The Capacity of the Weighted Read Channel

Omer Yerushalmi, **Tuvi Etzion** and **Eitan Yaakobi**
Department of Computer Science, Technion — Israel Institute of Technology, Haifa, 3200003 Israel
Email: {omeryer, etzion, yaakobi}@cs.technion.ac.il

***Abstract*— One of the primary sequencing methods gaining prominence in DNA storage is nanopore sequencing, attributed to various factors. In this work, we consider a simplified model of the sequencer, characterized as a channel. This channel takes a sequence and processes it using a sliding window of length ℓ, shifting the window by δ characters each time. The output of this channel, which we refer to as the *read vector*, is a vector containing the sums of the entries in each of the windows. The capacity of the channel is defined as the maximal information rate of the channel. Previous works have already revealed capacity values for certain parameters ℓ and δ. In this work, we show that when $\delta < \ell < 2\delta$, the capacity value is given by $\frac{1}{\delta}\log_2 \frac{1}{2}(\ell+1+\sqrt{(\ell+1)^2-4(\ell-\delta)(\ell-\delta+1)})$. Additionally, we construct an upper bound when $2\delta < \ell$. Finally, we extend the model to the two-dimensional case and present several results on its capacity.**

I. Introduction

DNA storage is an emerging technology driven by the increasing demand for data storage. Consequently, there has been significant progress in both synthesis and sequencing technologies [1]–[4]. One particular sequencing methodology, named the nanopore sequencer, is mainly renowned for its support in long reads, low cost, and high probability [5]–[7]. The nanopore sequencing process proceeds as follows: when reading a DNA strand, its nucleotides traverse a pore sequentially. In this continuous process, a constant number of nucleotides, denoted by ℓ, pass through the pore simultaneously each time. The output of the reading process is determined by the values of each of the ℓ nucleotides. While this sequencing technique is efficient in multiple aspects, it also presents some challenges. Primarily, the output of the reading process experiences inter-symbol interference (ISI) due to the dependence on the values of ℓ nucleotides simultaneously rather than just one. Moreover, this process may frequently introduce random errors in the reading output, leading to occurrences like duplications or deletions of certain nucleotides Thus, various models have been proposed for the nanopore sequencer [8], [9].

In this work, we focus on a specific model corresponding to the transverse-read model as outlined and studied in [10], which was motivated by racetrack memories. This model has been proposed in [9], and also studied in [11]. It constitutes a specific instance of the ISI channel, which is characterized by parameters (ℓ, δ), and is denoted as (ℓ, δ)-*weighted read channel*. From this point forward, we refer to it as the (ℓ, δ)-*read channel*. This channel characterizes the reading operation of the sequencer as a sliding window of size ℓ, shifting over the sequence in increments of δ. Thus, for a sequence $(x_1, x_2, \ldots, x_n)$, the initial read examines the first ℓ characters, $(x_1, x_2, \ldots, x_\ell)$, while the subsequent read occurs with a shift of δ characters, i.e., $(x_{\delta+1}, x_{\delta+2}, \ldots, x_{\delta+\ell})$. Each read produces a value corresponding to the values of the ℓ characters. In our case, for simplification, we concentrate on the cases where the emitting value is the sum of all these ℓ characters. Distinct sequences may produce identical outputs under the (ℓ, δ)-read channel. Therefore, our focus in this work is to study the capacity of the (ℓ, δ)-read channel, denoted by $\mathsf{cap}(\ell, \delta)$, and is defined as the logarithmic ratio between the number of outputs and inputs of the (ℓ, δ)-read channel.

Several works have already studied this and similar models, focused on finding both the capacity [9], [10], and error-correcting codes [9]–[13] for the channel. In the subject of finding the capacity, both [9], [10] introduce algorithms for computing the capacity value with fixed parameters. In particular, [9] focused on the cases where $\delta = 1$, and the output of each read is a general function dependent on the ℓ characters. On the other hand, [10] focused on the case where the read function is the Hamming weight for any ℓ and δ, which is the (ℓ, δ)-read channel studied in this paper. More specifically, [10] solved $\mathsf{cap}(\ell, \delta)$ for the following cases: 1) $\ell \le \delta$, 2) ℓ is a multiple of δ, and 3) $\delta = 2$ and $\ell = 3, 5, 7$, using an algorithm that can be generalized for other values ℓ. The main goal of this paper is to determine the capacity value $\mathsf{cap}(\ell, \delta)$ for more parameters of ℓ and δ.

The rest of this paper is organized as follows. In Section II, we introduce the definitions describing the model. Section III is dedicated to solve the capacity $\mathsf{cap}(\ell, \delta)$ for $\delta < \ell < 2\delta$, where it is shown that $\mathsf{cap}(\ell, \delta) = \frac{1}{\delta}\log_2 \frac{1}{2}(\ell+1+\sqrt{(\ell+1)^2-4(\ell-\delta)(\ell-\delta+1)})$. Furthermore, when $\ell > \delta$, in Section IV, it is shown that $\mathsf{cap}(\ell, \delta) \le \frac{1}{\delta}\log_2 \frac{1}{2}(m+\sqrt{m^2+4m})$, where $m = (\ell \bmod \delta)((-\ell) \bmod \delta) + \delta$. Lastly, in Section V, we extend the model to the two-dimensional case and present several results on the capacity as well. Due to the lack of space, we omit some of the proofs and plots in the paper and they appear in the long version of this paper [14].

II. Definitions and Preliminaries

Let Σ_2 denote the binary alphabet $\{0, 1\}$. For every vector $\boldsymbol{x} \in \Sigma_2^n$, we refer to its sub-vector $(x_i, x_{i+1}, \ldots, x_{i+\ell-1})$, where $1 \le i \le n - \ell$, as $\boldsymbol{x}_{[i,\ell]}$. The *Hamming weight* of a vector $\boldsymbol{x}$ is denoted by $\mathrm{wt}(\boldsymbol{x})$.

Definition 1. The (ℓ, δ)-***read vector*** of $\boldsymbol{x} \in \Sigma_2^n$ is denoted by,

$$R_{\ell,\delta}(\boldsymbol{x}) \triangleq (\mathrm{wt}(\boldsymbol{x}_{[1,\ell]}), \mathrm{wt}(\boldsymbol{x}_{[\delta+1,\ell]}), \ldots, \mathrm{wt}(\boldsymbol{x}_{[t\cdot\delta+1,\ell]}))$$

The research was funded by the European Union (ERC, DNAStorage, 865630). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This work was also supported in part by NSF Grant CCF2212437.

979-8-3503-8284-6/24/$31.00 ©2024 IEEE

where $t = \frac{n-\ell}{\delta}$. For simplicity, we assume that $\delta | n - \ell$. For each binary vector $\boldsymbol{x}$, the (ℓ, δ)-***read channel*** produces the (ℓ, δ)-read vector of $\boldsymbol{x}$.

Example 1. Let $\boldsymbol{x} = (0,0,1,0,1,0,1,1,0,0,0,0)$, the $(4,2)$-read vector of $\boldsymbol{x}$ is $R_{4,2}(\boldsymbol{x}) = (1,2,3,2,0)$. We can notice that there exist other vectors, such as $\boldsymbol{y} = (0,0,0,1,0,1,1,0,0,0,0)$, $\boldsymbol{y} \neq \boldsymbol{x}$ that have the same read vector, i.e., $R_{4,2}(\boldsymbol{y}) = R_{4,2}(\boldsymbol{x})$.

Hence, a notable issue is that the read channel might have the same output for multiple inputs. As a result, our main focus will be on assessing and describing this reduction. To achieve this, we will establish the following definitions:

Definition 2. A code $\mathcal{C} \subseteq \Sigma_2^n$ is called an (ℓ, δ)-***read code*** if for all distinct $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ it holds that $R_{\ell,\delta}(\boldsymbol{x}) \neq R_{\ell,\delta}(\boldsymbol{y})$. The largest size of any (ℓ, δ)-read code of length n is denoted by $A(n, \ell, \delta)$. The ***capacity*** of the (ℓ, δ)-read channel is given by:

$$\mathsf{cap}(\ell, \delta) \triangleq \limsup_{n \to \infty} \frac{\log_2 A(n, \ell, \delta)}{n}.$$

A straightforward example of the (ℓ, δ)-read channel, occurs when $\delta = 1$, as explored in [10]. In this case, all distinct vectors, $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^n$, have distinct $(\ell, 1)$-read vector, for every ℓ and n. Consequently, $A(n, \ell, 1) = 2^n$ and the capacity of the $(\ell, 1)$-read channel is given by $\mathsf{cap}(\ell, 1) = \limsup_{n\to\infty} \frac{1}{n} \log_2 2^n = 1$. For the rest of the paper, it is assumed that $\delta > 1$.

The binary model described here can be extended to the q-ary case. Within this model, the L_1 *weight* is defined as the sum of all the entries in the vector and is also denoted by $\text{wt}(\boldsymbol{x})$. The definitions remain the same as in the binary model, and here we refer to this channel as the $(\ell, \delta)_q$-read channel for the q-ary alphabet, and we let $\mathsf{cap}_q(\ell, \delta)$ denote its capacity. The capacity value in the q-ary model can be directly deduced from the capacity in the binary case as is proved next in Theorem 1. Consequently, our focus in this work is directed towards the binary case.

Theorem 1. Let ℓ, δ be integers, for every integer $q \geq 2$ it holds that,

$$\mathsf{cap}_q(\ell, \delta) = \frac{q-1}{\log_2 q} \cdot \mathsf{cap}_2((q-1) \cdot \ell, (q-1) \cdot \delta).$$

This and similar models have been studied in several works, focusing on exploring the capacity [9], [10], which focuses on finding expressions and bounds for the capacity. Additionally, there is a concerted focus on finding error-correcting codes for the channel [9]–[11] which are mainly focused on finding constructions and bounds on the size of the code in the cases where there is one deletion. Our primary focus is on investigating the capacity of the (ℓ, δ)-read channel across various parameters. Multiple parameters of the (ℓ, δ)-read channel have already been studied in [10]. First, explicit expressions and bounds for the capacity within the following parameters have been revealed.

Theorem 2 ([10])**.** Let ℓ, δ be positive integers,

1) For $\ell \leq \delta$, $\mathsf{cap}(\ell, \delta) = \frac{1}{\delta} \log_2(\ell + 1)$.
2) If ℓ is a multiple of δ, then $\mathsf{cap}(\ell, \delta) = \frac{1}{\delta} \log_2(\delta + 1)$.
3) For $1 < \delta < \ell$, $\mathsf{cap}(\ell, \delta) \geq \frac{1}{\delta} \log_2(\delta + 1)$.

Second, the capacity value of the following cases, where $\delta = 2$ and $\delta < \ell$ have been calculated:

	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$	$\ell = 8$
$\delta = 2$	0.8857	0.7958	0.9258	0.7925	0.9361	0.7925

In this work, we present an explicit expression for $\mathsf{cap}(\ell, \delta)$, where $\delta < \ell < 2\delta$. Additionally, we establish an upper bound on the capacity for the rest of the cases. Table I presents the current results on $\mathsf{cap}(\ell, \delta)$, with entries in bold indicating new results derived from this work.

TABLE I

$\ell, \delta > 1$		$\mathsf{cap}(\ell, \delta)$
$\ell \leq \delta$		$\mathsf{cap}(\ell, \delta) = \frac{1}{\delta} \log_2(\ell + 1)$
$\delta < \ell < 2\delta$		$\mathsf{cap}(\ell, \delta) = \boldsymbol{\frac{1}{\delta} \log_2 f_1(\ell, \delta)}$
$\ell \geq 2\delta$	$\delta \mid \ell$	$\mathsf{cap}(\ell, \delta) = \frac{1}{\delta} \log_2(\delta + 1)$
	$\delta \nmid \ell$	$\frac{1}{\delta} \log_2(\delta + 1) \leq \mathsf{cap}(\ell, \delta) \leq \boldsymbol{\frac{1}{\delta} \log_2 f_2(\ell, \delta)}$

$f_1(\ell, \delta) = 0.5(\ell + 1 + \sqrt{(\ell+1)^2 - 4(\ell - \delta)(\ell - \delta + 1)})$,
$f_2(\ell, \delta) = 0.5(m - 1 + \sqrt{(m-1)^2 - 4(m-1)})$,
and $m = (\ell \bmod \delta + 1)(((-\ell) \bmod \delta + 1))$.

III. The Capacity for $\delta < \ell < 2\delta$

In this section, we study the value of $\mathsf{cap}(\ell, \delta)$ when $\delta < \ell < 2\delta$. First, we observe that the (ℓ, δ)-read channel is a regular language and therefore can be recognized by a non-deterministic transition state diagram. In the next definition, we present such a diagram for any $\ell \geq \delta$. This diagram was proposed in [10] and we present it here with an explanation of its correctness for the completeness of the results in the paper. All edges in such a graph are labeled. Thus, we refer to any directed edge (u, v) with a label α as $u \overset{\alpha}{\to} v$.

Definition 3. The graph $\mathcal{G}(\ell, \delta)$ is defined as follows.

- The nodes in $\mathcal{G}(\ell, \delta)$ are the set of all vectors $\boldsymbol{s}$ of length $\ell - \delta$, i.e., $V(\mathcal{G}(\ell, \delta)) = \{\boldsymbol{s} : \boldsymbol{s} \in \Sigma_2^{\ell-\delta}\}$.
- The set of directed labeled edges in $\mathcal{G}(\ell, \delta)$ is defined as

  $$E(\mathcal{G}(\ell,\delta)) = \{\boldsymbol{x}_{[1,\ell-\delta]} \overset{\alpha}{\to} \boldsymbol{x}_{[\delta+1,\ell-\delta]} : \boldsymbol{x} \in \Sigma_2^\ell, \alpha = \text{wt}(\boldsymbol{x})\}.$$

  That is, an edge between the nodes $\boldsymbol{u}$ and $\boldsymbol{v}$ with label α exists if there is a vector $\boldsymbol{x} \in \Sigma_2^\ell$ such that, $\boldsymbol{u} = \boldsymbol{x}_{[1,\ell-\delta]}$, $\boldsymbol{v} = \boldsymbol{x}_{[\delta+1,\ell-\delta]}$ and $\text{wt}(\boldsymbol{x}) = \alpha$.

For every n, t, where $n = t \cdot \delta + \ell$, every vector $\boldsymbol{x} \in \Sigma_2^n$, has the following (unique) path $\boldsymbol{x}_{[1,\ell-\delta]} \overset{\alpha_0}{\to} \boldsymbol{x}_{[\delta+1,\ell-\delta]} \overset{\alpha_1}{\to} \boldsymbol{x}_{[2\delta+1,\ell-\delta]} \overset{\alpha_2}{\to} \cdots \overset{\alpha_t}{\to} \boldsymbol{x}_{[(t+1)\cdot\delta+1,\ell-\delta]}$ in $\mathcal{G}(\ell, \delta)$, such that $(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_t)$ is the (ℓ, δ)-read vector of $\boldsymbol{x}$. Therefore, there is a path in $\mathcal{G}(\ell, \delta)$ for every read vector. On the other hand, every path in $\mathcal{G}(\ell, \delta)$ can correspond to more than one vector $\boldsymbol{x}$, but to only one read vector. That is, all vectors $\boldsymbol{x}$ that generate this path have the same read vector, and thus, $\mathcal{G}(\ell, \delta)$ is a state diagram of the (ℓ, δ)-read channel. Note that there might be two distinct vectors $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_2^\ell$, such that $\boldsymbol{x}_{[1,\ell-\delta]} = \boldsymbol{y}_{[1,\ell-\delta]} = \boldsymbol{v}$ and $\boldsymbol{x}_{[\delta+1,\ell-\delta]} \neq \boldsymbol{y}_{[\delta+1,\ell-\delta]}$, with the same Hamming weight, denoted by α. Therefore, the edges $\boldsymbol{v} \overset{\alpha}{\to} \boldsymbol{x}_{[\delta+1,\ell-\delta]}$, and $\boldsymbol{v} \overset{\alpha}{\to} \boldsymbol{y}_{[\delta+1,\ell-\delta]}$ exist in $\mathcal{G}(\ell, \delta)$, and thus, $\mathcal{G}(\ell, \delta)$ is not necessarily deterministic. In fact, for the case where $\delta < \ell \leq 2\delta$, $\mathcal{G}(\ell,\delta)$ is a regular graph, and between any two nodes, there exists the same number of parallel edges, which is $2\delta - \ell + 1$.

Example 2. For $\delta = 3, \ell = 5$, the nodes set in $\mathcal{G}(\ell,\delta)$, as shown in Fig. 1, are the vectors $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. The vectors $(0,0,0,0,1)$ and $(0,0,1,0,0)$ are both in Σ_2^ℓ, and therefore, $(0,0) \xrightarrow{1} (0,1)$ and $(0,0) \xrightarrow{1} (0,0)$ are edges in $\mathcal{G}(\ell,\delta)$. In conclusion, $\mathcal{G}(\ell,\delta)$ is not deterministic.

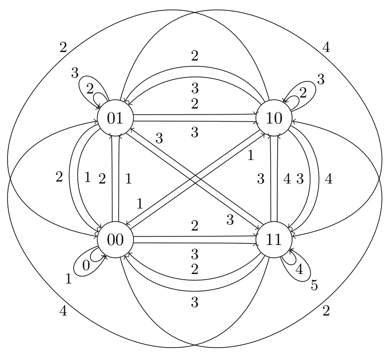

Fig. 1. The graph $\mathcal{G}(\ell,\delta)$ for $\ell = 5$ and $\delta = 3$.

A common approach to deriving the capacity from a non-deterministic state diagram is to convert the graph to a deterministic one. A known way to create a deterministic graph is the determinizing graph demonstrated in [15], where each vertex represents a subset of vertices in the non-deterministic graph. Here, we present a deterministic graph based on the determinizing graph with some changes in both the definition of the vertices and edges to enable analysis. To do so, we first introduce another useful definition.

Definition 4. For every subset V of $V(\mathcal{G}(\ell,\delta))$, and a label α. Let $\mathcal{E}(V,\alpha) \subseteq V(\mathcal{G}(\ell,\delta))$ be the set of all nodes with an incoming edge labeled α from any node in V, i.e.,

$$\mathcal{E}(V,\alpha) = \{\boldsymbol{u} : \exists \boldsymbol{v} \in V : \boldsymbol{v} \xrightarrow{\alpha} \boldsymbol{u} \in E(\mathcal{G}(\ell,\delta))\}.$$

Next, we introduce a deterministic graph of $\mathcal{G}(\ell,\delta)$.

Definition 5. The graph $\mathcal{H}(\ell,\delta)$ is defined as follows.

- The nodes of the graph $\mathcal{H}(\ell,\delta)$ are the set of all subsets of nodes in $\mathcal{G}(\ell,\delta)$, denoted as $V_{(a,b)}$, where the Hamming weights of the nodes are between a and b, $0 \leq a \leq b \leq \ell - \delta$, i.e.,
$$V(\mathcal{H}(\ell,\delta)) = \{V_{(a,b)} : 0 \leq a \leq b \leq \ell - \delta\},$$
where, $V_{(a,b)} = \{\boldsymbol{s} \in \Sigma_2^{\ell-\delta} : a \leq \text{wt}(\boldsymbol{s}) \leq b\}$.
- The set of directed labeled edges in $\mathcal{H}(\ell,\delta)$, denoted by $E(\mathcal{H}(\ell,\delta))$, is defined as
$$\{V_{(a_1,b_1)} \xrightarrow{\alpha} V_{(a_2,b_2)} : \mathcal{E}(V_{(a_1,b_1)},\alpha) \subseteq V_{(a_2,b_2)}, \exists \boldsymbol{u}^1,\boldsymbol{u}^2 \in \mathcal{E}(V_{(a_1,b_1)},\alpha), \text{wt}(\boldsymbol{u}^1)=a_2, \text{wt}(\boldsymbol{u}^2)=b_2\}.$$
That is, an edge between the nodes $V_{(a_1,b_1)}$ and $V_{(a_2,b_2)}$ with label α exists if, 1) all nodes of $\mathcal{G}(\ell,\delta)$ in $\mathcal{E}(V_{(a_1,b_1)},\alpha)$ belongs to $V_{(a_2,b_2)}$, and 2) there are nodes $\boldsymbol{u}^1, \boldsymbol{u}^2 \in \mathcal{E}(V_{(a_1,b_1)},\alpha)$ such that $\text{wt}(\boldsymbol{u}^1) = a_2$ and $\text{wt}(\boldsymbol{u}^2) = b_2$.

Example 3. For $\delta = 3, \ell = 5$, the graph $\mathcal{H}(\ell,\delta)$ of $\mathcal{G}(\ell,\delta)$ from Example 2, is shown in Fig. 2. The graph $\mathcal{H}(\ell,\delta)$ contains the edge $V_{(0,0)} \xrightarrow{1} V_{(0,1)}$, because, 1. $\mathcal{E}(V_{(0,0)},1) = \{(0,0),(0,1),(1,0)\} = V_{(0,1)}$, and thus, $\mathcal{E}(V_{(0,0)},1) \subseteq V_{(0,1)}$. 2. Both $(0,0) \xrightarrow{1} (0,0)$ and $(0,0) \xrightarrow{1} (0,1)$ are edges in $\mathcal{G}(\ell,\delta)$, while $\text{wt}(0,0) = 0$ and $\text{wt}(0,1) = 1$.

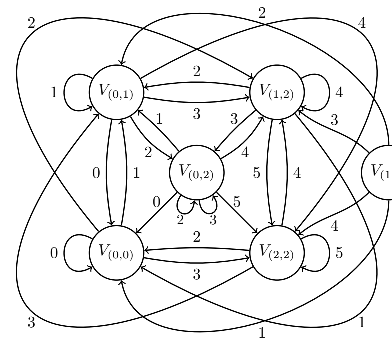

Fig. 2. The determinizing graph $\mathcal{H}(5,3)$ of $\mathcal{G}(5,3)$ in Fig. 1.

Next, we prove that $\mathcal{H}(\ell,\delta)$ is a determinizing graph of $\mathcal{G}(\ell,\delta)$.

Claim 1. The graph $\mathcal{H}(\ell,\delta)$ is a deterministic graph of $\mathcal{G}(\ell,\delta)$.

We have established that $\mathcal{H}(\ell,\delta)$ is a deterministic, finite state transition diagram of the regular language of the (ℓ,δ)-read channel. To find the adjacency matrix of $\mathcal{H}(\ell,\delta)$, we start by determining the number of edges from $V_{(a_1,b_1)}$ to $V_{(a_2,b_2)}$.

Claim 2.
1) If $a_2 = 0$ and $b_2 = \ell - \delta$, then the number of edges is $\max\{0, 3\delta - 2\ell + b_1 - a_1 + 1\}$.
2) If $a_2 = 0$ and $b_2 < \ell - \delta$, then the number of edges is 1 when $b_2 \leq b_1 - a_1 + 2\delta - \ell$ and otherwise 0.
3) If $a_2 > 0$ and $b_2 = \ell - \delta$, then the number of edges is 1 if $a_2 \geq 2\ell - 3\delta - b_1 + a_1$ and otherwise 0.
4) Otherwise, there is an edge only if $a_1 + b_2 = b_1 + a_2 + 2\delta - \ell$.

As observed in Example 3, the in or out degree of some nodes in $\mathcal{H}(\ell,\delta)$ might be zero. These nodes, as known, do not influence the value of the capacity and can thus be excluded from the graph. We can see by Claim 2, that the in and out degree of all the nodes $V_{(a,b)}$ where $a = 0$ or $b = \ell - \delta$ is at least one. We denote $\Lambda_{\ell,\delta}$ to be the set of all such nodes. The in-degree of all other remaining nodes is at least one if $b - a \geq 2\delta - \ell$, because in this case, there exist a_1 and b_1 such that $b_2 - a_2 = b_1 - a_1 + 2\delta - \ell$. All remaining nodes have an in-degree of zero. Let $\mathcal{H}^*(\ell,\delta)$ denote the graph $\mathcal{H}(\ell,\delta)$ without those nodes.

Example 4. For $\delta = 3, \ell = 5$, the graph $\mathcal{H}^*(\ell,\delta)$ of $\mathcal{H}(\ell,\delta)$ in Example 3, is the same graph as $\mathcal{H}(\ell,\delta)$ excluding the node $V_{(1,1)}$ which its in-degree is zero.

From now on, we focus on the graph $\mathcal{H}^*(\ell,\delta)$. We observe from Claim 2 that the number of edges between $V_{(a_1,b_1)}$ and $V_{(a_2,b_2)}$ is exclusively determined by whether $V_{(a_2,b_2)}$ is in $\Lambda_{\ell,\delta}$, as well as the values of $b_1 - a_1$ and $b_2 - a_2$. For every such node $V_{(a,b)}$, we define its *size* to be $b - a$. For every

$2\delta-\ell \leq d \leq \ell-\delta-1$, the number of nodes which are not in $\Lambda_{\ell,\delta}$ and are of size d, denoted by m_d, is the number of options for nodes of size d where $a \neq 0$ and $b \neq \ell-\delta$, i.e., $\ell-\delta-d-1$. Therefore, the total number of nodes in $\mathcal{H}^*(\ell,\delta)$, denoted by m, is $|\Lambda_{\ell,\delta}| + \sum_{d=2\delta-\ell}^{\ell-\delta-1} m_d = 1 + 2(\ell-\delta) + \binom{2\ell-3\delta}{2}$. Let $A_{\mathcal{H}^*}(\ell,\delta) \in \mathbb{N}^{m\times m}$ be the adjacency matrix of $\mathcal{H}^*(\ell,\delta)$ where its indices are ordered by the sizes of their nodes, while nodes with the same size are ordered lexicography. For shorthand, let $A_{i,j} \triangleq (A_{\mathcal{H}^*}(\ell,\delta))_{i,j}$. We denote $t:\{1,\ldots,m\}\rightarrow\{V_{(a,b)} : 0\leq a\leq b\leq \ell-\delta\}$ to be a mapping between an index in the matrix to its node, and let $d:\{1,\ldots,m\}\rightarrow\{0,\ldots,\ell-\delta\}$ be a mapping between an index in the matrix to the size of its node. For example, $t(2)=V_{(0,\ell-\delta-1)}$, and $d(2)=\ell-\delta-1$. From Claim 2, we get that for every $1\leq i\leq m$, $2\leq j\leq m$,

- $A_{i,1}=\max\{0, 3\delta-2\ell+d(i)+1\}$.
- If $t(j)\in\Lambda_{\ell,\delta}$, $d(j)\leq 2\delta-\ell+d(i)$, then $A_{i,j}=1$.
- If $t(j)\notin\Lambda_{\ell,\delta}$, $d(j)=2\delta-\ell+d(i)$, then $A_{i,j}=1$.
- Otherwise, $A_{i,j}=0$.

Example 5. For $\delta=3$, $\ell=5$, the adjacency matrix $A_{\mathcal{H}^*}(\ell,\delta)$ of $\mathcal{H}^*(\ell,\delta)$ from Example 4, is

$$A_{\mathcal{H}}(5,3) = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{matrix} - & (V_{(0,2)}) \\ - & (V_{(0,1)}) \\ - & (V_{(1,2)}) \\ - & (V_{(0,0)}) \\ - & (V_{(2,2)}) \end{matrix}.$$

To the right of the matrix, the values of $V_{(a,b)}$ indicate the values of $t(i)$ for each index i.

Claim 3. The characteristic polynomial of $A_{\mathcal{H}^*}(\ell,\delta)$, is

$$p_{\ell,\delta}(x) = (x^2-(\ell+1)x+(\ell-\delta)(\ell-\delta+1))x^{m-2}.$$

By Perron-Frobenius Theorem [15], we can establish the expression for the capacity.

Theorem 3. For every integers ℓ and δ, where $\delta<\ell<2\delta$ the capacity of the (ℓ,δ)-read channel is given by

$$\mathsf{cap}(\ell,\delta) = \frac{\log_2 \frac{\ell+1+\sqrt{(\ell+1)^2-4(\ell-\delta)(\ell-\delta+1)}}{2}}{\delta}.$$

IV. An Upper Bound on the Capacity for $\ell \geq 2\delta$

In this section, we propose an upper bound on the capacity for the rest of the cases where δ does not divide ℓ, and $\ell\geq 2\delta$. To accomplish this, we introduce a constrained system with a higher capacity than the one of the (ℓ,δ)-read channel. For the rest of this section, for every ℓ and δ, such that $\delta \nmid \ell$, $\ell\geq 2\delta$ and n,t integers such that $n=\delta t+\ell$. Let $a=\lfloor \frac{\ell}{\delta}\rfloor$, $b=\ell \bmod \delta$, and $d=\delta-b$, note that $\ell=a\delta+b$. We introduce the following claims and definitions.

Observation 1. For every $\boldsymbol{v}\in\Sigma_2^n$, and $0\leq i\leq t$, it holds that $\mathrm{wt}(\boldsymbol{v}_{[\delta\cdot i+1,\ell]})$ equals to

$$\mathrm{wt}(\boldsymbol{v}_{[\delta\cdot i+1,b]})+\sum_{j=1}^{a}\big(\mathrm{wt}(\boldsymbol{v}_{[\delta\cdot(i+j)-d+1,d]})+\mathrm{wt}(\boldsymbol{v}_{[\delta\cdot(i+j)+1,b]})\big).$$

First, for shorthand, let $\Delta(i)\triangleq \delta i-d+1$, for every $1\leq i\leq t+a$, and $\beta(i)\triangleq\delta i+1$ for every $0\leq i\leq t+a$. Next, as a result from Observation 1, the (ℓ,δ)-read vector depends only on the weights of the sub-vectors $\boldsymbol{v}_{[\beta(i),b]}$ and $\boldsymbol{v}_{[\Delta(j),d]}$, where $0\leq i\leq t+a$, $1\leq j\leq t+a$. Therefore, the sequence of ones and zeros within these sub-vectors has no impact on the value of the read vector. Consequently, to establish an upper bound on the number of read vectors, we focus exclusively on vectors where all zeros appear before the ones in all sub-vectors $\boldsymbol{v}_{[\beta(i),b]}$ and $\boldsymbol{v}_{[\Delta(i),d]}$. Thus, we concentrate only on vectors in

$$\Pi_{\ell,\delta}^n \triangleq \Pi_b\times\Pi_d\times\cdots\times\Pi_b\subseteq\Sigma_2^n,$$

where $\Pi_m \triangleq \{0^m1^{m-\alpha} : 0\leq\alpha\leq m\}$. Note that $|\Pi_{\ell,\delta}^n|\geq A(n,\ell,\delta)$, and thus, $\Pi_{\ell,\delta}$ provides an upper bound on $\mathsf{cap}(\ell,\delta)$. To find a tighter bound, we introduce the following lemma.

Lemma 4. For every $\boldsymbol{v},\boldsymbol{u}\in\Pi_{\ell,\delta}^n$ and $1\leq i\leq t-1$, if

$$\mathrm{wt}(\boldsymbol{v}_{[\Delta(i),\delta]})=\mathrm{wt}(\boldsymbol{u}_{[\Delta(i),\delta]}),\quad \mathrm{wt}(\boldsymbol{v}_{[\beta(i+a),\delta]})=\mathrm{wt}(\boldsymbol{u}_{[\beta(i+a),\delta]}),$$
$$\mathrm{wt}(\boldsymbol{v}_{[\beta(i),b]})+\mathrm{wt}(\boldsymbol{v}_{[\beta(i+a),b]})=\mathrm{wt}(\boldsymbol{u}_{[\beta(i),b]})+\mathrm{wt}(\boldsymbol{u}_{[\beta(i+a),b]}),$$

while all other sub-vectors are equal, then $R_{\ell,\delta}(\boldsymbol{v})=R_{\ell,\delta}(\boldsymbol{u})$.

Thus, we introduce the following mapping designed to maintain the value of the (ℓ,δ)-read vector.

Definition 6. Let $\phi_{\ell,\delta}^n:\Pi_{\ell,\delta}^n\rightarrow\Pi_{\ell,\delta}^n$ be a function that changes the value of any $\boldsymbol{v}$ according to the following steps: For every $i=1,\ldots,t-1$, if there exists $\boldsymbol{u}\in\Sigma_2^{\ell-2b}$ such that $\boldsymbol{v}_{[\Delta(i),\ell+2d]}=1^d0^b\boldsymbol{u}1^b0^d$, then

$$\phi_{\ell,\delta}(\boldsymbol{v})_{[\Delta(i),\ell+2d]}=01^{d-1}0^{b-1}1\boldsymbol{u}01^{b-1}0^{d-1}1.$$

Example 6. For $\ell=8$, $\delta=3$, and $n=14$, we have that, $a=2$, $b=2$, $d=1$. Let $\boldsymbol{v}=(\mathbf{0},\mathbf{1},1,\mathbf{0},\mathbf{0},0,\mathbf{0},\mathbf{1},1,\mathbf{1},\mathbf{1},0,\mathbf{1},\mathbf{1})$, the bold values in $\boldsymbol{v}$ are the sub-vectors $\boldsymbol{v}_{[\beta(i),b]}$, while the rest are the sub-vectors $\boldsymbol{v}_{[\Delta(i),d]}$. For $i=1$, $\Delta(i)=3$, $\beta(i)=4$, $\beta(i+a)=10$, and $\Delta(i+a+1)=12$, and we have that $\boldsymbol{v}_{[\Delta(i),\ell+2d]}=100\boldsymbol{u}110=1^d0^b\boldsymbol{u}1^b0^d$. Thus, $\phi_{\ell,\delta}^n(\boldsymbol{v})_{[\Delta(i),\ell+2d]}=01^{d-1}0^{b-1}1\boldsymbol{u}01^{b-1}0^{d-1}0$ and

$$\phi_{\ell,\delta}^n(\boldsymbol{v})=(\mathbf{0},\mathbf{1},1,0,\mathbf{0},\mathbf{1},0,\mathbf{0},\mathbf{1},1,\mathbf{0},\mathbf{1},1,\mathbf{1},\mathbf{1}).$$

As we can see $R_{\ell,\delta}(\phi_{\ell,\delta}^n(\boldsymbol{v}))=R_{\ell,\delta}(\boldsymbol{v})=(3,4,6)$.

Note that, every change of the function $\phi_{\ell,\delta}^n$, maintains the values of $\mathrm{wt}(\boldsymbol{v}_{[\Delta(i),\delta]})$, $\mathrm{wt}(\boldsymbol{v}_{[\beta(i+a),\delta]})$, and $\mathrm{wt}(\boldsymbol{v}_{[\beta(i),b]})+\mathrm{wt}(\boldsymbol{v}_{[\beta(i+a),b]})$, while all other sub-vectors are not changed. Thus, by Lemma 4, we have that $R_{\ell,\delta}(\boldsymbol{v})=R_{\ell,\delta}(\phi_{\ell,\delta}^n(\boldsymbol{v}))$. In addition, $\phi_{\ell,\delta}^n$ ensures that there are no sub-vectors $\boldsymbol{v}_{[\Delta(i),\ell+2d]}=1^d0^b\boldsymbol{u}1^b0^d$. Thus, we construct the following code from $\phi_{\ell,\delta}^n$.

Definition 7. The code $\mathcal{C}_{\ell,\delta}(n)$ is denoted to be the co-domain of $\phi_{\ell,\delta}^n(\boldsymbol{v})$,

$$\mathcal{C}_{\ell,\delta}(n)\triangleq\{\phi_{\ell,\delta}^n(\boldsymbol{v}):\boldsymbol{v}\in\Pi_{\ell,\delta}^n\}.$$

Using the last observation on $\phi_{\ell,\delta}^n$, we can derive an upper bound on the value of $A(n,\ell,\delta)$.

Lemma 5. For every $n\in\mathbb{N}$, $A(n,\ell,\delta)\leq|\mathcal{C}_{\ell,\delta}(n)|$.

In Lemma 5, we establish an upper bound for $A(n,\ell,\delta)$, however, determining the exact values of the bounds remains challenging. To address this, we introduce a more relaxed bound through the following constraint.

Definition 8. Let $\mathcal{L}_{b,\delta}$ be the following constraint. First, the vector must be in $\Pi^n_{\ell,\delta}$, and second, every sub-vector of length 2δ, that starts in an index of the form of $\Delta(i)$, is not in the form of $1^d0^b1^d0^b$, i.e., $\mathcal{L}_{b,\delta} = \{\boldsymbol{v} : \forall i, \boldsymbol{v}_{[\Delta(i),2\delta]} \neq 1^d0^b1^d0^b\}$. Let $A(n, \mathcal{L}_{b,\delta})$, be the number of length-n vectors that satisfy the constraint. The capacity of the constraint is denoted by $\mathsf{cap}(\mathcal{L}_{b,\delta})$, i.e., $\mathsf{cap}(\mathcal{L}_{\ell,\delta}) = \lim_{n\to\infty} \frac{1}{n}\log_2 A(n, \mathcal{L}_{b,\delta})$.

Let $\mathcal{W}_{b,\delta}(n) \subseteq \Pi^n_{\ell,\delta}$ be a code that satisfies the $\mathcal{L}_{b,\delta}$ constraint such that $|\mathcal{W}_{b,\delta}(n)| = A(n, \mathcal{L}_{b,\delta})$. To show that $\mathsf{cap}(\ell,\delta) \leq \mathsf{cap}(\mathcal{L}_{b,\delta})$, we first define the following function.

Definition 9. Let $g : \mathcal{C}_{\ell,\delta}(n + a\delta) \to \{\boldsymbol{u} \in \Pi^{n+a\delta}_{\ell,\delta} : \boldsymbol{u}_{[1,n]} \in \mathcal{W}_{b,\delta}(n)\}$ be a function, such that for each $\boldsymbol{v}$, $g(\boldsymbol{v})$ is the value of $\boldsymbol{v}$ after the following steps:
For $i = 1, 2, \ldots, t + a - 1$, if $\boldsymbol{v}_{[\Delta(i),2\delta]} = 1^d0^b1^d0^b$, then

$$\boldsymbol{v}_{[\beta(i+a),\delta]} = 1^b0^d, \boldsymbol{v}_{[\Delta(i+1),d]} = 0^\gamma 1^{d-\gamma}, \boldsymbol{v}_{[\beta(i),b]} = 0^\alpha 1^{b-\alpha}$$

where $\alpha = \mathrm{wt}(\boldsymbol{v}_{[\beta(i+a),b]})$, and $\gamma = \mathrm{wt}(\boldsymbol{v}_{[\Delta(i+a+1),d]})$. The changes are well defined since $a \geq 2$, ensuring that there is no overlap between the changed sub-vectors. Note that, after every step, $\boldsymbol{v}_{[\Delta(i),2\delta]} \neq 1^d0^b1^d0^b$, and if $\boldsymbol{\nu}_{[\Delta(i),2\delta]}$ was equal to $1^d0^b1^d0^b$ then $\boldsymbol{v}_{[\Delta(i),\ell+2d]} \neq 1^d0^b\boldsymbol{u}1^b0^d$.

Example 7. For the parameters as in Example 6 and $n = 8$, let $\boldsymbol{v} = (\mathbf{0}, \mathbf{1}, 1, \mathbf{0}, \mathbf{0}, 1, \mathbf{0}, \mathbf{0}, 1, \mathbf{0}, \mathbf{1}, 1, \mathbf{1}, \mathbf{1})$. We can see that there is no i such that $\boldsymbol{v}_{[\Delta(i),\ell+2d]} = 1^d0^b\boldsymbol{u}1^b0^d$, and thus $\boldsymbol{v} \in \mathcal{C}_{\ell,\delta}(n+a\delta)$. We can see that for $i = 1$, $\boldsymbol{v}_{[\Delta(i),2\delta]} = 1^d0^b1^d0^b$, and in addition, $\boldsymbol{v}_{[\beta(i+a),\delta]} = 011$. Therefore, by applying g on $\boldsymbol{v}$, we get that after the first iteration $\boldsymbol{v}_{[\Delta(i+1),\delta]} = 001$ and $\boldsymbol{v}_{[\beta(i+a),\delta]} = 110$, i.e.,

$$g(\boldsymbol{v}) = (\mathbf{0}, \mathbf{1}, 1, \mathbf{0}, \mathbf{0}, 0, \mathbf{0}, \mathbf{1}, 1, \mathbf{1}, \mathbf{1}, 0, \mathbf{1}, \mathbf{1}).$$

We notice that now, $g(\boldsymbol{v})_{[\Delta(i),2\delta]} \neq 1^d0^b1^d0^b$, while $g(\boldsymbol{v})_{[\Delta(i),\ell+2d]} = 1^d0^b\boldsymbol{u}1^b0^d$, and thus, $g(\boldsymbol{v})$ is in $\mathcal{W}_{b,\delta}(n)$ and not in $C_{\ell,\delta}(n + a\delta)$.

Claim 4. For every $\boldsymbol{v} \in \mathcal{C}_{\ell,\delta}(n + a\delta)$, $g(\boldsymbol{v})_{[1,n]} \in \mathcal{W}_{b,\delta}(n)$.

Claim 5. The function g is injective.

From Claims 4 and 5, $|\{\boldsymbol{v} \in \Pi^{n+a\delta}_{\ell,\delta} : \boldsymbol{v}_{[1,n]} \in \mathcal{W}_{b,\delta}(n)\}| \geq |\mathcal{C}_{\ell,\delta}(n + a\delta)|$, leading to the following upper bound.

Theorem 6. For every ℓ, δ, such that $2\delta < \ell$ and $\delta \not| \ell$, $\mathsf{cap}(\mathcal{L}_{b,\delta})$ is an upper bound of the capacity of the (ℓ,δ)-read channel, i.e., $\mathsf{cap}(\ell,\delta) \leq \mathsf{cap}(\mathcal{L}_{b,\delta})$, and

$$\mathsf{cap}(\mathcal{L}_{b,\delta}) = \frac{\log_2 \frac{m-1+\sqrt{(m-1)^2+4(m-1)}}{2}}{\delta},$$

where $m = (b + 1)(d + 1)$.

We conjecture that for every δ and $1 \leq b \leq \delta$, the capacity value $\mathsf{cap}(\delta \cdot a + b, \delta)$ increases with a. Under this conjecture, the limit $L(\delta, b) = \lim_{a\to\infty} \mathsf{cap}(\delta \cdot a + b, \delta)$ is exists. From Theorem 6 we get that $L(\delta, b) \leq \mathsf{cap}(\mathcal{L}_{b,\delta})$. For example, as it can be observed from Table II when $\delta = 2$ and ℓ is odd the capacity is increasing as ℓ increases. Yet, this conjecture has not been proven and we hope to explore it in future works.

TABLE II

ℓ,δ	$3,2$	$5,2$	$7,2$	$9,2$
$\mathsf{cap}(\ell,\delta)$	0.8857	0.9257	0.9361	0.9399

V. The Two-dimensional Weighted Channel

Coding for multiple dimensions, particularly two-dimensional storage systems, has gained significant attention in recent years due to its potential applications in diverse fields. This is mainly due to the unique properties of the information which can be more accurately described in multiple dimensions. Examples of such coding schemes were explored in [16], [17], [18], and [19]. Thus, we are interested in extending the read channel to its two-dimensional version. The two-dimensional read channel is defined as follows.

For a matrix $B \in \Sigma_q^{n_1\times n_2}$, let $B_{[k_1,\ell_1;k_2,\ell_2]}$ be the $\ell_1 \times \ell_2$ sub-matrix of B with entries between rows k_1 and $k_1 + \ell_1 - 1$, and columns between k_2 and $k_2 + \ell_2 - 1$. The *weight* of $B_{[k_1,\ell_1;k_2,\ell_2]}$ is denoted by $\mathrm{wt}(B_{[k_1,\ell_1;k_2,\ell_2]})$ and is defined as the sum of entries in the window.

Definition 10. The $((\ell_1,\ell_2),(\delta_1,\delta_2))$-***read matrix***, $R^{\ell_1,\ell_2}_{\delta_1,\delta_2}(B)$, of B is a $(t_1 + 1) \times (t_2 + 1)$ matrix, where $t_1 = \frac{n_1-\ell_1}{\delta_1}$ and $t_2 = \frac{n_2-\ell_2}{\delta_2}$, and its (i,j)-th entry is defined by $\mathrm{wt}(B_{[k_1,\ell_1;k_2,\ell_2]})$, where $k_1 = \delta_1 i + 1$, $k_2 = \delta_2 j + 1$, $0 \leq i \leq t_1$, and $0 \leq j \leq t_2$.

Definition 11. A code $\mathbb{C} \subseteq \Sigma_q^{n_1\times n_2}$ is called an $((\ell_1,\ell_2),(\delta_1,\delta_2))_q$-***read code*** if for all distinct $B, D \in \mathbb{C}$ we have that $R^{\ell_1,\ell_2}_{\delta_1,\delta_2}(B) \neq R^{\ell_1,\ell_2}_{\delta_1,\delta_2}(D)$. The largest size of any $n_1 \times n_2$ $((\ell_1,\ell_2),(\delta_1,\delta_2))_q$-read code is denoted by $A_q(n_1, n_2, (\ell_1,\ell_2),(\delta_1,\delta_2))$. The ***capacity*** of the $((\ell_1,\ell_2),(\delta_1,\delta_2))_q$-read channel denoted by $\mathsf{cap}_q((\ell_1,\ell_2),(\delta_1,\delta_2))$ is defined by:

$$\limsup_{n,m\to\infty} \frac{\log_q A_q(n_1, n_2, (\ell_1,\ell_2),(\delta_1,\delta_2))}{n_1 \cdot n_2}.$$

Theorem 7. For positive integers δ_1, δ_2, ℓ_1, ℓ_2, we have that for every two integers $q_1, q_2 \geq 1$ such that $q = q_1q_2 + 1$, the capacity of the $((\ell_1,\ell_2),(\delta_1,\delta_2))_q$-read channel, equals

$$\frac{q_1 \cdot q_2}{\log_2(q_1 \cdot q_2 + 1)} \cdot \mathsf{cap}_2((q_1 \cdot \ell_1, q_2 \cdot \ell_2),(q_1 \cdot \delta_1, q_2 \cdot \delta_2)).$$

Theorem 7 implies that all the results in the two-dimensional binary model can be generalized to any q-ary model as well. Furthermore, we demonstrate in the next theorem the connection between the capacity of the two-dimensional read channel and that of the one-dimensional channel.

Theorem 8. For positive integers $q, \delta, \ell_1, \ell_2$, we have that

1) If $\delta_2 \geq \ell_2$
$$\mathsf{cap}_q((\ell_1,\ell_2),(\delta_1,\delta_2)) = \frac{\ell_2}{\delta_2}\mathsf{cap}_q(\ell_2 \cdot \ell_1, \ell_2 \cdot \delta_1).$$
2) If δ_2 divides ℓ_2
$$\mathsf{cap}_q((\ell_1,\ell_2),(\delta_1,\delta_2)) = \mathsf{cap}_q(\delta_2 \cdot \ell_1, \delta_2 \cdot \delta_1).$$

In addition, for all cases where $\delta_1 < \ell_1$

$$\mathsf{cap}_q(\delta_1 \cdot \ell_2, \delta_1 \cdot \delta_2) \leq \mathsf{cap}_q((\ell_1,\ell_2),(\delta_1,\delta_2)) \leq \mathsf{cap}_q(\ell_2,\delta_2).$$

REFERENCES

[1] P. L. Antkowiak, J. Lietard, M. Z. Darestani, M. M. Somoza, W. J. Stark, R. Heckel, and R. N. Grass, "Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction," *Nature communications*, vol. 11, no. 1, p. 5345, 2020.

[2] S. Palluk *et al.*, "De novo DNA synthesis using polymerase-nucleotide conjugates," *Nature biotechnology*, vol. 36, no. 7, pp. 645–650, 2018.

[3] H. H. Lee, R. Kalhor, N. Goela, J. Bolot, and G. M. Church, "Terminator-free template-independent enzymatic DNA synthesis for digital information storage," *Nature communications*, vol. 10, no. 1, p. 2383, 2019.

[4] M. Jain, H. E. Olsen, B. Paten, and M. Akeson, "The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community," *Genome biology*, vol. 17, pp. 1–11, 2016.

[5] D. Deamer, M. Akeson, and D. Branton, "Three decades of nanopore sequencing," *Nature biotechnology*, vol. 34, no. 5, pp. 518–524, 2016.

[6] A. H. Laszlo, Derrington *et al.*, "Decoding long nanopore sequencing reads of natural DNA," *Nature biotechnology*, vol. 32, no. 8, pp. 829–833, 2014.

[7] J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, "Characterization of individual polynucleotide molecules using a membrane channel," *Proceedings of the National Academy of Sciences*, vol. 93, no. 24, pp. 13 770–13 773, 1996.

[8] W. Mao, S. N. Diggavi, and S. Kannan, "Models and information-theoretic bounds for nanopore sequencing," *2018 IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 3216–3236.

[9] R. Hulett, S. Chandak, and M. Wootters, "On coding for an abstracted nanopore channel for DNA storage," in *proceeding 2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 2465–2470.

[10] Y. M. Chee, A. Vardy, V. K. Vu, and E. Yaakobi, "Transverse-read-codes for domain wall memories," *IEEE Journal on Selected Areas in Information Theory*, vol. 4, pp. 784–793, 2023.

[11] A. Banerjee, Y. Yehezkeally, A. Wachter-Zeh, and E. Yaakobi, "Error-Correcting Codes for Nanopore Sequencing," *arXiv preprint arXiv:2305.10214*, 2023.

[12] Y. Sun and G. Ge, "Bounds and Constructions of ℓ-Read Codes under the Hamming Metric," *arXiv preprint arXiv:2403.11754*, 2024.

[13] A. Banerjee, Y. Yehezkeally, A. Wachter-Zeh, and E. Yaakobi, "Correcting a single deletion in reads from a nanopore sequencer," *arXiv preprint arXiv:2401.15939*, 2024.

[14] O. Yerushalmi, T. Etzion, and E. Yaakobi, "The Capacity of the Weighted Read Channel," *arXiv preprint arXiv:2401.15368*, 2024.

[15] B. H. Marcus, R. M. Roth, and P. H. Siegel, "An introduction to coding for constrained systems," *Lecture notes*, 2001.

[16] S. Marcovich and E. Yaakobi, "The zero cubes free and cubes unique multidimensional constraints," in *proceeding 2022 IEEE Information Theory Workshop (ITW)*, pp. 220–225.

[17] T. Etzion and E. Yaakobi, "Error-correction of multidimensional bursts," *proceeding 2009 IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 961–976.

[18] A. Sharov and R. M. Roth, "Two-dimensional constrained coding based on tiling," *proceeding 2010 IEEE transactions on information theory*, vol. 56, no. 4, pp. 1800–1807.

[19] I. Tal and R. M. Roth, "Bounds on the Rate of 2-D Bit-Stuffing Encoders," *2010 IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2561–2567.