2019 19th International Symposium on Communications and Information Technologies (ISCIT)

On the Number of Distinct k-Decks: Enumeration and Bounds

Johan Chrisnata*, Han Mao Kiah*, Sankeerth Rao†, Alexander Vardy†, Eitan Yaakobi‡, and Hanwen Yao†
*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
†Department of Electrical & Computer Engineering, University of California San Diego, LA Jolla, CA, USA
†Department of Computer Science, Technion Israel Institute of Technology, Haifa, Israel
Emails: johanchr001@ntu.edu.sg, hmkiah@ntu.edu.sg, sankeerth1729@gmail.com, avardy@ucsd.edu,
yaakobi@cs.technion.ac.il, hay125@eng.ucsd.edu

***Abstract*—The *k-deck* of a sequence is defined to be the multiset of all its subsequences of length k and let $D_k(n)$ denote the number of distinct k-decks for binary sequences of length n. In this paper, we determine the exact value of $D_k(n)$ for small values of k and n and provide asymptotic estimates of $D_k(n)$ when k is fixed.**

Specifically, for fixed k, we provide a trellis-based method to compute $D_k(n)$ in time polynomial in n. We then compute $D_k(n)$ for $k \in \{3,4,5,6\}$ and $k \le n \le 30$. We also improve the asymptotic upper bound on $D_k(n)$ and in particular, show $D_k(n) = O\left(n^{(k-1)2^{k-1}+1}\right)$. For the specific case when $k = 3$, we show $D_3(n) = \Omega(n^6)$ while the upper bound states that $D_3(n) = O(n^9)$.

I. Introduction

A protein macromolecule is a long string of amino acids. However, current sequencing technology either is unable to determine the *long* sequence directly or reads the sequence at a high error rate. Therefore, most sequencing methods obtain information about its *short* substrings or subsequences and attempt to infer or *reconstruct* the original string from this partial information. This gives rise to a myriad of combinatorial problems, known as string reconstruction problems [1], [2], [5], [6], [8], [10].

In this paper, we study the reconstruction problem involving k-decks. First described by Kalashnik [7], the *k-deck* of a sequence is defined to be the multiset of all its subsequences of length k. Traditionally, the k-deck problem is to determine $S(k)$, the smallest value of n such that *all* sequences of length n have unique k-decks. The exact values of $S(k)$ are known for $k \le 5$ and both upper and lower bounds for $S(k)$ have been extensively studied [3], [4], [7], [9], [12], [15].

Motivated by applications in DNA-based data storage (see [16] for a broad overview), we study the *coded* version of the k-deck problem. Consider sequences or words of length n. Instead of requiring *all* words to have different k-decks, we choose a subset of these words, or a *codebook*, such that every codeword in this codebook can be uniquely identified by its k-deck. In this setting, we consider the following fundamental problem: *how large can this codebook be*? Equivalently, this problem can be restated as an enumeration problem:

> Consider all words of length n. How many distinct k-decks are there?

Let $D_k(n)$ denote this quantity and $D_k(n)$ is the object of study for this paper. In another context, Rigo and Salimov used the term *k-binomial equivalence* to describe words with the same k-deck and provided rudimentary upper bounds on $D_k(n)$ [14]. In particular, Rigo and Salimov determined $D_2(n)$ and showed that $D_k(n) = O\left(n^{\Delta(k)}\right)$, where $\Delta(k) = 2((k-1)2^k+1)$. The values of $D_3(n)$ for $n \le 16$ are listed on the On-Line Encyclopedia of Integer Sequences [13].

In this paper, we provide a trellis-based method (see [11] for the definition of a trellis) to compute $D_k(n)$ and determine the exact values of $D_k(n)$ for $k \in \{3,4,5,6\}$ and $n \le 30$. We also provide asymptotic estimates for the case where k is fixed. In particular, we improve the asymptotic upper bound to $O(n^{\Delta(k)/4+1/2})$ for $k \ge 2$.

The paper is organized as follows. Section II formally defines the problem, summarizes previous results and states our contributions. Section III details a polynomial-time algorithm that computes $D_k(n)$. Section IV then provides an upper bound on $D_k(n)$ for general k, while Section V provides a lower bound for the case $k = 3$.

II. Problem Statement and Contributions

Let $\boldsymbol{X} = x_1x_2\cdots x_n$ be a binary word of length n. For $A \subseteq \{1,2,\ldots,n\}$, we use $\boldsymbol{X}_A$ to denote the subsequence with indices in A. In other words, $\boldsymbol{X}_A = x_{a_1}x_{a_2}\cdots x_{a_k}$ where $a_1 < a_2 < \cdots < a_k$ and $A = \{a_1,a_2,\ldots,a_k\}$. For $k \le n$, the *k-deck* of $\boldsymbol{X}$, denoted by $\boldsymbol{D}_k(\boldsymbol{X})$, refers to the multiset of all $\binom{n}{k}$ subsequences of length k. We represent the k-deck of a word $\boldsymbol{X}$ by an integer-valued vector of length 2^k. Specifically, $\boldsymbol{D}_k(\boldsymbol{X}) \triangleq (\boldsymbol{X}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\{0,1\}^k}$, where $\boldsymbol{X}_{\boldsymbol{\alpha}}$ denotes the number of occurrences of $\boldsymbol{\alpha}$ as a subsequence of $\boldsymbol{X}$ and the indices in $\{0,1\}^k$ are presented in an increasing lexicographic order.

Example 1. Let $\boldsymbol{X} = 110011$. Then $\boldsymbol{X}_{\{3,5,6\}} = \boldsymbol{X}_{\{4,5,6\}} = 011$ and we check that $\boldsymbol{X}_{011} = 2$. Furthermore, $\boldsymbol{D}_1(\boldsymbol{X}) = (2,4)$, $\boldsymbol{D}_2(\boldsymbol{X}) = (1,4,4,6)$ and $\boldsymbol{D}_3(\boldsymbol{X}) = (0,2,0,2,2,8,2,4)$. □

Two words $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be *k-equivalent*, or $\boldsymbol{X} \underset{k}{\sim} \boldsymbol{Y}$ if their k-decks are the same, i.e. $\boldsymbol{D}_k(\boldsymbol{X}) = \boldsymbol{D}_k(\boldsymbol{Y})$.

Example 2. Let $\boldsymbol{Y} = 101101$. Then $\boldsymbol{D}_1(\boldsymbol{Y}) = (2,4)$, $\boldsymbol{D}_2(\boldsymbol{Y}) = (1,4,4,6)$ and $\boldsymbol{D}_3(\boldsymbol{Y}) = (0,1,2,3,1,6,3,4)$. Hence, $\boldsymbol{X} \underset{k}{\sim} \boldsymbol{Y}$ for $k \in \{1,2\}$, but $\boldsymbol{X} \underset{3}{\nsim} \boldsymbol{Y}$. □

It can be shown that $\underset{k}{\sim}$ defines an equivalence relationship on all binary words. Furthermore, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same k-deck, then the lengths of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are necessarily the same. Hence, we fix n and partition the binary words of length n using the relation $\underset{k}{\sim}$. Then we set $D_k(n)$ to be the resulting number of equivalence classes. In this paper, we determine the exact value of $D_k(n)$ for $k \in \{3,4,5,6\}$ and $k \le n \le 30$ and provide asymptotic estimates of $D_k(n)$ when k is fixed.

For $k \in \{1,2\}$, the exact values on $D_k(n)$ and characterization of $\boldsymbol{D}_k(\boldsymbol{X})$ have been determined [14, Lemma 4].

Proposition 1. *Suppose that $\boldsymbol{X}$ is a binary word of length n.*

(i) Then $\boldsymbol{D}_1(\boldsymbol{X}) = (n-w, w)$, where $\boldsymbol{X}_1 = w$. Therefore, $D_1(n) = n+1$.

(ii) Then $\boldsymbol{D}_2(\boldsymbol{X}) = (\binom{n-w}{2}, t, w(n-w)-t, \binom{n-w}{2})$, where $\boldsymbol{X}_1 = w$ and $\boldsymbol{X}_{01} = t$. Therefore, $D_2(n) = (n^3+5n+6)/6$.

For $k \geq 3$, the best known upper bound on $D_k(n)$ is below.

Theorem 2 (Rigo and Salimov [14, Proposition 5]). *For all $n \geq k$, we have that $D_k(n) \leq \prod_{\ell=1}^{k}\left(\binom{n}{\ell}+1\right)^{2^{\ell}-1}$. When k is fixed, we have that $D_k(n) = O\left(n^{2((k-1)2^k+1)}\right)$.*

In addition to $D_k(n)$, we define $S(k) \triangleq \min\{n : D_k(n) < 2^n\}$. The exact values of $S(k)$ have been determined for $k \in \{3,4,5\}$ (see [4] for a survey). The first open case is $S(6)$ and the best known upper bound is given by Manvel et al. who constructed a pair of words of length thirty with the same 6-deck [12, Example 4].

Theorem 3. $S(6) \leq 30$.

For completeness, we present the best known upper and lower bounds for $S(k)$ that were summarized in [4].

Theorem 4. *We have that $S(k) = \Omega(k^2)$ and that*

$$S(k) \leq \begin{cases} 1.75 \cdot 1.62^k, & \textit{for } 7 \leq k \leq 28, \\ 0.25 \cdot 1.17^k k^3 \log k, & \textit{for } 29 \leq k \leq 84, \\ 3^{(3/2+o(1))\log_3^2 k}, & \textit{for } k \geq 85. \end{cases}$$

A. Our Contributions

(A) In Section III, we use a trellis structure to describe the k-decks and using this insight, we then provide an algorithm that enumerates all k-decks efficiently. When k is a constant, the technique computes $D_k(n)$ in polynomial time. We then compute $D_k(n)$ for $k \in \{3,4,5,6\}$ and $k \leq n \leq 30$ and establish that $S(6) = 30$.

(B) In Section IV, we improve the asymptotic upper bound on $D_k(n)$. In particular, we show that

$$D_k(n) = O\left(n^{(k-1)2^{k-1}+1}\right). \tag{1}$$

(C) In Section V, we look at the specific case for $k=3$ and show that $D_3(n) = \Omega(n^6)$. On the other hand, we note that (1) shows that $D_3(n) = O(n^9)$.

III. Polynomial-Time Enumeration

In this section, we introduce a trellis-based algorithm that calculates $D_k(n)$ for a fixed value of k. To compute $D_k(n)$, we construct a trellis with levels $i \in \{k, k+1, \ldots, n\}$ such that we are able to compute $D_k(i)$ at level i. At level $i+1$, instead of naively enumerating all k-decks for all possible $\boldsymbol{X} \in \{0,1\}^{i+1}$, our algorithm runs recursively and calculates the set $\{\boldsymbol{D}_k(\boldsymbol{X}) : \boldsymbol{X} \in \{0,1\}^{i+1}\}$ from the set $\{\boldsymbol{D}_k(\boldsymbol{X}) : \boldsymbol{X} \in \{0,1\}^{i}\}$, which reduces the complexity from 2^{i+1} down to $D_k(i)$.

To do so, we make two important combinatorial observations. In [14], it was observed that $\boldsymbol{X} \underset{k}{\sim} \boldsymbol{Y}$ implies $\boldsymbol{X} \underset{s}{\sim} \boldsymbol{Y}$ for all $1 \leq s < k$. The next proposition gives an explicit method to compute the s-deck from a k-deck.

Proposition 5. *Let $\boldsymbol{X} \in \{0,1\}^n$, $\boldsymbol{\alpha} \in \{0,1\}^s$ with $1 \leq s < k$, then*

$$\binom{n-s}{k-s}\boldsymbol{X}_{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta} \in \{0,1\}^k} \boldsymbol{\beta}_{\boldsymbol{\alpha}} \boldsymbol{X}_{\boldsymbol{\beta}} \tag{2}$$

Proof. For $A \subset [n]$, recall that $\boldsymbol{X}_A$ denotes the subsequence of $\boldsymbol{X}$ with indices in A. To demonstrate (2), we consider the two collections $\mathcal{A}$ and $\mathcal{B}$ of tuples. First, we set

$$\mathcal{A} \triangleq \left\{(A;S) : A, S \subset [n],\ \boldsymbol{X}_A = \boldsymbol{\alpha},\ |S| = k-s,\ A \cap S = \varnothing\right\}.$$

For each occurrence of $\boldsymbol{\alpha}$ in $\boldsymbol{X}$, we fix A and have $\binom{n-s}{k-s}$ choices for S. Therefore, the left hand side of (2) counts the number of tuples in $\mathcal{A}$.

Next, we set

$$\mathcal{B} \triangleq \left\{(\boldsymbol{\beta}; B; T) : B, T \subset [n],\ \boldsymbol{X}_B = \boldsymbol{\beta},\ T \subseteq B,\ \boldsymbol{X}_{B\setminus T} = \boldsymbol{\alpha}\right\}.$$

For each $\boldsymbol{\beta} \in \{0,1\}^k$ and each occurrence of $\boldsymbol{\beta}$ in $\boldsymbol{X}$, we have $\boldsymbol{\beta}_{\boldsymbol{\alpha}}$ choices for T. Therefore, the right hand side of (2) counts the number of tuples in $\mathcal{B}$.

To establish (2), it remains to exhibit a bijection between $\mathcal{A}$ and $\mathcal{B}$. Consider the map $\phi : \mathcal{A} \to \mathcal{B}$ such that $\phi(A;S) = (\boldsymbol{\beta}; B; S)$, where $B = A \cup S$ and $\boldsymbol{\beta} = X_B$.

For the inverse, we consider the $\psi : \mathcal{B} \to \mathcal{A}$ such that $\psi(\boldsymbol{\beta}; B; T) = (A; T)$, where $A = B \backslash T$. It is not difficult to verify that both $\phi \circ \psi$ and $\psi \circ \phi$ are identity maps on their respective domains. Therefore, we establish (2). $\square$

For $\boldsymbol{X} \in \{0,1\}^n$ and $a \in \{0,1\}$, let $(\boldsymbol{X}|a)$ denote the concatenation of $\boldsymbol{X}$ and a. Our second observation states that we can compute $\boldsymbol{D}_k(\boldsymbol{X}|a)$ from $\boldsymbol{D}_{k-1}(\boldsymbol{X})$ and $\boldsymbol{D}_k(\boldsymbol{X})$.

Proposition 6. *Let $\boldsymbol{X} \in \{0,1\}^n$, $\boldsymbol{\alpha} \in \{0,1\}^k$, then*

$$(\boldsymbol{X}|0)_{\boldsymbol{\alpha}} = \boldsymbol{X}_{\boldsymbol{\alpha}} + \boldsymbol{X}_{\boldsymbol{\beta}},\ \textit{where } (\boldsymbol{\beta}|0) = \boldsymbol{\alpha}, \tag{3}$$

$$(\boldsymbol{X}|1)_{\boldsymbol{\alpha}} = \boldsymbol{X}_{\boldsymbol{\alpha}} + \boldsymbol{X}_{\boldsymbol{\beta}},\ \textit{where } (\boldsymbol{\beta}|1) = \boldsymbol{\alpha}. \tag{4}$$

Proof. Consider the collection of index subsets:

$$S = \{A \subset [n] : (\boldsymbol{X}|0)_A = \boldsymbol{\alpha}\}.$$

Then S can be written as a disjoint union of S_1 and S_2 where

$$S_1 = \{A \in S : n+1 \notin A\}, \quad S_2 = \{A \in S : n+1 \in A\}.$$

Since $(\boldsymbol{X}|0)_{\boldsymbol{\alpha}} = |S|$, $\boldsymbol{X}_{\boldsymbol{\alpha}} = |S_1|$ and $\boldsymbol{X}_{\boldsymbol{\beta}} = |S_2|$, we have (3). Equation (4) can be proved in the same manner. $\square$

A more general version of Proposition 6 was given in [14]. As the authors did not furnish a proof, we provide one here for completeness.

We are ready to present our trellis. As mentioned earlier, the trellis has levels $i \in \{k, k+1, \ldots, n\}$. Each vertex at level i represents a k-deck of some word of length i and we denote the vertices at level i with $\mathbb{D}_k(i)$. Therefore,

$$\mathbb{D}_k(i) = \{\boldsymbol{D}_k(\boldsymbol{X}) : \boldsymbol{X} \in \{0,1\}^i\}.$$

Using (2), (3), and (4), each vertex in $\mathbb{D}_k(i)$ is extended by two edges labeled '0' and '1' to two vertices in $\mathbb{D}_k(i+1)$. The resulting trellis is *biproper*, in other words, every vertex has exactly two outgoing arcs with distinct labels and at most two incoming arcs with distinct labels. Furthermore, $\mathbb{D}_k(i)$ is the set of all k-decks of words of length i and the set of paths to a vertex,

Algorithm 1: Compute $D_k(n)$

```
1  initialize 𝔻_k(k) as follow:
     𝔻_k(k) = {(1,0,···,0),(0,1,···,0),···,(0,···,0,1)}
2  for i = k to n − 1 do
3      initialize 𝔻_k(i+1) as an empty set
4      for every k-deck D ∈ 𝔻_k(i) do
5          for a ∈ {0,1} do
6              Let X be a word such that D_k(X) = D.
7              Using equation (2) with (3) or (4),
                 compute D′ = D_k(X|a)
8              if D′ ∉ 𝔻_k(i+1) then
9                  insert D′ to the set 𝔻_k(i+1)
10 return D_k(n) = |𝔻_k(n)|
```

or a k-deck, is the set of all binary sequences having this k-deck. See Figure 1 for a trellis section for $k = 2$ and levels $i \in \{4, 5\}$. A formal description of the enumeration method is detailed in Algorithm 1.

We discuss the computational complexity of Algorithm 1. First, we note that $\boldsymbol{X}$ need not be explicitly found in Line 6. To compute $\boldsymbol{D}_k(\boldsymbol{X}|a)$, it suffices to apply (2), (3) and (4) to $\boldsymbol{D} = \boldsymbol{D}_k(\boldsymbol{X})$. Also, since Equations (2), (3) and (4) involve sums with at most 2^k terms, Lines 6 and 7 take constant time.

The time complexity for Line 7 depends on the data structure we used for $\mathbb{D}_k(n)$. If we use a binary search tree, we can insert each "new" k-deck in $O(D_k(n) \log D_k(n))$ time using $O(D_k(n))$ space. Therefore, Algorithm 1 runs in $O(nD_k(n) \log D_k(n))$ time using $O(nD_k(n))$ space. For fixed k, since $D_k(n)$ is polynomial in n by Theorem 2, the algorithm has space and time complexity polynomial in n.

To conclude this section, we compute the values of $D_k(n)$ for $k \in \{3, 4, 5, 6\}$ and $k \le n \le 30$ and present them in Table I. In particular, we computed that $D_6(n) = 2^n$ for $n \le 29$. Together with Theorem 3, we established the following.

Theorem 7. $S(6) = 30$.

IV. Upper Bounds on $D_k(n)$

Fix $k \ge 3$. In this section, we derive an upper bound on $D_k(n)$ To this end, we fix some $(k-1)$-deck $\boldsymbol{D}'$ and consider the collection $\mathcal{F}$ of all words of length n whose $(k-1)$-deck is given by $\boldsymbol{D}'$. Suppose that the number of k-equivalence classes in $\mathcal{F}$ is at most U for all choice of $(k-1)$-decks. Then an upper bound for $D_k(n)$ is simply given by $D_{k-1}(n)U$.

To find U, we consider an additional parameter $1 \le m \le k-1$ and define $J(k, m)$ to be all binary words of length k and weight m. Similar to [12], we relax the notion of k-equivalence and define the (k, m)-*equivalence relation*: $\boldsymbol{X} \underset{(k,m)}{\sim} \boldsymbol{Y}$ if and only if $(\boldsymbol{X}_{\boldsymbol{\beta}})_{\boldsymbol{\beta} \in J(k,m)} = (\boldsymbol{Y}_{\boldsymbol{\beta}})_{\boldsymbol{\beta} \in J(k,m)}$. Suppose that the number of (k, m)-equivalence classes in $\mathcal{F}$ is at most $U(m)$ for all choice of $(k-1)$-decks. Then we can obtain $U = \prod_{m=1}^{k-1} U(m)$.

We now proceed to estimate $U(m)$. Now, suppose $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{F}$. Since $\boldsymbol{X} \underset{k-1}{\sim} \boldsymbol{Y}$, we have that $\boldsymbol{X} \underset{1}{\sim} \boldsymbol{Y}$. In other words, $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same weight. Hence, we let w denote the weight of any word in $\mathcal{F}$.

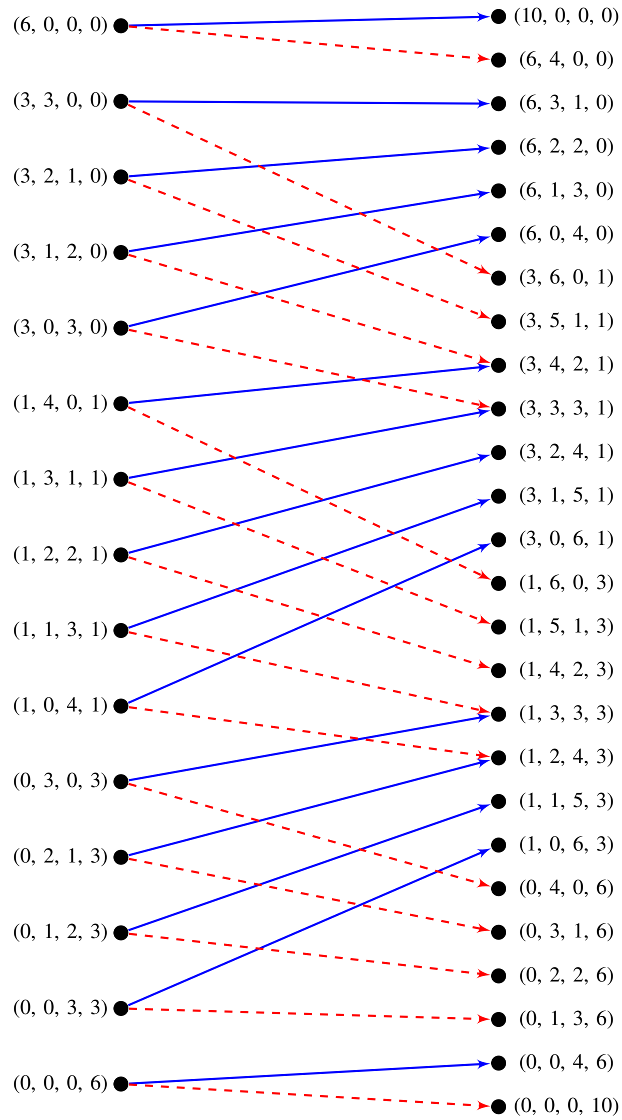

Fig. 1. Trellis section for $k = 2$ and levels $i \in \{4, 5\}$. Left vertices belong to $\mathbb{D}_2(4)$, while right vertices belong to $\mathbb{D}_2(5)$. Blue solid edges corresponds to label '0', while red dashed edges corresponds to label '1'.

Let $\boldsymbol{X} \in J(n, w)$, the set of all words of length n and weight w. Suppose that $\boldsymbol{\beta} \in J(k, m)$ and let $\boldsymbol{\alpha}$ be a subsequence of length $k-1$ of $\boldsymbol{\beta}$. Then $\boldsymbol{\alpha}$ necessarily belongs to $J(k-1, m-1) \cup J(k-1, m)$. Recall that $\boldsymbol{\beta}_{\boldsymbol{\alpha}}$ is the number of occurrences of $\boldsymbol{\alpha}$ in $\boldsymbol{\beta}$. Below we obtain a combinatorial relationship between $\boldsymbol{X}_{\boldsymbol{\alpha}}$, $\boldsymbol{X}_{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_{\boldsymbol{\alpha}}$, which is a refinement of Proposition 5.

Proposition 8. *Let* $\boldsymbol{X} \in J(n, w)$ *and* $1 \le m \le k-1$.
If $\boldsymbol{\alpha} \in J(k-1, m-1)$, *then*

$$(w - m + 1)\boldsymbol{X}_{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta} \in J(k,m)} \boldsymbol{\beta}_{\alpha} \boldsymbol{X}_{\boldsymbol{\beta}}. \tag{5}$$

If $\boldsymbol{\alpha} \in J(k-1, m)$, *then*

$$(n - k + m - w + 1)\boldsymbol{X}_{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta} \in J(k,m)} \boldsymbol{\beta}_{\boldsymbol{\alpha}} \boldsymbol{X}_{\boldsymbol{\beta}}\,. \tag{6}$$

Proof. Let $\boldsymbol{X} = x_1 x_2 \cdots x_n$. Suppose that $\boldsymbol{\alpha} \in J(k-1, m-1)$. To demonstrate (5), we proceed in a similar manner as in the

$n\backslash k$	3	4	5	6
3	8	—	—	—
4	16	16	—	—
5	32	32	32	—
6	64	64	64	64
7	**126**	128	128	128
8	247	256	256	256
9	480	512	512	512
10	926	1024	1024	1024
11	1764	2048	2048	2048
12	3337	**4092**	4096	4096
13	6208	8176	8192	8192
14	11408	16328	16384	16384
15	20608	32604	32768	32768
16	36649	65075	**65534**	65536
17	63976	129824	131064	131072
18	109866	258906	262120	262144
19	185012	516168	524212	524288
20	306285	1028448	1048360	1048576
21	497190	2048272	2096586	2097152
22	792920	4077316	4192896	4194304
23	1241936	8111400	8385216	8388608
24	1913566	16124458	16769254	16777216
25	2898574	32034016	33536094	33554432
26	4323980	63579386	67067294	67108864
27	6353060	126076522	134124596	134217728
28	9206137	249736704	268228914	268435456
29	13158574	494124382	536416730	536870912
30	18576644	976302888	1072750464	**1073741820**

TABLE I
VALUES OF $D_k(n)$ FOR $3 \le k \le 6$ AND $k \le n \le 30$. VALUES HIGHLIGHTED IN BOLD CORRESPOND TO $D_k(S(k))$.

proof of Proposition 5 and consider the following two collections $\mathcal{A}$ and $\mathcal{B}$ of tuples. Set

$$\mathcal{A}^* \triangleq \left\{(A;s) : A \subset [n],\ \boldsymbol{X}_A = \boldsymbol{\alpha},\ x_s = 1,\ s \notin A\right\}.$$

Since $\boldsymbol{\alpha}$ has weight $m-1$ and $\boldsymbol{X}$ has weight w, we have $(w-m+1)$ choices for s for each occurrence of $\boldsymbol{\alpha}$ in $\boldsymbol{X}$. Therefore, the left hand side of (5) counts the number of tuples in $\mathcal{A}$.

Next, set

$$\mathcal{B}^* \triangleq \left\{(\boldsymbol{\beta}; B; t) : B \subset [n],\ \boldsymbol{X}_B = \boldsymbol{\beta},\ t \in B,\ \boldsymbol{X}_{B\setminus\{t\}} = \boldsymbol{\alpha}\right\}.$$

For each $\boldsymbol{\beta} \in J(k,m)$ and each occurrence of $\boldsymbol{\beta}$ in $\boldsymbol{X}$, we have $\boldsymbol{\beta}_{\boldsymbol{\alpha}}$ choices for t. Therefore, the right hand side of (5) counts the number of tuples in $\mathcal{B}$.

To establish (5), it remains to exhibit a bijection between $\mathcal{A}^*$ and $\mathcal{B}^*$. Consider the maps ϕ and ψ defined in the proof of Proposition 5. When we restrict the domains of ϕ and ψ to $\mathcal{A}^*$ and $\mathcal{B}^*$, respectively, the maps are well-defined bijections from $\mathcal{A}^*$ to $\mathcal{B}^*$ and vice versa. Hence, we obtain (5). When $\boldsymbol{\alpha} \in J(k-1,m)$, Equation (6) can be similarly established by requiring $x_s = 0$ in the definition of $\mathcal{A}^*$. $\square$

Define $\boldsymbol{H}^{(k,m)}$ to be the $\binom{k}{m} \times \binom{k}{m}$ matrix whose rows and columns are indexed by $J(k-1,m-1) \cup J(k-1,m)$ and $J(k,m)$, respectively. The entries of $\boldsymbol{H}^{(k,m)}$ are given by $\boldsymbol{H}^{(k,m)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \triangleq \boldsymbol{\beta}_{\boldsymbol{\alpha}}$.

Further define a column vector $\boldsymbol{Z}$ of length $\binom{k}{m}$ such that the first $\binom{k-1}{m-1}$ entries are given by $\left((w-m+1)\boldsymbol{X}_{\boldsymbol{\alpha}}\right)_{\boldsymbol{\alpha}\in J(k-1,m-1)}$ and the next $\binom{k-1}{m}$ entries are given by $\left((n-k+m-w+1)\boldsymbol{X}_{\boldsymbol{\alpha}}\right)_{\boldsymbol{\alpha}\in J(k-1,m)}$. Then (5) and (6) imply that

$$\boldsymbol{H}^{(k,m)} \left(\boldsymbol{X}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)} = \boldsymbol{Z}. \tag{7}$$

Example 3. Let $k=3$ and $m=2$. Then

$$\boldsymbol{H}^{(3,2)} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}.$$

Consider $\boldsymbol{X} = 110011$ and so, $n=6$, $w=4$. Also,

$$\left(\boldsymbol{X}_{\boldsymbol{\beta}}\right)^T_{\boldsymbol{\beta}\in J(3,2)} = (2,8,2),$$
$$\boldsymbol{Z}^T = (3\cdot 4, 3\cdot 4, 2\cdot 6) = (12,12,12).$$

We verify that (7) holds. $\square$

The following lemma then characterizes the (k,m)-equivalence of two words when they share the same $(k-1)$-deck.

Lemma 9. *Let $\boldsymbol{H}^{(k,m)}$ be as defined above. If $\boldsymbol{X} \underset{k-1}{\sim} \boldsymbol{Y}$, then $\left(\boldsymbol{X}_{\boldsymbol{\beta}} - \boldsymbol{Y}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)}$ belongs to the nullspace of $\boldsymbol{H}^{(k,m)}$.*

Proof. Since $\boldsymbol{X} \underset{k-1}{\sim} \boldsymbol{Y}$, we have that $\boldsymbol{H}^{(k,m)} \left(\boldsymbol{X}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)} = \boldsymbol{H}^{(k,m)} \left(\boldsymbol{Y}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)}$. Hence, $\boldsymbol{H}^{(k,m)} \left(\boldsymbol{X}_{\boldsymbol{\beta}} - \boldsymbol{Y}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)} = 0$ and the lemma follows. $\square$

Therefore, it remains to provide an upper bound on the nullity of $\boldsymbol{H}^{(k,m)}$.

Proposition 10. *The nullity of $\boldsymbol{H}^{(k,m)}$ is at most $\binom{k-2}{m-1}$.*

Proof. We write $\boldsymbol{H} = \boldsymbol{H}^{(k,m)}$ for short. Recall that the columns of $\boldsymbol{H}$ are indexed by $J(k,m)$ and we arrange the columns in an increasing lexicographic order as in $J(k,m)$. We demonstrate that the nullity of $\boldsymbol{H}$ is at most $\binom{k-2}{m-1}$ by exhibiting $\binom{k}{m} - \binom{k-2}{m-1}$ columns with leading coefficients.

We have the following cases.

- Let $\boldsymbol{\beta} = \beta_1\beta_2\cdots\beta_k \in J(k,m)$ with $\beta_1 = 0$. Then consider the row $\boldsymbol{\alpha} \triangleq \beta_2\cdots\beta_k \in J(k-1,m)$ and clearly, $\boldsymbol{H}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ge 1$. Suppose $\boldsymbol{\beta}' \in J(k,m)$ and $\boldsymbol{\beta}'_{\boldsymbol{\alpha}} \ge 1$. Then $\boldsymbol{\beta}$ is necessarily lexicographically smaller than or equal to $\boldsymbol{\beta}'$. In other words, $\boldsymbol{H}_{\boldsymbol{\alpha},\boldsymbol{\beta}''} = 0$ for all words $\boldsymbol{\beta}''$ that are lexicographically smaller than $\boldsymbol{\beta}$. Therefore, $\boldsymbol{H}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is the leading coefficent of row $\boldsymbol{\alpha}$.
- Let $\boldsymbol{\beta} = \beta_1\beta_2\cdots\beta_k \in J(k,m)$ with $\beta_k = 1$. Then consider the row $\boldsymbol{\alpha} \triangleq \beta_1\beta_2\cdots\beta_{k-1} \in J(k-1,m-1)$ and as before, $\boldsymbol{H}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ge 1$. Proceeding as before, we observe that $\boldsymbol{H}_{\boldsymbol{\alpha},\boldsymbol{\beta}''} = 0$ for all words $\boldsymbol{\beta}'' \in J(k,m)$ that are lexicographically smaller than $\boldsymbol{\beta}$. Therefore, $\boldsymbol{H}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is the leading coefficent of row $\boldsymbol{\alpha}$.

Hence, the columns with possibly no leading coefficients start with a one and end with a zero. Therefore, there are $\binom{k-2}{m-1}$ such columns and the proposition follows. $\square$

Finally, we state the main theorem for this section and provide an upper bound on $D_k(n)$. Recall that $\boldsymbol{X} \underset{(k,m)}{\sim} \boldsymbol{Y}$ if and only if $\left(\boldsymbol{X}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)} = \left(\boldsymbol{Y}_{\boldsymbol{\beta}}\right)_{\boldsymbol{\beta}\in J(k,m)}$.

Theorem 11. *The number of (k,m)-equivalence classes for words of length n with the same $(k-1)$-deck is $O\left(n^{k\binom{k-2}{m-1}}\right)$.*

Therefore, the number of distinct k-decks with the same $(k-1)$-deck is $O\left(n^{k2^{k-2}}\right)$ and hence, $D_k(n) = O\left(n^{(k-1)2^{k-1}+1}\right)$.

Proof. Fix $\boldsymbol{X}$ to be of length n. Suppose that $\boldsymbol{Y} \underset{k-1}{\sim} \boldsymbol{X}$. Then Lemma 9 states that $(\boldsymbol{Y_\beta} - \boldsymbol{X_\beta})_{\boldsymbol{\beta} \in J(k,m)}$ belongs to the nullspace of $\boldsymbol{H}^{(k,m)}$. Since the nullity of $\boldsymbol{H}^{(k,m)}$ is at most $\binom{k-2}{m-1}$ and every entry of $(\boldsymbol{Y_\beta})_{\boldsymbol{\beta} \in J(k,m)}$ is at most $\binom{n}{k} = O(n^k)$, the number of choices for $(\boldsymbol{Y_\beta})_{\boldsymbol{\beta} \in J(k,m)}$ is $O\left(n^{k\binom{k-2}{m-1}}\right)$.

Therefore, the number of distinct k-decks with the same $(k-1)$-deck is

$$O\left(\prod_{m=1}^{k-1} n^{k\binom{k-2}{m-1}}\right) = O\left(n^{k(\sum_{m=1}^{k-1}\binom{k-2}{m-1})}\right) = O\left(n^{k2^{k-2}}\right).$$

Finally, it follows from simple induction that

$$D_k(n) = D_{k-1}(n) \cdot O\left(n^{k2^{k-2}}\right) = O\left(n^{(k-1)2^{k-1}+1}\right). \quad \square$$

V. Lower Bounds on $D_3(n)$

In this section, we focus on the case $k = 3$ and prove that $D_3(n) = \Omega(n^6)$. As with Section IV, we consider the words with the same $(k-1)$-deck, or 2-deck, and determine the number of 3-equivalence classes amongst these words.

Let $\boldsymbol{X} \in J(n,w)$. Following [12], we consider the notion of zero-vectors. The *zero-vector* of $\boldsymbol{X}$, denoted by $\boldsymbol{u_X} = (u_0, u_1, \ldots, u_w)$, is the vector of length $w+1$, where u_0 is the number of zeroes in front of the first one, u_w is the number of zeroes after the last one, and u_j is the number of zeroes between the jth one and the $(j+1)$th one for any $1 \le j \le w-1$. In other words, if $\boldsymbol{u_X} = (u_0, u_1, \ldots, u_w)$, then $\boldsymbol{X} = 0^{u_0}10^{u_1}1\cdots10^{u_w}$.

Recall that $\boldsymbol{X_\alpha}$ is the number of occurrences of $\boldsymbol{\alpha}$ as a subsequence of $\boldsymbol{X}$. Set $\boldsymbol{X}_{01} = t$ and $\boldsymbol{X}_1 = w$. Our objective is to estimate the possible values of $\boldsymbol{X}_{011}$. To this end, we have the following lemma from [12].

Lemma 12 ([12, Lemma 13])**.** *For $k \ge 1$, define the following $k \times (w+1)$-integer-valued matrix:*

$$\boldsymbol{M}_k \triangleq \begin{pmatrix} \binom{w}{k-1} & \binom{w-1}{k-1} & \binom{w-2}{k-1} & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \binom{w-1}{k-2} & 2\binom{w-2}{k-2} & \cdots & (w-k+1)\binom{k-1}{k-2} & (w-k+2) & \cdots & 0 \\ 0 & 0 & \binom{w-2}{k-3} & \cdots & \binom{w-k+1}{2}\binom{k-1}{k-3} & \binom{w-k+2}{2}\binom{k-2}{k-3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{w-k+1}{k-1} & \binom{w-k+2}{k-1} & \cdots & \binom{w}{k-1} \end{pmatrix}.$$

Then $\boldsymbol{M}_k\boldsymbol{u_X} = (\boldsymbol{X_\alpha})_{\alpha \in J(k,k-1)}$.

Example 4. Let $\boldsymbol{X} = 110011$ and $k = 3$. So, $\boldsymbol{u_X} = (0,0,2,0,0)$ and

$$\boldsymbol{M}_k = \begin{pmatrix} 6 & 3 & 1 & 0 & 0 \\ 0 & 3 & 4 & 3 & 0 \\ 0 & 0 & 1 & 3 & 6 \end{pmatrix}.$$

We verify that $(\boldsymbol{X_\alpha})^T_{\alpha \in J(3,2)}$ is indeed given by $\boldsymbol{M}_3\boldsymbol{u_X} = (2,8,2)^T$.

Lemma 13. *Set Γ to be the following set of vectors of length $w+1$:*

$$\Gamma \triangleq \{(1,-2,1,0,0,\ldots,0),(0,1,-2,1,0,\ldots,0), (0,0,1,-2,1,\ldots,0),\ldots,(0,\ldots,1,-2,1)\}.$$

If $\boldsymbol{u_{X'}} = \boldsymbol{u_X} + \boldsymbol{c}$ for some $\boldsymbol{c} \in \Gamma$, then $\boldsymbol{X}_1 = \boldsymbol{X}'_1$, $\boldsymbol{X}_{01} = \boldsymbol{X}'_{01}$ and $\boldsymbol{X}'_{011} = \boldsymbol{X}_{011} + 1$.

Proof. It follows from the definition of zero-vectors that $\boldsymbol{X}_1 = \boldsymbol{X}'_1$. For the other two equalities, we have that $\boldsymbol{M}_k\boldsymbol{u_{X'}} = \boldsymbol{M}_k\boldsymbol{u_X} + \boldsymbol{M}_k\boldsymbol{c}$. Applying Lemma 12 for $k \in \{2,3\}$ and considering the first coordinate, we have

$$\boldsymbol{X}'_{01} = \boldsymbol{X}_{01} + \boldsymbol{M}_2\boldsymbol{c} \text{ and } \boldsymbol{X}'_{011} = \boldsymbol{X}_{011} + \boldsymbol{M}_3\boldsymbol{c}.$$

Then $\boldsymbol{M}_2\boldsymbol{c} = 0$ because $a - 2(a+1) + (a+2) = 0$ for all a. On the other hand, $\boldsymbol{M}_3\boldsymbol{c} = 1$ because $\binom{a}{2} - 2\binom{a+1}{2} + \binom{a+2}{2} = 1$ for all a. The lemma is then immediate. $\square$

Example 5. As before, let $\boldsymbol{X} = 110011$ and $k = 3$. Further set $\boldsymbol{X}' = 101101$ and so, $\boldsymbol{u_{X'}} = (0,1,0,1,0)$ and $\boldsymbol{X}' = \boldsymbol{X} + (0,1,-2,1,0)$. We verify that $\boldsymbol{X}_1 = \boldsymbol{X}'_1 = 4$, $\boldsymbol{X}_{01} = \boldsymbol{X}'_{01} = 4$, and $\boldsymbol{X}'_{011} = \boldsymbol{X}_{011} + 1 = 3$.

Next, we have the following proposition.

Proposition 14. *Fix n, $0 \le w \le n$ and $0 \le t \le w(n-w)$. Then there exists $\boldsymbol{X}$ and $\boldsymbol{Y}$ of length n such that the following hold:*

(i) $\boldsymbol{X}_1 = \boldsymbol{Y}_1 = w$ and $\boldsymbol{X}_{01} = \boldsymbol{Y}_{01} = t$;
(ii) $\boldsymbol{X}_{011} \le (n-w)\binom{q+1}{2}$ where $q = \lfloor t/(n-w) \rfloor$;
(iii) $\boldsymbol{Y}_{011} \ge \frac{w-1}{2}\left(t - \binom{w}{2}\right)$;
(iv) for any $\boldsymbol{X}_{011} \le s \le \boldsymbol{Y}_{011}$, there exists $\boldsymbol{Z}$ such that $\boldsymbol{Z}_1 = w$, $\boldsymbol{Z}_{01} = t$ and $\boldsymbol{Z}_{011} = s$.

Proof. Write that $t = q(n-w) + r$. Set $\boldsymbol{X}$ to be the word whose zero vector is $(A_0, A_1, \ldots, A_w)$, where $A_{w-q-1} = r$, $A_{w-q} = n-w-r$ and all others are 0. Recall that $\boldsymbol{X}_{01} = \sum_{i=0}^{w} u_i(w-i)$ and indeed, $\boldsymbol{X}_{01} = r(q+1) + (n-w-r)q = t$. Furthermore, $\boldsymbol{X}_{011} = r\binom{q+1}{2} + (n-w-r)\binom{q}{2} \le (n-w)\binom{q+1}{2}$.

To construct $\boldsymbol{Y}$, we iteratively add some $\boldsymbol{c}$ from Γ to $\boldsymbol{u_X}$. Specifically, we start with $\boldsymbol{X}^{(0)} = \boldsymbol{X}$ and $\boldsymbol{u}^{(0)} = \boldsymbol{u_X}$ and suppose that we have $\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(i)}$. If $\boldsymbol{u}^{(i)}$ has a component with value at least two at index j with $1 \le j \le w-1$, we choose $\boldsymbol{c}^{(i)} = (0,0,\ldots,0,1,-2,1,0,\ldots,0)$ with the minus two at index j. Then we set $\boldsymbol{u}^{(i+1)} = \boldsymbol{u}^{(i)} + \boldsymbol{c}^{(i)}$ and $\boldsymbol{X}^{(i+1)}$ to be the corresponding word. It follows from Lemma 13 that the two-deck of $\boldsymbol{X}^{(i+1)}$ is the same as $\boldsymbol{X}^{(i)}$ and the number of 011 in $\boldsymbol{X}^{(i)}$ is given by $\boldsymbol{X}_{011} + i$.

Hence, we terminate the process when the components of $\boldsymbol{u}^{(i)}$ are at most one on the indices from 1 to $w-1$. Let $(B_0, B_1, \ldots, B_w)$ be the final zero-vector and $\boldsymbol{Y}$ be the corresponding word. It then remains to show Proposition 14(iii).

Again, we have $0 \le B_i \le 1$ for all $1 \le i \le w-1$. Furthermore, we have $\boldsymbol{Y}_{01} = \boldsymbol{X}_{01} = t$, which means

$$t = wB_0 + \sum_{i=1}^{w-1}(w-i)B_i \le wB_0 + \sum_{i=1}^{w-1} w-i \le wB_0 + \binom{w}{2}.$$

Therefore, $B_0 \ge \frac{t}{w} - \frac{w-1}{2}$ and so, $\boldsymbol{Y}_{011} \ge B_0\binom{w}{2} \ge \frac{w-1}{2}\left(t - \binom{w}{2}\right)$. $\square$

Following this proposition, for purposes of brevity, we write

$$s' \triangleq (n-w)\binom{q+1}{2},$$
$$s'' \triangleq \frac{w-1}{2}\left(t - \binom{w}{2}\right).$$

Therefore, a lower bound for $D_3(n)$ is

$$\sum_{w=0}^{n}\sum_{t=0}^{w(n-w)} \max\{0, s''-s'\}. \tag{8}$$

We estimate the terms of (8). Note that

$$\begin{aligned}
\sum_{w=0}^{n}\sum_{t=0}^{w(n-w)} s'' &= \sum_{w=0}^{n}\sum_{t=0}^{w(n-w)} \frac{w-1}{2}\left(t-\binom{w}{2}\right)\\
&= \sum_{w=0}^{n}\frac{w-1}{2}\sum_{t=1}^{w(n-w)} t-\binom{w}{2}\\
&\geq \sum_{w=0}^{n}\frac{w-1}{2}\left(\frac{1}{2}w^2(n-w)^2-\binom{w}{2}w(n-w)\right)\\
&\geq \sum_{w=0}^{n}\frac{w-1}{4}\left(w^2(n-w)^2-w^3(n-w)\right)\\
&= \sum_{w=0}^{n}\frac{w-1}{4}\left(w^2(n-w)(n-2w)\right)\\
&= \sum_{w=0}^{n}\frac{1}{4}w^3(n-w)(n-2w)+O\left(n^4\right).
\end{aligned} \tag{9}$$

On the other hand,

$$\begin{aligned}
\sum_{w=0}^{n}\sum_{t=0}^{w(n-w)} s' &= \sum_{w=0}^{n}\sum_{t=0}^{w(n-w)} (n-w)\binom{q+1}{2}\\
&= \sum_{w=0}^{n}\sum_{t=0}^{w(n-w)} \frac{1}{2}(n-w)q^2+O\left(n^2\right)\\
&= \sum_{w=0}^{n}\sum_{q=0}^{w} \frac{1}{2}(n-w)^2q^2+O\left(n^3\right)\\
&\quad \text{since for each } q,\ r \text{ can go from } 0 \text{ to } n-w-1\\
&= \sum_{w=0}^{n}\frac{1}{6}(n-w)^2w^3+O\left(n^4\right).
\end{aligned} \tag{10}$$

Combining (9) and (10) into (8), we have that the number of 3-decks is at least

$$\begin{aligned}
&\sum_{w=0}^{n/4}\frac{1}{4}w^3(n-w)(n-2w)-\frac{1}{6}(n-w)^2w^3+O\left(n^4\right)\\
&= \sum_{w=0}^{n/4} w^3(n-w)\left(\frac{1}{4}n-\frac{1}{2}w-\frac{1}{6}n+\frac{1}{6}w\right)+O\left(n^4\right)\\
&= \sum_{w=0}^{n/4}\frac{1}{12}w^3(n-w)\left(n-4w\right)+O\left(n^4\right)=\Omega(n^6)\,.
\end{aligned} \tag{11}$$

We summarize our discussion in the following theorem.

Theorem 15. $D_3(n)=\Omega(n^6)$.

Remark 6. The statement in Theorem 15 can be made more precise. We have demonstrated that the number of $(3,2)$-equivalence classes amongst all words of length n is $\Omega(n^6)$. It then follows from Theorem 11 that this estimate is tight.

Remark 7. Implicit in the proof of Proposition 14 is an efficient method that encodes messages into words with distinct 3-decks. Specifically, let the message set be

$$\mathcal{M} \triangleq \{(w,t,s) : 0\leq w\leq n,\ 0\leq t\leq w(n-w),\ s'\leq s\leq s''\}.$$

Given any triple $(w,t,s)\in\mathcal{M}$, we can construct $\boldsymbol{X}$ in linear time such that $\boldsymbol{X}_1=w$, $\boldsymbol{X}_{01}=t$ and $\boldsymbol{X}_{011}=s'$.

Following the procedure described in the proof of Proposition 14, we choose a sequence of $\boldsymbol{c}^{(0)},\boldsymbol{c}^{(1)},\ldots\in\Gamma$ to add to $\boldsymbol{u}_{\boldsymbol{X}}$. Since $s'\leq s\leq s''$, there is a sequence such that the resulting zero-vector corresponds to $\boldsymbol{Y}$ and $\boldsymbol{Y}_{011}=s$. Therefore, $\boldsymbol{Y}$ is the codeword encoding the message (w,t,s) and $\boldsymbol{Y}$ can be computed in $O(n^3)$ time.

VI. Conclusion

We provide an efficient trellis-based method to compute the number of distinct k-decks and determined the exact value of $D_k(n)$ for $k\in\{3,4,5,6\}$ and $k\leq n\leq 30$. An interesting consequence is that we established the fact that $S(6)=30$.

We also establish an asymptotic upper bound on $D_k(n)$ for general k and an asymptotic lower bound for $D_3(n)$. In summary, we have $D_3(n)=O(n^9)$ and $D_3(n)=\Omega(n^6)$. It remains open to determine tight bound on the asymptotic growth rate of $D_3(n)$.

References

[1] J. Acharya, H. Das, O. Milenkovic, A. Orlitsky, and S. Pan, "String reconstruction from substring compositions," SIAM Journal on Discrete Mathematics 29, no. 3, 1340-1371, 2015.

[2] Z. Chang, J. Chrisnata, M. F. Ezerman and H. M. Kiah, "Rates of DNA sequence profiles for practical values of read lengths," IEEE Transactions on Information Theory, vol. 63, no. 11, pp. 7166-7177, Nov. 2017.

[3] C. Choffrut, J. Karhumäki, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, vol. I, Springer, Berlin, 1997, pp. 329-438.

[4] M. Dudik and L. J. Schulman, "Reconstruction from subsequences," Journal of Combinatorial Theory, vol. 103, no. 2, pp. 337-348, 2003.

[5] R. Gabrys and O. Milenkovic, "The hybrid k-deck problem: Reconstructing sequences from short and long traces," in 2017 IEEE International Symposium on Information Theory (ISIT), Jun. 2017, pp. 1306-1310.

[6] R. Gabrys and O. Milenkovic, "Unique reconstruction of coded sequences from multiset substring spectra," in 2018 IEEE International Symposium on Information Theory (ISIT), Jun. 2018, pp. 2540-2544.

[7] Kalashnik, L. O. "The reconstruction of a word from fragments," Numerical Mathematics and Computer Technology, Akad. Nauk. Ukrain. SSR Inst. Mat., Preprint IV (1973): 56-57.

[8] H. M. Kiah, G. J. Puleo, and O. Milenkovic, "Codes for DNA sequence profiles," IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3125-3146, 2016.

[9] I. Krasikov and Y. Roditty, "On a reconstruction problem for sequences," Journal of Combinatorial Theory, vol. 77, no. 2, pp. 344-348, 1997.

[10] P. Ligeti and P. Sziklai, "Reconstruction from subwords," in 6th International Conference on Applied Informatics, Jan. 2004, pp. 1-7.

[11] J. L. Massey, "Foundation and methods of channel encoding," in Proc. Int. Conf. Information Theory and Systems, vol. 65, NTG-Fachberichte, 1978.

[12] B. Manvel, A. Meyerowitz, A. Schwenk, K. Smith, and P. Stockmeyer, "Reconstruction of sequences," Discrete Math, vol. 94, no. 3, pp. 209-219, 1991.

[13] OEIS Foundation Inc. (2019), The On-Line Encyclopedia of Integer Sequences, http://oeis.org/A258585

[14] M. Rigo, P. Salimov, "Another generalization of abelian equivalence: Binomial complexity of infinite words," Theoretical Computer Science, vol. 601, pp. 47-57, 2015.

[15] A.D. Scott, "Reconstructing sequences," Discrete Mathematics, vol. 175, no. 1-3, pp. 231-238, 1997.

[16] S. M. H. T. Yazdi, H. M. Kiah, E. Garcia-Ruiz, J. Ma, H. Zhao, and O. Milenkovic, "DNA-based storage: Trends and methods," IEEE Transactions on Molecular, Biological and Multi-Scale Communications, vol. 1, no. 3, pp. 230-248, Sep. 2015.