

On shortening u-cycles and u-words for permutations

Sergey Kitaev*, Vladimir N. Potapov[†] and Vincent Vajnovszki[‡]

July 19, 2017

Abstract

This paper initiates the study of shortening universal cycles (u-cycles) and universal words (u-words) for permutations either by using incomparable elements, or by using non-deterministic symbols. The latter approach is similar in nature to the recent relevant studies for the de Bruijn sequences. A particular result we obtain in this paper is that u-cycles for permutations of length n exist of lengths $n!$, $n! - (n - 1)$, $n! - 2(n - 1)$, \dots , $n! - (n - 1)!$.

1 Introduction

Chung et al. [3] introduced the notion of a *universal cycle*, or *u-cycle*, for permutations, which is a cyclic word such that any permutation of fixed length is order-isomorphic to exactly one factor (that is, to an interval of consecutive elements) in the word. In fact, the notion of a u-cycle for permutations can be extended to that of a u-cycle for any combinatorial class of objects admitting encoding by words [3]. In particular, universal cycles for sets of words are nothing else but the celebrated *de Bruijn sequences* [3]. De Bruijn sequences are

*Department of Computer and Information Sciences, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, United Kingdom, sergey.kitaev@strath.ac.uk

[†]Sobolev Institute of Mathematics, 4 Acad. Koptyug Ave, 630090 Novosibirsk, Russia, vpotapov@math.nsc.ru

[‡]LE2I, Université de Bourgogne, BP 47870, 21078 Dijon Cedex, France, vvajnov@u-bourgogne.fr

a well studied direction in discrete mathematics, and over the years they found widespread use in real-world applications, e.g. in the areas of molecular biology [2], computer security [6], computer vision [5], robotics [7] and psychology [8].

The existence of u-cycles (of length $n!$) for n -permutations (that is, permutations of length n) was shown in [3] for any n via *clustering the graph of overlapping n -permutations*. This graph has $n!$ vertices labelled by n -permutations, and there is an edge $x_1x_2 \cdots x_n \rightarrow y_1y_2 \cdots y_n$ if and only if the words $x_2x_3 \cdots x_n$ and $y_1y_2 \cdots y_{n-1}$ are order-isomorphic.

Each cluster collects all n -permutations whose first $n - 1$ elements form the same *pattern*. That is, the permutations inside a cluster are order-isomorphic to the same $(n - 1)$ -permutation. We call such a pattern the *signature* of a cluster, and we denote a signature by “ π ” where π is an $(n - 1)$ -permutation. See Figure 1 for the case of $n = 3$, and Figure 2 for the case of $n = 4$ where clusters are thought of as “super nodes”. There is exactly one edge associated with each permutation $x_1x_2 \cdots x_n$, which goes to the cluster with the signature that is order-isomorphic to $x_2x_3 \cdots x_n$. The edges are also viewed as edges between clusters.

Any Eulerian cycle in a graph formed by clusters can be extended to a Hamiltonian cycle in the graph of overlapping permutations (since each edge corresponds to exactly one permutation and we know this permutation), which, in turn, can be extended to a u-cycle for permutations via linear extensions of partially ordered sets as described in [3].

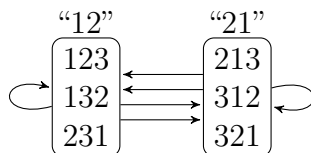


Figure 1: Clustering the graph of overlapping permutations of order 3

Removing the requirement for a u-cycle to be a cyclic word, while keeping the other properties, we obtain a *universal word*, or *u-word*. In this paper we deal both with the cyclic and non-cyclic cases related to the objects introduced below. This will cause no confusion though as from the context, it will always be clear which case we mean.

U-cycles provide an optimal encoding of a set of combinatorial objects in the sense that such an encoding is shortest possible. However, as is discussed in [1] for the case of *de Bruijn sequences*, one can still shorten u-cycles by using non-deterministic symbols. The studies in [1], mainly related to binary alphabets, were extended in [4] to the case of non-binary alphabets. In this paper, we will utilise the “shortening” idea, approaching the problem of shortening u-cycles for permutations from two different angles discussed next.

- Our non-determinism will be in using *incomparable elements* and considering *linear extensions of partial orders*, and we will study compression possibilities for u-cycles for permutations.
- Our second approach is a plain extension of the studies in [1, 4] to the case of permutations. However, using the “wildcard” symbol \diamond seems to be inefficient in the context (it is dominated by non-existence results; see Section 3.1), so we consider its refinement \diamond_S , where S is a subset of the alphabet in question (see Section 3.2).

1.1 Using linear extensions of partially ordered sets (posets) for shortening

To illustrate our idea, consider the word 112, which is claimed by us to be a u-cycle¹ for all permutations of length 3, thus shortening a “classical” u-cycle for these permutations, say, 145243. Indeed, we treat equal elements as *incomparable elements*, while the relative order of these incomparable elements to the other elements must be respected. Thus, 112 encodes all permutations whose last element is the largest one, namely, 123 and 213; starting at the second position (and reading the word cyclically), we obtain the word 121 encoding the permutations 132 and 231, and finally, starting at the third position, we (cyclically) read the word 211 encoding the permutations 312 and 321. More generally, it is clear that the word $\underbrace{11 \cdots 1}_{n-1 \text{ times}} 2 = 1^{n-1}2$

encodes all permutations and is of length n (instead of length $n!$ for earlier defined u-cycles for permutations). However, there are other compression possibilities creating u-cycles of lengths between n and

¹We modify the notion of a u-cycle for n -permutations introduced in [3] by allowing equal elements in a factor of length n and declaring them to be incomparable. Note that we still call the obtained object a “u-cycle for permutations”.

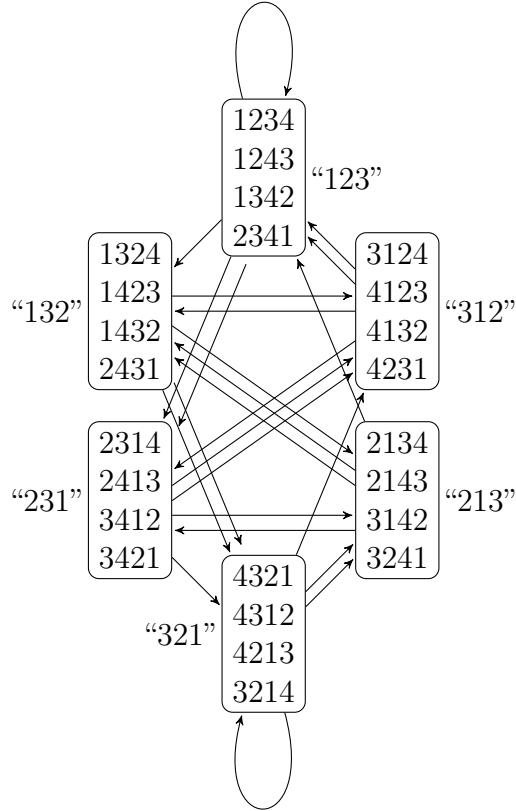


Figure 2: Clustering the graph of overlapping permutations of order 4

$n!$. For example, the word 1232 is also a u -cycle for permutations of length 3. Note that the word of the form $11 \cdots 1$ is a (trivial) u -word for all permutations of the respective length (when words are not read cyclically), while this word is not a u -cycle because the definition of a u -cycle cannot be applied to it.

The main goal of this paper is to study compression possibilities for (classical) u -cycles for permutations. In particular, we will show that such u -cycles exist of lengths $n! - kn$ for $k = 0, 1, \dots, (n-1)!$ (see Theorem 5). More specifically, our concern will be in existence of u -cycles for permutations in which equal elements do not stay closer than a fixed number of elements $d \geq 1$ from each other, that is, when there are at least $d - 1$ other elements between any pair of equal elements.

Note that the case of $d \geq n$ is not interesting when dealing with n -permutations since then equal elements cannot appear in the same factor of length n , and therefore, such a problem would be equivalent to constructing classical u-cycles for n -permutations, that has already been solved. Thus, the interesting values for d for us are between 1 and $n - 1$.

Finally, note that the problem can be modified by requiring from equal elements to stay *exactly*, rather than *at least*, at distance d , $1 \leq d \leq n - 1$, from each other, and then one can study the lengths of possible u-cycles for permutations, if any. The both problems are, of course, equivalent for the case $d = n - 1$, which we deal with in Section 2.1.

1.2 Using \diamond s for shortening

In [1, 4] u-cycles for words (de Bruijn sequences) and u-words for words are shortened using the \diamond symbol playing the role of a “wild-card” symbol, or a “universal symbol”. Any word containing a \diamond is called a *partial word*, or *p-word* in [1, 4], and the universal cycles/words obtained by shortening with \diamond s are called, respectively, *universal partial cycles*, or *u-p-cycles*, and *universal partial words*, or *u-p-words*. For example, $u = \diamond\diamond 0111$ is a u-p-word for binary words of length 3, since

- $\diamond\diamond 0$ covers 000, 010, 100 and 110;
- $\diamond 01$ covers 001 and 101; and
- the remaining factors in u cover 011 and 111.

As a straightforward extension of the objects in [1, 4] to the case of permutations, our u-cycles and u-words will contain $\diamond(s)$, whose meaning needs to be redefined though to avoid factors not order-isomorphic to permutations. In analogy with [1, 4], we call u-cycles and u-words for permutations containing at least one \diamond universal partial cycles (u-p-cycles) and universal partial words (u-p-words) for permutations, respectively. Introducing these notions helps us to distinguish between shortening using linear extensions of posets (when the resulting objects are still called by us u-cycles and u-words; see Section 1.1), and shortening using \diamond s, in which case the obtained objects are called u-p-cycles and u-p-words.

To see which of n -permutations are covered by a factor of length n , we keep the same relative order of non- \diamond elements, and insert all

possible elements instead of the $\diamond(s)$ that will result in the reduced form (see Subsection 1.3 for definitions) in an n -permutation. Following this definition, for $n = 3$, $1\diamond 2$ covers the permutations 213, 123 and 132, while for $n = 4$, $1\diamond 2\diamond$ covers the following 12 permutations: 3142, 3241, 2143, 2341, 2134, 2431, 1243, 1342, 1234, 1432, 1324 and 1423. Any factor of length n with k \diamond s covers $\frac{n!}{(n-k)!}$ permutations. Indeed, the number of ways to pick values for the \diamond s is $\binom{n}{k}$, and there are $k!$ ways to arrange these values.

We say that a u-p-word for n -permutations is *trivial* if it either does not contain any \diamond s, or contains only \diamond s. Obviously, \diamond is the only u-p-word for the permutation of length 1. Also, $\diamond 1$ is a u-p-word for 2-permutations. Proposition 14 below shows that if $n \geq 3$ then there is no u-p-word containing a single \diamond that is placed in position 1. This result, along with Propositions 15 and 16 and Corollaries 10 and 18, led us to the observation that usage of \diamond s in u-p-cycles, or u-p-words, for permutations may be too restrictive to be of practical use, and instead of a \diamond , one should use a *restricted* \diamond denoted \diamond_S , where S is a subset of $\{1, 2, \dots, n\}$ and n is the size of permutations in question. Indeed, even though no u-p-word for 3-permutations of the form $\diamond x_1 x_2 \dots x_k$ exists by Proposition 14, for example, $\diamond_{1,2} 243231$ is a u-p-word for 3-permutations (in particular, the factor $\diamond_{1,2} 24$ covers the permutations 123 and 213). See Theorem 19 for a result in this direction.

So, \diamond_S gives the permissible extensions out of n possible extensions given by \diamond . However, note that the notion of a \diamond_S is well-defined only if there is *at most* one \diamond_S in any factor of length n , since there is no meaning of, for example, the factor $\diamond_{1,2} \diamond_{1,2} \diamond_{1,2} 1$ for $n = 4$. Having said that, it is always acceptable to have $\diamond_{S_1}, \diamond_{S_2}, \dots, \diamond_{S_k}$ inside the same factor of length n as long as $S_1 \cap S_2 \cap \dots \cap S_k = \emptyset$.

1.3 Some basic definitions

For a word $w = w_1 \dots w_n$ over an ordered alphabet, we let $\text{red}(w)$ denote the word that is obtained from w by replacing each copy of the i -th smallest element in w by i . For example, $\text{red}(2547) = 1324$, $\text{red}(5470) = 3241$ and $\text{red}(436326) = 324214$.

Let π be a permutation of $\{1, \dots, n\}$ and x its element. For $x < n$, we let x^+ denote an element y such that $x < y < x + 1$, while for $x = n$, $x^+ = n + 1$. Also, for $x > 1$, we let x^- denote an element y such that $x - 1 < y < x$, while for $x = 1$, $x^- = 0$. The definitions of

x^+ and x^- can be generalized to any word instead of a permutation π in a straightforward way, namely, x^+ refers to an element larger than x but less than next largest element (if it exists), while x^- refers to an element smaller than x but larger than next smallest element (if it exists).

The *complement* of an n -permutation $\pi_1\pi_2\cdots\pi_n$ is the permutation obtained by replacing π_i by $n + 1 - \pi_i$. For example, the complement of 2314 is 3241. The *reverse* of a permutation is the permutation written in the reverse order. For example, the reverse of 2341 is 1432. In Section

1.4 Organization of the paper

This paper is organized as follows. In Section 2 we discuss shortening u-cycles for permutations via linear extensions of posets and present a key result, Theorem 5, giving possible lengths of u-cycles for permutations. An extension of this theorem in the case of $n = 4$ is discussed in Section 2.2. In Section 3 we discuss the usage of \diamond (see Section 3.1) and \diamond_S for the special case of $S = \{a, b\}$ (see Section 3.2) in the context of shortening u-cycles and u-words for permutations. Finally, in Section 4 we give some concluding remarks and state some problems for further research.

2 Shortening u-cycles for permutations via linear extensions of posets

In Section 2.1 we will derive Theorem 5 showing possible lengths of u-cycles when incomparable elements are allowed at distance $n - 1$ for n -permutations. In Section 2.2 we will provide an example for $n = 4$ of a shorter u-cycle than those given by Theorem 5. The example was obtained by allowing incomparable elements to be closer to each other (to be at distance 2 rather than at distance 3).

2.1 Incomparable elements at distance $n - 1$ for n -permutations

Definition 1. *Two permutations, $\pi_1 \cdots \pi_n$ and $\sigma_1 \cdots \sigma_n$, are called twin permutations, or twins, if*

- $\text{red}(\pi_1 \cdots \pi_{n-1}) = \text{red}(\sigma_1 \cdots \sigma_{n-1})$, and
- $|\pi_n - \pi_1| = |\sigma_n - \sigma_1| = 1$.

Examples of twins are 3124 and 4123, 2413 and 3412, and 23451 and 13452.

We refer the Reader to Figures 1 and 2 to check their understanding of the following four lemmas in the cases of $n = 3$ and $n = 4$, respectively.

Lemma 1. *Each cluster has exactly one pair of twins.*

Proof. Let the signature (the first $n-1$ elements of the permutations in the reduced form) of a cluster be “ $x_1 \cdots x_{n-1}$ ”. The only possibilities to create twin permutations are to adjoin x_1^+ or x_1^- at the end of $x_1 \cdots x_{n-1}$, and these possibilities always exists. \square

By *parallel edges* between clusters we mean multiple edges oriented in the same way. In particular, a pair (resp., a triple) of parallel edges is called a *double edge* (resp., a *triple edge*). In what follows, double and triple edges from a cluster X to a cluster Y will be denoted, respectively, by $\boxed{X} \rightarrow \boxed{Y}$ and $\boxed{X} \Rightarrow \boxed{Y}$.

Lemma 2. *For any cluster X , there exists the unique cluster Y such that $\boxed{X} \rightarrow \boxed{Y}$. Also, for no clusters X and Y , we have $\boxed{X} \Rightarrow \boxed{Y}$.*

Proof. Both of the statements follow from the fact that parallel edges can only be produced by twins, but by Lemma 1, there is only one such pair in each cluster. \square

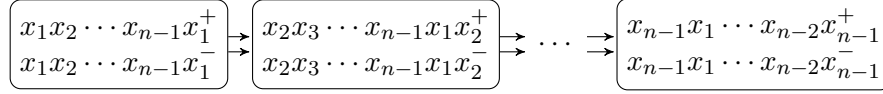
Lemma 3. *For any cluster Y , there exists the unique cluster X such that $\boxed{X} \rightarrow \boxed{Y}$.*

Proof. Let the signature of Y be “ $x_1 \cdots x_{n-1}$ ”. Then the only double edge that can come to Y is given by the permutations $x_{n-1}^- x_1 \cdots x_{n-1}$ and $x_{n-1}^+ x_1 \cdots x_{n-1}$ (both belonging to the same cluster with the signature “ $x_{n-1} x_1 \cdots x_{n-2}$ ”). \square

By Lemmas 2 and 3, the clustered graph of overlapping permutations can be partitioned into disjoint union of cycles formed by double edges.

Lemma 4. *Any of the disjoint cycles formed by the double edges goes through exactly $n - 1$ distinct clusters.*

Proof. Since double edges are formed by twin permutations, we can assume that any such cycle is of the form:



where the last cluster is linked to the first one by a double edge. Since all x_i s are distinct, the cycle must involve exactly $n - 1$ clusters. \square

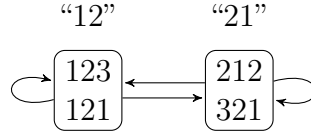
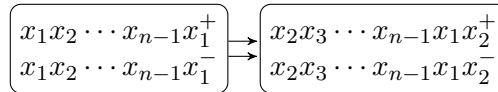


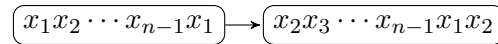
Figure 3: Applying incomparable elements on distance 2 for 3-permutations

Theorem 5. *Using incomparable elements at distance $n - 1$, one can obtain u -cycles for n -permutations of lengths $n!$, $n! - (n - 1)$, $n! - 2(n - 1)$, \dots , $n! - (n - 1)!$.*

Proof. There are $(n - 1)!$ clusters. By Lemma 4, there are $(n - 2)! = (n - 1)! / (n - 1)$ disjoint cycles formed by double edges, and we can decide in which cycles to replace each double edge by a single edge thus maintaining the property of the graph (whose nodes are clusters) being balanced. This action will correspond to replacing each double edge of the form



by



and thus introducing incomparable elements inside some of clusters. Strong connectivity in the graph will clearly be maintained as well, since our action is in simply replacing a pair of equivalent edges by a single edge.

So, by removing double edges in such a way we guarantee existence of an Eulerian cycle going through the clusters, which gives existence of the respective Hamiltonian cycle (recall that to each word or permutation there corresponds exactly one edge), and thus existence of a respective u-cycle for n -permutations. \square

See Figure 3 for an illustration of the proof of Theorem 5 in the case of $n = 2$, and Figure 4 for that in the case of $n = 3$ when both of double edges cycles were replaced. Examples of u-cycles that can be obtained from Figures 3 and 4, respectively, are 1232 and 123746576536738623.

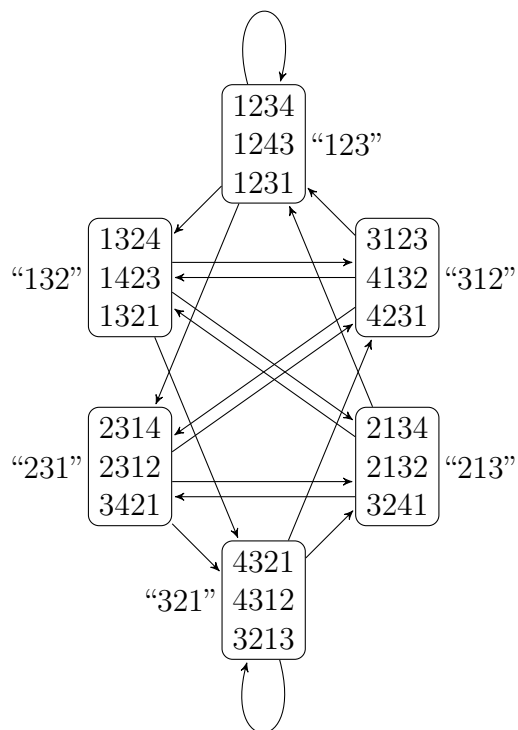


Figure 4: Applying incomparable elements at distance 3 for 4-permutations

2.2 Lengths of u-cycles for n -permutations different from n and those in Theorem 5

Theorem 5 gives a number of possible lengths for u-cycles for n -permutations, the shortest of which is $n! - (n - 1)!$. The trivial u-word $1^{n-2}2$ is of length n . A natural question is whether there exist u-cycles for n -permutations of length larger than n but smaller than $n! - (n - 1)!$. Clearly, such u-cycles could only be obtained if incomparable elements would be allowed on the distance smaller than $n - 1$. An example of such a u-cycle is 34321432345234 of length 14 for 4-permutations (note that $4! - 3! = 18$ is the shortest u-cycle given by Theorem 5). Clustering the graph of overlapping 4-permutations with incomparable elements used is shown in Figure 5.

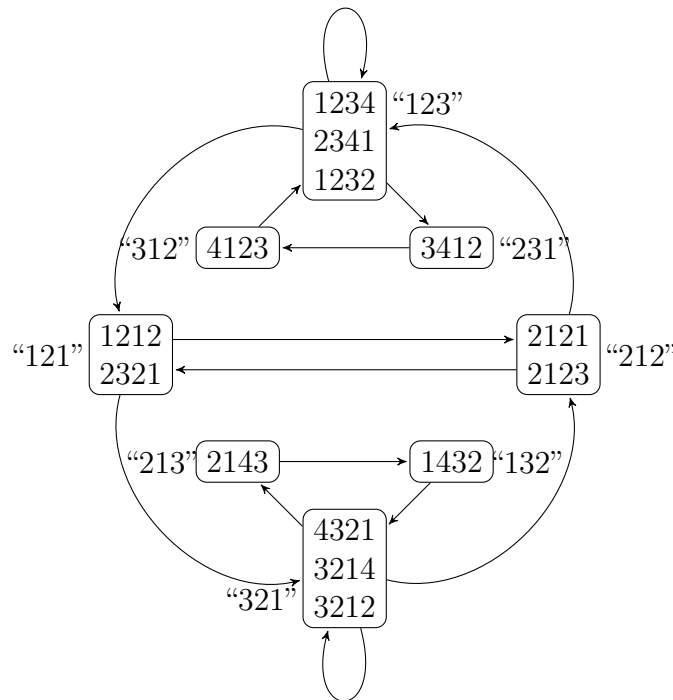


Figure 5: Clustering the graph of overlapping 4-permutations that corresponds to the u-cycle 34321432345234

3 Shortening u-cycles and u-words for permutations via usage of \diamond s

In this section we consider the shortening problem via usage of \diamond s. While the usage of the plain symbol \diamond seems to be dominated by various non-existence results (see Section 3.1), the usage of \diamond_S may potentially result in interesting classification theorems, an example of which is given in Section 3.2 (see Theorem 19).

3.1 Usage of \diamond s

The following lemma is an analogue in the case of permutations of Theorem 4.1 in [4] and Lemma 14 in [1] obtained for words.

Lemma 6. *Let $n \geq 3$ and $u = u_1 u_2 \cdots u_N$ be a u-p-cycle, or u-p-word, for n-permutations. If $u_k = \diamond$ then $u_{k+n} = u_{k-n} = \diamond$ assuming $k+n$ and/or $k-n$ exist in the case of u-p-words, and taking these numbers modulo N in the case of u-p-cycles.*

Proof. In what follows, the indices are taken modulo N in the circular case. Suppose that $u_k = \diamond$ and $u_{k+n} \neq \diamond$. Further, suppose that $\pi = \pi_1 \cdots \pi_{n-1}$ is one of the permutations obtained from $u_{k+1} \cdots u_{k+n-1}$ by substituting all the \diamond s, if any, by any permissible values and taking the reduced form.

For the circular case, because $u_k = \diamond$, the permutation π cannot be covered by any other factor of u (or else, some permutation ending with π in the reduced form would be covered twice). However, this means that if π is not monotone, at least one of the n -permutations $\text{red}(\pi 0)$ or πn is not covered by u ; contradiction. On the other hand, if π is monotone, then we use the fact that $n \geq 3$, so even though both $\text{red}(\pi 0)$ and πn can be covered by u , there is still at least one n -permutation not covered by u ; contradiction.

For the non-circular case, there is a possibility for π to occur one more time in u , namely, at its very beginning (that is, it is possible that $\text{red}(u_1 u_2 \cdots u_{n-1}) = \pi$). However, since $n \geq 3$, we know that at least one of the n -permutations $\text{red}(\pi 0)$, $\text{red}(\pi 1^+)$ or πn is not covered by u ; contradiction.

One can use similar arguments, or use the fact that the reverse of a u-p-cycle/u-p-word is a u-p-cycle/u-p-word, to show that $u_{k-n} = \diamond$. \square

By the previous lemma, for any \diamond in a u-p-cycle or u-p-word u , the other two symbols in distance n from it must be \diamond s as well. Thus the positions of \diamond s are periodic in u with period n , and any factor of u of length n contains equal number of \diamond s. It follows that the notion of the *diamondicity* introduced next is well defined (see also [3] where this notion was introduced in the context of u-p-words over non-binary alphabets).

Definition 2. For a u-p-cycle or u-p-word u for n -permutations, the *diamondicity* of u is the number of \diamond s in any length n factor in u .

3.1.1 U-p-cycles for permutations with $\diamond(s)$

Lemma 6 yields Corollary 8 below, which captures various rather restrictive conditions on relations between n and N to be satisfied by any u-p-cycle for permutations. In the proof of Corollary 8, we need the following easy number theoretical fact.

Lemma 7. If n and N are two positive integers, $c = \gcd(n, N)$, and $I = \{0, 1, \dots, \frac{N}{c} - 1\}$, then

$$\{i \cdot \frac{n}{c} \pmod{\frac{N}{c}} : i \in I\} = I.$$

Proof. We show that the integers of the form $i \cdot \frac{n}{c} \pmod{\frac{N}{c}}$, $i \in I$, are all different. If $i, i' \in I$ with $i \neq i'$, then $i \cdot \frac{n}{c} \pmod{\frac{N}{c}} \neq i' \cdot \frac{n}{c} \pmod{\frac{N}{c}}$. Indeed, otherwise $(i - i') \cdot \frac{n}{c}$ is a multiple of $\frac{N}{c}$, or equivalently $(i - i') \cdot n$ is a common multiple of n and N , which yields a contradiction since $|i - i'| < \frac{N}{c}$ and $\text{lcm}(n, N) = \frac{N}{c} \cdot n$. Thus, the sets in question have the same cardinality, which completes the proof. \square

Corollary 8. Let $u = u_1 u_2 \cdots u_N$ be a u-p-cycle (with or without $\diamond(s)$) for n -permutations. Then we have

(i) $N = k!$, where $n - k$ is the diamondicity of u .

In addition, if $c = \gcd(n, N)$, then

(ii) any length c factor of u contains the same number of \diamond s, and

(iii) $\frac{n}{c}$ divides $n - k$, so $c \neq 1$ for $1 \leq k \leq n - 1$.

Proof. (i) The number of \diamond s in each factor of u of length n is $n - k$, and thus such a factor covers $n(n - 1) \cdots (n - k + 1)$ permutations of length n , and there must be $k!$ length n factors (read cyclicly) to cover all $n!$ permutations.

(ii) Factoring u as $\underbrace{u_1 u_2 \cdots u_c}_{v_1} \underbrace{u_{c+1} u_{c+2} \cdots u_{2c}}_{v_2} \cdots \underbrace{u_{N-c+1} u_{N-c+2} \cdots u_N}_{v_{\frac{N}{c}}}$,

we have $u = v_1 v_2 \cdots v_{\frac{N}{c}}$ where v_i , $1 \leq i \leq \frac{N}{c}$, is the length c factor $u_{c \cdot (i-1)+1} u_{c \cdot (i-1)+2} \cdots u_{c \cdot i}$. With this notation, it follows that the number of \diamond s in v_1 is the same as that in $v_{\frac{n}{c}+1}$. Indeed, $v_1 v_2 \cdots v_{\frac{n}{c}}$ and $v_2 v_3 \cdots v_{\frac{n}{c}+1}$ are two length n factors of u which overlap when $c \neq n$, and by Lemma 6 they have the same number of \diamond s, and so do v_1 and $v_{\frac{n}{c}+1}$. Similarly, and taking the indices modulo $\frac{N}{c}$, the length c factors $v_{\frac{n}{c}+1}$ and $v_{\frac{2n}{c}+1}$ have the same number of \diamond s. And generally, each of the length c factors $v_{i \cdot \frac{n}{c}+1}$, $0 \leq i < \frac{N}{c}$, has the same number of \diamond s. By Lemma 7, the set $\{i \cdot \frac{n}{c} + 1 : 0 \leq i < \frac{N}{c}\}$ is precisely the set $\{i : 1 \leq i \leq \frac{N}{c}\}$, and thus each of the length c factors v_i , $1 \leq i \leq \frac{N}{c}$, has the same number of \diamond s.

Clearly, $u_2 u_3 \cdots u_N u_1$ is a u - p -cycle for n -permutations too, and factoring it as $\underbrace{u_2 u_3 \cdots u_{c+1}}_{v'_1} \underbrace{u_{c+2} u_{c+3} \cdots u_{2c+1}}_{v'_2} \cdots \underbrace{u_{N-c+2} u_{N-c+3} \cdots u_1}_{v'_{\frac{N}{c}}}$,

and reasoning as previously, we have that each v'_i has the same number of \diamond s (which is the same as that of v_i s). Repeating this process, we have finally that each length c factor of u has the same number of \diamond s.

(iii) By (ii) it follows that $\frac{n}{c}$ divides the number of \diamond s in each factor of length n . \square

The following corollary of Corollary 8 refines Lemma 6 in the case of u - p -cycles for permutations.

Corollary 9. *With the notations in Corollary 8, if u is a u - p -cycle for n -permutations, the positions of \diamond s in u are periodic with period c .*

In the next corollary, we give two proofs for the case when n is a prime number.

Corollary 10. *If n is a prime number, or $n = 4$, then there exists no u - p -cycle for n -permutations.*

Proof. If $n = 4$, it follows from Corollary 8 that the admissible values of N are 2 and 6, corresponding to $k = 2, 3$, respectively. Clearly only $N = 6$ can give the length of a u - p -cycle, thus $k = 3$ and $c = \gcd(n, N) = 2$. By (iii) in Corollary 8, $\frac{n}{c} = 2$ divides $n - k = 1$, contradiction.

If n is prime, from (ii) and (iii) in Corollary 8 (and with the notations therein) $\gcd(n, N) = n$, which contradicts (i) in Corollary 8, namely $N = k!$, with $k < n$.

An alternative proof for the case when n is prime is as follows. The total number of \diamond s counted in *all* factors of length n is $k!(n - k)$. However, each \diamond was counted exactly n times, so n must divide $k!(n - k)$, which is impossible if n is a prime number since $k < n$. \square

We conclude the subsection with two more non-existence results, the first of which is also applicable to u-p-words for permutations to be considered in the next subsection. Recall that by Lemma 6, \diamond s in a u-p-word or a u-p-cycle must occur periodically.

Theorem 11. *For any $n \geq 1$, there are no non-trivial u-p-word, or u-p-cycle for n -permutations in which \diamond s occur periodically with period 2.*

Proof. Suppose that such a u-p-word, or u-p-cycle $u = u_1 u_2 \cdots u_N$ for permutations exists, where $N \geq n + 1$ because u is non-trivial. Then $u_1 u_2 \cdots u_n$ is of one of the following four forms:

1. $u_1 \diamond u_3 \diamond u_5 \cdots u_n$;
2. $u_1 \diamond u_3 \diamond u_5 \cdots \diamond$;
3. $\diamond u_2 \diamond u_4 \diamond \cdots u_n$;
4. $\diamond u_2 \diamond u_4 \diamond \cdots \diamond$.

In either case, we claim that there exists an n -permutation that is covered by both $u_1 u_2 \cdots u_n$ and $u_2 u_3 \cdots u_{n+1}$, contradicting u 's properties. Next we provide such permutations for the first two cases; the remaining two cases are similar and their considerations are omitted.

1. $u_1 a_3 u_3 a_5 u_5 \cdots a_n u_n$, where $\text{red}(a_3 a_5 \cdots a_n) = \text{red}(u_3 u_5 \cdots u_n)$ and each of a_i s is larger than any u_j (clearly, the \diamond s can be assigned in such values). This permutation is also covered by $u_2 u_3 \cdots u_{n+1}$ by choosing the values of the \diamond s from left to right to be $b_3 b_5 \cdots b_{n+2}$, such that $\text{red}(b_3 b_5 \cdots b_{n+2}) = \text{red}(u_1 u_3 \cdots u_n)$, and each of b_i s is smaller than any u_j .
2. $u_1 a_1 u_3 a_3 \cdots u_n a_n$, where $\text{red}(a_1 a_3 \cdots a_n) = \text{red}(u_1 u_3 \cdots u_n)$ and each of a_i s is larger than any u_j . This permutation is also covered by $u_2 u_3 \cdots u_{n+1}$ by choosing the values of the \diamond s from left to right to be $b_3 b_5 \cdots b_{n+1}$, such that $\text{red}(b_3 b_5 \cdots b_{n+1}) = \text{red}(u_1 u_3 \cdots u_{n-1})$, and each of b_i s is smaller than any u_j .

□

Theorem 12. *For any $n \geq 1$, there are no non-trivial u - p -cycle for n -permutations in which \diamond s occur periodically with period 3.*

Proof. Using Corollary 8, 3 divides n , and we have three cases to consider based on which factor covers the increasing permutation. In each of the cases it is crucial that our universal word is cyclic, because we do not know the location of the factor covering the increasing permutation. Without loss of generality, we assume that in the factor covering the increasing permutation, the non- \diamond symbols are $1, 2, 3, \dots$

- The increasing permutation is covered by the factor

$$\diamond 12 \diamond 34 \dots \diamond \left(\frac{2n}{3} - 1 \right) \frac{2n}{3}.$$

Then this permutation is covered one more time starting from the letter 1, since the value of the \diamond next to $\frac{2n}{3}$ (cyclically) can be chosen $\left(\frac{2n}{3} + 1\right)$; contradiction.

- The increasing permutation is covered by the factor

$$12 \diamond 34 \diamond \dots \left(\frac{2n}{3} - 1 \right) \frac{2n}{3} \diamond.$$

Picking the value of the \diamond immediately to the left (cyclically) of the letter 1 be 1^- we see that the increasing permutation is covered one more time starting from this position.

- The increasing permutation is covered by the factor

$$1 \diamond 23 \diamond 4 \dots \left(\frac{2n}{3} - 1 \right) \diamond \frac{2n}{3}. \quad (1)$$

Consider the factor

$$23 \diamond 4 \dots \left(\frac{2n}{3} - 1 \right) \diamond \frac{2n}{3} x \diamond \quad (2)$$

of length n , where x is some letter. No matter what x is, we cover some permutation (not necessarily increasing) twice. Indeed, the rightmost \diamond in (2) can be chosen to be maximum in the permutation, while the rightmost \diamond in (1) can be chosen to be equivalent to x in (2).

□

Remark 13. *Unfortunately, the arguments in Theorems 11 and 12 do not seem to be possible to extend to periods of length 4, or more.*

3.1.2 U-p-words for permutations with $\diamond(s)$

Clearly, \diamond and $\diamond 1$ are, respectively, u-p-words for the 1-permutation and 2-permutations. The following proposition shows that these are the only u-p-words with a single \diamond placed at the beginning of the word. Before stating the proposition, we introduce a notion related to the clustered graph of overlapping permutations that will be used in some of our proofs.

Definition 3. Let $u_i u_{i+1} \cdots u_{i+n}$ be a factor of a u-p-word $u_1 u_2 \cdots u_N$ for n -permutations. We say that the edge coming out from the permutation $\text{red}(u_i u_{i+1} \cdots u_{i+n-1})$ in the clustered graph of overlapping permutations is used to reach the permutation $\text{red}(u_{i+1} \cdots u_{i+n})$.

Proposition 14. Let $n \geq 3$. No u-p-word for n -permutations with a single \diamond of the form $u = \diamond u_2 u_3 \cdots u_N$ exists.

Proof. Since $n \geq 3$, it is clear that $N \geq n + 1$. We can now apply Lemma 6 to obtain the desired. \square

The case $n = 4$ in the next proposition follows from our more general Corollary 18 below. However, we keep this case in Proposition 15 for yet another illustration of our straightforward approach to prove some of the non-existence statements.

Proposition 15. For $n = 3, 4$ there is no u-p-word for n -permutations of the form $u = u_1 \diamond u_3 u_4 \cdots u_N$.

Proof. Let $n = 3$. Without loss of generality (using the complement operation, if necessary), we can assume that u begins with $1 \diamond 2$. Then the possible continuations of u are $1 \diamond 22^+$, $1 \diamond 22^-$ and $1 \diamond 21^-$. But then the following permutations are covered twice, respectively, 123, 132 and 132.

Let $n = 4$. Without loss of generality (using the complement operation, if necessary), we can assume that there are three cases of beginning of u to consider.

- $1 \diamond 23$. Possible continuations are as follows.
 - $1 \diamond 234$. The permutation 1234 is covered twice; contradiction.
 - $1 \diamond 232^+ x$ for some x . Note that so far three permutations, namely, 1324, 1423, and $\text{red}(232^+ x)$ from the cluster with the signature “132” were covered. But the fourth permutation

from that cluster will never be covered (or else, because of $\diamond 232^+$, some permutation ending with the pattern 132 will be covered twice).

- $1\diamond 231x$ for some x . Because of the factor $\diamond 231$, the permutation $\text{red}(231x)$ will be the only one covered in the cluster with the signature “231” (no such permutation can be covered starting at the leftmost position, or at the \diamond). Contradiction with the cluster having four permutations.

- $1\diamond 32$.

- $1\diamond 324x$ for some x . Note that so far three permutations, namely, 2143, 3142, and $\text{red}(324x)$ from the cluster with the signature “213” were covered. But the fourth permutation from that cluster will never be covered (or else, because of $\diamond 324$, some permutation ending with the pattern 213 will be covered twice).
- $1\diamond 322^+x$ for some x . Because of the factor $\diamond 322^+$, the permutation $\text{red}(322^+x)$ will be the only one covered in the cluster with the signature “312” (no such permutation can be covered starting at the leftmost position, or at the \diamond). Contradiction with the cluster having four permutations.
- $1\diamond 321$. The permutation 1432 is covered twice; contradiction.

- $2\diamond 13$.

- $2\diamond 134$. The permutation 3124 is covered twice; contradiction.
- $2\diamond 132x$ for some x . Because of the factor $\diamond 132$, the permutation $\text{red}(132x)$ will be the only one covered in the cluster with the signature “132” (no such permutation can be covered starting at the leftmost position, or at the \diamond). Contradiction with the cluster having four permutations.
- $2\diamond 131^-x$ for some x . Note that so far three permutations, namely, 2314, 2413, and $\text{red}(131^-x)$ from the cluster with the signature “231” were covered. But the fourth permutation from that cluster will never be covered (or else, because of $\diamond 131^-$, some permutation ending with the pattern 231 will be covered twice).

□

Proposition 16. *For $n = 3$ there is no u - p -word for n -permutations of the form $u = u_1u_2\Diamond u_4u_5 \cdots u_N$.*

Proof. Because we need to cover six permutations, using the complement operation, if necessary, we can assume that $u = 12\Diamond u_4u_5$. Further, if $u_4 > 2$ then 123 would be covered twice, so $u_4 < 2$ and we can assume that $u = 12\Diamond 1u_5$ since the relevant order of the elements in positions 1 and 4 is not important. If $u_5 < 1$ then u would cover 132 twice (starting at positions 1 and 3). Thus, $u = 12\Diamond 12$ since the relevant order of the last element and the first two is not important. But then, the permutation 123 is covered twice (starting at positions 1 and 3). \square

Let u be a u - p -word for n -permutations with diamondicity d . It follows (see also the proof of the first part of Corollary 8) that u must contain exactly $(n - d)!$ different factors, and thus the length of u is $(n - d)! + n - 1$.

Theorem 17. *Let u be a non-trivial u - p -word for n -permutations, and let f be the number of \Diamond s in u . Then $n \leq 3f + 1$.*

Proof. Let $d \geq 1$ be the diamondicity of u . Thus, the length of u is $(n - d)! + n - 1$, and the number f of \Diamond symbols in u satisfies:

$$\begin{aligned} f &\geq \lfloor \frac{(n - d)! + n - 1}{n} \rfloor \cdot d \\ &\geq \frac{(n - d)!}{n} \cdot d \\ &\geq \frac{(n - d) \cdot (n - (d + 1))}{n} \\ &\geq n - (2d + 1). \end{aligned}$$

It follows that $n \leq f + 2d + 1 \leq 3f + 1$, and the statement follows. \square

As is mentioned above, \Diamond is a trivial u - p -word for the 1-permutation, and $\Diamond 1$ is a u - p -word for 2-permutations. These are the only u - p -words for permutations with a single \Diamond as shown by the following corollary.

Corollary 18. *For $n \geq 3$, there is no u - p -words for n -permutations with a single \Diamond .*

Proof. By Theorem 17, if u is a u - p -word for n -permutations with a single \Diamond , then $n \leq 4$.

Using the reverse operation, if necessary, one can assume that the single \diamond in a u-p-word for permutations is in its first half. Thus, by Proposition 14, Propositions 15 and 16, no u-p-word exists for 3-permutations.

If $n = 4$, then by Lemma 6, since we have exactly one \diamond , the length of a u-p-word must be at most 7, and if it is 7 then the \diamond must be in position 4. However, that means that such a u-p-word can cover at most $4 \times 4 = 16$ permutations instead of 24; contradiction. \square

3.2 Usage of $\diamond_{a,b}$

Theorem 19. *Let $n \geq 2$ and $a < b$. Then necessary and sufficient conditions for existence of a u-p-word for n -permutations of the form $\diamond_{a,b}u_2u_3 \cdots u_n$ are*

- $a = 1$ and $\text{red}(u_2u_3 \cdots u_n) = 12 \cdots (n-1)$, or
- $b = n$ and $\text{red}(u_2u_3 \cdots u_n) = (n-1)(n-2) \cdots 1$.

Proof. Similarly to the proof of Proposition 14, consider the cluster C corresponding to the signature “ $\text{red}(u_2u_3 \cdots u_n)$ ”. If $u_2u_3 \cdots u_n$ is not monotone (increasing or decreasing) then because of the factor $\diamond_{a,b}u_2u_3 \cdots u_n$ we see that reaching the permutation $\text{red}(u_2u_3 \cdots u_{n+1})$ in C (recall Definition 3) uses two edges, so that at least one permutation in C will never be covered by u . On the other hand, one can see that exactly the same situation occurs if $\text{red}(u_2u_3 \cdots u_n) = 12 \cdots (n-1)$ and $a \neq 1$ (if $a = 1$ then one of the two edges mentioned above is a loop and there is no contradiction), and if $\text{red}(u_2u_3 \cdots u_n) = (n-1)(n-2) \cdots 1$ and $b \neq n$ (again, if $b = n$ then one of the two edges is a loop giving no contradiction).

On the other hand, if one of the two conditions are satisfied, then we have that the $\diamond_{a,b}$ is responsible for removing an edge coming to the respective cluster C with a monotone signature and the loop connected to C from the clustered graph of overlapping permutations, as well as covering two permutations, one from C and one from another cluster C' . The rest of the word $u_2u_3 \cdots u_n$ corresponds to an Eulerian path beginning at C and ending at C' , which exists because each cluster is balanced, except for C (one extra out-edge) and C' (one extra in-edge), and the graph is clearly still strongly connected. \square

4 Concluding remarks

This agenda-setting paper is just scratching the surface of the new research direction of shortening u-cycles and u-words for permutations that naturally extends analogous studies conducted for the celebrated de Bruijn sequences [1, 4]. We were able to offer two different ways to approach the problem, namely, via linear extensions of posets, and via usage of (restricted) \diamond s, and we discussed several existence and non-existence results related to the context.

Out of possible directions for further research, it would be interesting to prove or disprove our guess that no distribution of \diamond s in a word can result in a non-trivial construction of a u-p-word for permutations. Also, we would like to see some characterization theorems involving (more than one) \diamond_S for S not necessarily of size 2, thus extending the result of Theorem 19.

Some enumerative questions can be raised as well. For example, one should be able to count u-cycles of various lengths in Theorem 5, that should be based on the choice of k cycles formed by double edges to be replaced by single edges (out of the total number of $(n - 2)!$ cycles formed by double edges).

Finally, we end up our paper with a discussion over extending the results of Section 2.2 to the case of permutations of length 5.

The clustered graph in Figure 5 uses incomparable elements only at distance 2, and it has a particularly nice symmetric structure. This motivated us to finding a similarly looking symmetric structure for 5-permutations, with a further hope to be able to find a nice structure for the general case. While looking at 5-permutations, to extend the construction of the graph in Figure 5, it is natural to either require incomparable elements to be at distance $n - 2 = 3$, or them to be at constant distance 2. In our experiments, we have not properly followed any of these assumptions. For example, while trying to assume the constant distance 2, we have used words 12121 and 21212 having incomparable elements at distance 4 as well. In either case, we can only state an attempt to find a u-cycle for 5-permutations of length less than $5! - 4! = 96$ rather than to give a solution in this case. We record a possible partial clustering of the graph of overlapping 5-permutations in Figure 6. The words/permutations presented in Figure 6 do not cover the same permutation twice, but we were not able to extend the graph in a balanced and strongly connected way to cover all 5-permutations exactly once, and we do not know whether

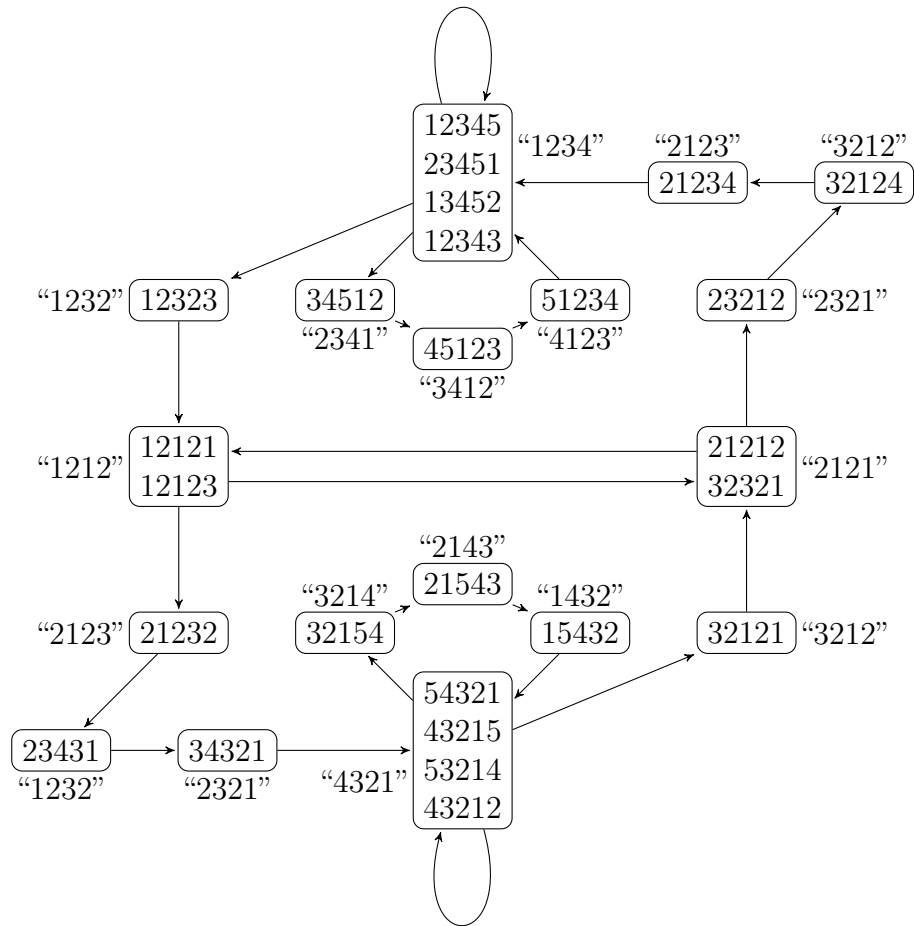


Figure 6: Partial clustering of the graph of overlapping 5-permutations

such an extension is possible for our partial clustering.

Acknowledgments

The authors are grateful to Sergey Avgustinovich for useful discussions related to the paper. Also, the first author is thankful to the Université de Bourgogne for its hospitality during the author's visit of it in November 2016. The second author is grateful to the Presidium of the Russian Academy of Sciences for supporting his research (Grant 0314-2015-0011).

References

- [1] H. Z.Q. Chen, S. Kitaev, T. Mütze, B. Y. Sun. On universal partial words. *Discrete Mathematics and Theoretical Computer Science*, to appear.
- [2] P. E. C. Compeau, P. A. Pevzner, G. Tesler. How to apply De Bruijn graphs to genome assembly. *Nature biotechnology*, **29(11)** (2011) 987–991.
- [3] F. Chung, P. Diaconis, R. Graham. Universal cycles for combinatorial structures. *Discrete Math.* **110** (1992) 43–59.
- [4] B. Goeckner, C. Groothuis, C. Hettle, B. Kell, P. Kirkpatrick, R. Kirsch, R. Solava. Universal Partial Words over Non-Binary Alphabets. arXiv:1611.03928, 2016.
- [5] J. Pagès, J. Salvi, C. Collewet, J. Forest. Optimised De Bruijn patterns for one-shot shape acquisition. *Image and Vision Computing*, **23(8)** (2005) 707 – 720.
- [6] A. Ralston. De Bruijn sequences—a model example of the interaction of discrete mathematics and computer science. *Math. Mag.*, **55(3)** (1982) 131–143.
- [7] E. R. Scheinerman. Determining planar location via complement-free De Bruijn sequences using discrete optical sensors. *IEEE Transactions on Robotics and Automation*, **17(6)** (2001) 883–889.
- [8] H. Sohn, D. L. Bricker, J. R. Simon, Y. Hsieh. Optimal sequences of trials for balancing practice and repetition effects. *Behavior*

Research Methods, Instruments, & Computers, **29(4)** (1997) 574–581.