\text{Greedy}(\mathbf{S})$ must be the lexicographically smallest universal cycle. Also by Theorem 5, $\text{FKM}(\mathbf{S})$ and $\text{Greedy}(\mathbf{S})$ produce the same universal cycle. Therefore, both $\text{FKM}(\mathbf{S})$ and $\text{Greedy}(\mathbf{S})$ produce the lexicographically smallest universal cycle for \mathbf{S} . \square

6 Final Remarks

Although the language $\mathcal{C}(n, k)$ includes a broad class of combinatorial objects, there are still sets that are not in $\mathcal{C}(n, k)$ while their universal cycles can be constructed by the FKM algorithm. As an example, consider the set $\mathbf{S}_4 \subseteq \mathbf{T}(4, 3)$ which contains the following length 4 strings:

$$\begin{array}{l} 1112, 1121, 1122, 1212, 1211, 1221, 1222, 1322, 2111, \\ 2112, 2121, 2122, 2132, 2211, 2212, 2213, 2221, 3221. \end{array}$$

The set \mathbf{S}_4 is closed under rotation, but $\mathbf{N}(\mathbf{S}_4) = \{1112, 1122, 1212, 1222, 1322\}$ is not a k -suffix language. However, $\text{FKM}(\mathbf{S}_4) = 1112 \cdot 1122 \cdot 12 \cdot 1222 \cdot 1322$ is a universal cycle for \mathbf{S}_4 . Naturally, we would like to characterize the sets of strings \mathbf{S} in which $\text{FKM}(\mathbf{S})$ is a universal cycle. Similarly, we are interested in characterizing when the greedy algorithm works.

Acknowledgements

Joe Sawada's research is supported by NSERC grant number 214599.

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 (TCBB)*, 4(1):98–107, 2007.
- [2] Y. H. Au. Generalized de Bruijn words for primitive words and powers. *Discrete Mathematics*, 338(12):2320–2331, 2015.
- [3] A. Bechel, B. LaBounty-Lay, and A. Godbole. Universal cycles of discrete functions. In *Proceedings of the Thirty-Ninth Southeastern International Conference on Combinatorics, Graph Theory and Computing. Congressus Numerantium 189*, pages 121–128, 2008.
- [4] G. Brockman, B. Kay, and E. Snively. On universal cycles of labeled graphs. *Electronic Journal of Combinatorics*, 17(1):9 pp, 2010.
- [5] F. Chung, P. Diaconis, and R. Graham. Universal cycles for combinatorial structures. *Discrete Mathematics*, 110:43–59, December 1992.
- [6] N. G. de Bruijn. A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 49:758–764, 1946.
- [7] N. G. de Bruijn. Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2^n zeros and ones that show each n -letter word exactly once. *T.H. Report 75-WSK-06*, page 13, 1975.
- [8] R. Ehrenborg, S. Kitaev, and E. Steingrímsson. Number of cycles in the graph of 312-avoiding permutations. *Journal of Combinatorial Theory, Series A*, 129:1–18, 2015.
- [9] P. Fraigniaud and P. Gauron. D2B: A de Bruijn based content-addressable network. *Theoretical Computer Science*, 355(1):65–79, 2006.
- [10] H. Fredricksen. The lexicographically least de Bruijn cycle. *Journal of Combinatorial Theory*, 9(1):1 – 5, 1970.
- [11] H. Fredricksen. A class of nonlinear de Bruijn cycles. *Journal of Combinatorial Theory, Series A*, 19(2):192 – 199, 1975.
- [12] H. Fredricksen and I. J. Kessler. An algorithm for generating necklaces of beads in two colors. *Discrete Mathematics*, 61:181–188, 1986.
- [13] H. Fredricksen and J. Maiorana. Necklaces of beads in k colors and k -ary de Bruijn sequences. *Discrete Mathematics*, 23:207–210, 1978.
- [14] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley Professional, 2nd edition, 1994.
- [15] A. E. Holroyd, F. Ruskey, and A. Williams. Shorthand universal cycles for permutations. *Algorithmica*, 64(2):215–245, 2012.
- [16] G. Hurlbert. On universal cycles for k -subsets of an n -element set. *SIAM Journal on Discrete Mathematics*, 7:598–604, 1994.
- [17] B. Jackson. Universal cycles of k -subsets and k -permutations. *Discrete Mathematics*, 117:114–150, 1993.
- [18] R. Johnson. Universal cycles for permutations. *Discrete Mathematics*, 309:5264–5270, 2009.
- [19] D. Knuth. *The art of computer programming. Volume 4, fascicule 2. , Generating all tuples and permutations*. The art of computer programming. Addison-Wesley, Upper Saddle River (N.J.), 2005. Autre tirage : 2010.

- [20] C. E. Leiserson, H. Prokop, and K. H. Randall. Using de Bruijn sequences to index a 1 in a computer word. 1998, [Online; accessed June 2009].
- [21] A. Leitner and A. Godbole. Universal cycles of classes of restricted words. *Discrete Mathematics*, 310:3303–3309, 2010.
- [22] L. Levine. Sandpile groups and spanning trees of directed line graphs. *Journal of Combinatorial Theory, Series A*, 118:350–364, 2011.
- [23] M. H. Martin. A problem in arrangements. *Bulletin of the American Mathematical Society*, 40:859–864, 1934.
- [24] E. Moreno. On the theorem of Fredricksen and Maiorana about de Bruijn sequences. *Advances in Applied Mathematics*, 33(2):413–415, 2004.
- [25] F. Ruskey, C. Savage, and T. M. Y. Wang. Generating necklaces. *Journal of Algorithms*, 13:414–430, 1992.
- [26] F. Ruskey, J. Sawada, and A. Williams. Binary bubble languages and cool-lex order. *Journal of Combinatorial Theory, Series A*, 119(1):155 – 169, 2012.
- [27] F. Ruskey, J. Sawada, and A. Williams. De Bruijn sequences for fixed-weight binary strings. *SIAM Journal on Discrete Mathematics*, 26(2):605–617, 2012.
- [28] F. Ruskey and A. Williams. An explicit universal cycle for the $(n - 1)$ -permutations of an n -set. *ACM Transactions on Algorithms*, 6(3):12 pp, 2010.
- [29] C. Flye Sainte-Marie. Solution to question nr. 48. *L’intermédiaire des Mathématiciens*, 1:107–110, 1894.
- [30] J. Sawada, A. Williams, and D. Wong. Universal cycles for weight-range binary strings. In *Proceedings of 24th International Workshop on Combinatorial Algorithms (IWOCA 2013)*, LNCS 8288, pages 388–401. Springer, 2013.
- [31] J. Sawada, A. Williams, and D. Wong. The lexicographically smallest universal cycle for binary strings with minimum specified weight. *Journal of Discrete Algorithms*, 28(0):31 – 40, 2014. String-Masters 2012 & 2013 Special Issue (Volume 1).
- [32] B. Stevens, G. Hurlbert, and B. Jackson. Preface. *Discrete Mathematics*, 309(17):5255 – 5258, 2009. Generalisations of de Bruijn Cycles and Gray Codes/Graph Asymmetries/Hamiltonicity Problem for Vertex-Transitive (Cayley) Graphs.
- [33] B. Stevens and A. Williams. On the existence of universal cycles for languages that are closed under rotation, submitted, 2015.
- [34] B. Stevens and A. Williams. The coolest way to generate binary strings. *Theory of Computing Systems*, pages 1–27, 2013.
- [35] T. van Aardenne-Ehrenfest and N. G. de Bruijn. Circuits and trees in oriented linear graphs. *Mathematical Reviews*, 28:203–217, 1951.
- [36] D. Wong. *Novel universal cycle constructions for a variety of combinatorial objects*. PhD thesis, University of Guelph, Canada, 2015.