Quantized-Constraint Concatenation and the Covering Radius of Constrained Systems

Dor Elimelech, *Student Member, IEEE*, Tom Meyerovitch, and Moshe Schwartz, *Fellow, IEEE*

***Abstract*— We introduce a novel framework for implementing error-correction in constrained systems. The main idea of our scheme, called Quantized-Constraint Concatenation (QCC), is to employ a process of embedding the codewords of an error-correcting code in a constrained system as a (noisy, non-invertible) quantization process. This is in contrast to traditional methods, such as concatenation and reverse concatenation, where the encoding into the constrained system is reversible. The possible number of channel errors QCC is capable of correcting is linear in the block length n, improving upon the $O(\sqrt{n})$ possible with the state-of-the-art known schemes. For a given constrained system, the performance of QCC depends on a new fundamental parameter of the constrained system – its covering radius. Motivated by QCC, we study the covering radius of constrained systems in both combinatorial and probabilistic settings. We reveal an intriguing characterization of the covering radius of a constrained system using ergodic theory. We use this equivalent characterization in order to establish efficiently computable upper bounds on the covering radius.**

***Index Terms*— Constrained systems, covering radius, error-correcting codes, Markov chains, sliding-block codes.**

I. Introduction

Constrained codes are often employed in communication and storage systems in order to mitigate the occurrence of data-dependent errors. In many channels, some words are more prone to error than others, and therefore by avoiding them, the number of errors is reduced. Such codes are called constrained codes. While the use of constrained codes may significantly reduce the occurrence of data-dependent errors, in many realistic scenarios, the transmitted data may still be corrupted by data-independent errors.

A well-known strategy for handling the corruption of data is to combine error-correcting codes with constrained codes. This has been extensively studied over the past 40 years (see for example [7], [11], [15], [18], [22], [27]), and has recently regained attention due to the increased interest in DNA storage systems. Over the last years, error-correcting constrained codes for DNA storage have been studied in numerous works [5], [8], [9], [24], [26], [32], [33], [34], [38], with particular attention given to the GC-content constraint and the run-length (homopolymer) constraint.

Despite the considerable recent progress made in the construction and analysis of error-correcting constrained codes for specific families of constraints, only a few general frameworks for implementing error correction in constrained systems are known (see [29, Ch. 8] for a survey). An important example of such a framework is the method of reverse concatenation, sometimes called modified concatenation (see [7], [15], [22], [27]), in which an error-correction encoding follows a constrained encoder. Recently, an improvement of the reverse-concatenation method called segmented reverse concatenation, was suggested [18]. A principal limitation of these methods is their error-correction capability. While the state-of-the-art method presented in [18] allows for a correction of $O(\sqrt{n})$ errors (where n is the block length), a general technique for correcting $\Theta(n)$ errors in constrained systems is unknown.

Motivated by this gap, we propose an alternative strategy, *quantized-constraint concatenation (QCC)*, for the implementation of error correction in constrained systems, which also works in the presence of $\Theta(n)$ errors. The basic idea behind our proposed method is simple: we suggest to consider the embedding process of information in the constrained media as a *quantization* process rather than a coding process. In traditional methods (including concatenation and reverse concatenation), a constrained word represents the data to be transmitted and protected against errors. Thus, the constrained encoder is reversible and incurs a rate penalty on top of the rate penalty for the error-correcting code. In QCC, we consider the constrained word as a corrupted version of the input message, obtained by a quantization procedure. Thus, the constrained quantizer incurs no rate penalty. Instead, the parameters of the error-correcting code are designed to handle both errors caused by the channel and by the quantization process.

Let $\mathcal{B}_n \subseteq \Sigma^n$ be some set of constrained words of length n over some finite alphabet Σ. Assume furthermore that $r < n$ is an integer such that for any word $\overline{y} \in \Sigma^n$ there exists a corresponding word $\overline{x} \in \mathcal{B}_n$ with Hamming distance $d(\overline{x}, \overline{y}) \leqslant r$. Given an error-correcting code $C \subseteq \Sigma^n$ that can correct $t > r$

Manuscript received 10 May 2023; revised 29 November 2023; accepted 8 December 2023. Date of publication 14 December 2023; date of current version 21 May 2024. This work was supported in part by the Israel Science Foundation under Grant 1052/18 and Grant 859/22. An earlier version of this paper was presented in part at the 2023 ISIT [DOI: 10.1109/ISIT54713.2023.10206533 and DOI: 10.1109/ISIT54713.2023.10206920]. *(Corresponding author: Dor Elimelech.)*

Dor Elimelech is with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva 8410501, Israel (e-mail: doreli@post.bgu.ac.il).

Tom Meyerovitch is with the Department of Mathematics, Ben-Gurion University of the Negev, Be'er Sheva 8410501, Israel (e-mail: mtom@bgu.ac.il).

Moshe Schwartz is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada, on leave from the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva 8410501, Israel (e-mail: schwartz.moshe@mcmaster.ca).

Communicated by M. Kovačević, Associate Editor for Coding and Decoding.

Digital Object Identifier 10.1109/TIT.2023.3343187

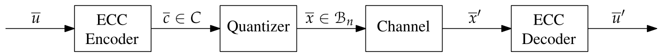

Fig. 1. A block diagram describing our proposed error correction procedure.

errors, we propose the following constrained error-correction procedure (see Figure 1):

- *Encoding:* Given an information word $\overline{u}$, use an encoder for an error-correcting code to map it to a codeword $\overline{c} \in C$.
- *Quantization:* Given $\overline{c} \in C$, find a constrained word $\overline{x} \in \mathcal{B}_n$ such that $d(\overline{c}, \overline{x}) \leqslant r$, and transmit $\overline{x}$.
- *Channel:* At the channel output, $\overline{x}' \in \Sigma^n$, a corrupted version of $\overline{x}$, is observed.
- *Decoding:* Use the decoder for C on $\overline{x}'$ and obtain $\overline{u}'$.

If the channel does not introduce more than $t-r$ errors, i.e., $d(\overline{x}, \overline{x}') \leqslant t-r$, then $d(\overline{c}, \overline{x}') \leqslant t$. Since C can correct t errors, we have $\overline{u} = \overline{u}'$. Namely, it is possible to correct $t-r$ channel errors. We are therefore interested in the minimal number r, such that any word in the space can be quantized to a word in $\mathcal{B}_n$ with at most r coordinates changed. In coding-theory terminology, this quantity, denoted by $R(\mathcal{B}_n)$, is called the *covering radius* of $\mathcal{B}_n$. Using this technique, it is now possible to correct $\Theta(n)$ errors: assume that we have a constrained system such that for all n we have $R(\mathcal{B}_n) \leqslant \rho \cdot n$, and $(C_n)_{n \in \mathbb{N}}$ is a sequence of codes capable of correcting $\delta \cdot n$ errors for some $\delta > \rho$. Using the scheme presented above, it is therefore possible to correct $(\delta - \rho) \cdot n$ channel errors, which is linear in n.

Certain ad-hoc coding strategies in the presence of constraints that use some sort of quantization have already been discussed in the literature in the context of balanced codes e.g. [17], [20], and [30]. However, not only are these examples limited to specific error-correcting codes or specific constraints, they also have a major difference with our work: the amount of quantization noise plays no role in these constructions.

To further understand our proposed scheme, we must study the covering radius of constrained systems, which is the goal of this paper. We outline the contributions we make. In Section III, we provide a combinatorial definition for the covering radius of a constrained system, and investigate some of its fundamental properties. We also observe an intriguing phenomenon: We present an example of a constrained system with positive capacity that has the same covering radius as the repetition code, which has zero capacity.

Inspired by this phenomenon, in Section IV we take a probabilistic approach and define the *essential* covering radius. We show that this version disregards the extreme cases causing the unwanted phenomenon described above. We use the framework of ergodic theory to give an alternative characterization of the essential covering radius. In Section V we use our ergodic-theoretic definition of the essential covering radius to establish upper bounds on the essential covering radius in typical scenarios. Using a Markov-chain approach, we derive a general upper bound that is efficiently computable as a solution of a linear program. We also provide bounds using sliding-block-code functions. We show that in the primitive case, these bounds asymptotically attain the essential covering radius.

II. Preliminaries

Throughout this paper, we shall use lower-case letters, x, to denote scalars and symbols, overlined lower-case letters, $\overline{x}$, to denote finite-length words, and bold lower-case letters, $\mathbf{x}$, to denote bi-infinite sequences. We use upper-case letters, X, for constrained systems. For a bi-infinite sequence $\mathbf{x} = \dots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \dots$ and $n \leqslant m$ we denote the subword $\mathbf{x}_n^m \triangleq \mathbf{x}_n, \dots, \mathbf{x}_m$ (and similarly $\overline{x}_n^m$ for finite words). We use Σ to denote a finite alphabet, and $[n] \triangleq \{0, 1, \dots, n-1\}$. We denote the string of i consecutive 0's (or 1's) by $\overline{0}^i$ (or $\overline{1}^i$, respectively). Whenever the length of the string is clear from the context, we may omit the subscript i from the notation.

The set of words of length n over Σ is denoted by Σ^n. If $\overline{u} \in \Sigma^n$, we shall index its letters by $[n]$, i.e., $\overline{u} = \overline{u}_0, \overline{u}_1, \dots, \overline{u}_{n-1}$. For any $\overline{v}, \overline{u} \in \Sigma^n$, we define the Hamming distance as

$$d(\overline{u}, \overline{v}) \triangleq |\{i \in [n] \ : \ \overline{u}_i \neq \overline{v}_i\}|.$$

The ball of radius r (with respect to the Hamming distance) centered in $\overline{x}$ is denoted by $\mathsf{Ball}(r, \overline{x})$. The covering radius of a code $C \subseteq \Sigma^n$ is the minimal integer r such that the union of balls of radius r, centered at the codewords of C, covers the whole space. That is,

$$R(C) \triangleq \min \left\{ r \in \mathbb{N} \cup \{0\} \ : \ \bigcup_{\overline{c} \in C} \mathsf{Ball}(r, \overline{c}) = \Sigma^n \right\}.$$

Elements in Σ^n whose distance to the closest codeword of C is $R(C)$, are often called *deep holes* (e.g., see [12, Definition 2.1.3]).

We turn to discuss constrained systems. These are often studied in the framework of symbolic dynamics (see for example [25], [29]). In a typical (one dimensional) setting we have a finite alphabet Σ, and the space of bi-infinite sequences of Σ, denoted $\Sigma^{\mathbb{Z}}$, is considered as a compact metrizable topological space, equipped with the product topology (where Σ has the discrete topology). The dynamics on the system $\Sigma^{\mathbb{Z}}$ are realized by the shift transformation, $T : \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$, defined by

$$(T\mathbf{x})_n \triangleq \mathbf{x}_{n+1},$$

which is a topological homeomorphism of the system. For a finite word $\overline{x} \in \Sigma^n$ we let $[\overline{x}]$ denote the cylinder set defined by $\overline{x}$, which is

$$[\overline{x}] \triangleq \left\{ \mathbf{x} \in \Sigma^{\mathbb{Z}} \ : \ \mathbf{x}_0^{n-1} = \overline{x} \right\}. \quad (1)$$

A subshift (or shift space) $X \subseteq \Sigma^{\mathbb{Z}}$ is a compact subspace, which is invariant under the shift transformation. For a subshift X, the language of X is the set of all finite words that appear as subwords of some element in X. That is

$$\mathcal{B}(X) \triangleq \left\{ \overline{x} = (x_0 \dots x_k) \ : \ \exists \mathbf{x} \in X, n \in \mathbb{Z} \text{ such that } \mathbf{x}_n^{n+k} = \overline{x}, k \in \mathbb{N} \cup \{0\} \right\}.$$

The set of words of length n in the language is denoted by $\mathcal{B}_n(X) \triangleq \mathcal{B}(X) \cap \Sigma^n$. The topological entropy, also called capacity, of X is defined to be the following limit (which exists by Fekete's Lemma)

$$h(X) \triangleq \lim_{n \to \infty} \frac{\log_{|\Sigma|} |\mathcal{B}_n(X)|}{n}.$$

In our setting, constrained systems are those shift spaces which can be realized by walks on some labeled graph.

Definition 1: A shift space $X \subseteq \Sigma^{\mathbb{Z}}$ is called a constrained system (or a sofic shift) if there exists a finite directed graph $G = (V, E)$ and a labeling function $L : E \to \Sigma$ such that

$$X = X_G \triangleq \left\{ (L(e_i))_{i\in\mathbb{Z}} \ : \ (e_i)_{i\in\mathbb{Z}} \text{ is a bi-infinite directed path in } G \right\}.$$

A labeled graph $G = (V, E, L)$ is called irreducible if any two vertices are connected by a directed path. An irreducible graph is called primitive if the greatest common divisor of all cycle lengths is 1. It is well known (e.g., see [25, Theorem 4.5.8]) that an irreducible graph is primitive if and only if there exists $n \in \mathbb{N}$ such that for any two vertices $v, v' \in V$ there exists a directed path of length n from v to v'.

Definition 2: A constrained system $X \subseteq \Sigma^{\mathbb{Z}}$ is called irreducible (respectively: primitive), if there exists an irreducible (respectively: primitive) labeled graph G such that $X = X_G$.

A special family of constrained systems of particular interest is the family of systems defined by a finite set of local constraints. These are referred to as *systems of finite type* and are formally defined as follows:

Definition 3: A constrained system $X \subseteq \Sigma^{\mathbb{Z}}$ is said to be a system of finite type (SFT) if there exists some $m \in \mathbb{N}$ and a finite set of forbidden words $\mathcal{F} \subseteq \Sigma^m$ such that X is the set of all bi-infinite sequences not containing any forbidden pattern from $\mathcal{F}$. That is

$$X = X_{\mathcal{F}} \triangleq \left\{ \mathbf{x} \in \Sigma^{\mathbb{Z}} \ : \ \forall n \in \mathbb{Z}, \mathbf{x}_n^{n+m-1} \notin \mathcal{F} \right\}.$$

We shall now formally describe the QCC scheme for a given constrained system X and block length n. Let C be an $[n, k, d]_q$ linear code, namely, a k-dimensional subspace of $\mathbb{F}_q^n$ with the property that the minimal Hamming distance between two distinct codewords in C is d. We define the components of the block diagram from Figure 1 as follows:

- A pair of functions (E, D) is called an error-correction encoding-decoding scheme for C if $E : \mathbb{F}_q^k \to C$ (the ECC-encoder) is an injection and $D : \mathbb{F}_q^n \to \mathbb{F}_q^k$ (the ECC-decoder) satisfies that for any $\overline{u} \in \mathbb{F}_q^k$ and $\overline{x} \in \mathbb{F}_q^n$ such that $d(E(\overline{u}), \overline{x}) \leqslant \lfloor \frac{d-1}{2} \rfloor \triangleq t$ we have $D(E(\overline{u})) = D(\overline{x}) = \overline{u}$.
- An (n, r)-quantizer for X is a function $Q : \mathbb{F}_q^n \to \mathcal{B}_n(X)$ such that for any $\overline{y} \in \mathbb{F}_q^n$ we have $d(\overline{y}, Q(\overline{y})) \leqslant r$.

In our coding procedure, the user-supplied information word $\overline{u} \in \mathbb{F}_q^k$ is encoded using the ECC-encoder to obtain $\overline{c} = E(\overline{u}) \in C$. Next, $\overline{c}$ is quantized to a constrained word $\overline{x} = Q(\overline{c}) \in \mathcal{B}_n(X)$, which is then transmitted. After passing through the channel, a possibly corrupted version of $\overline{x}$, denoted $\overline{x}' \in \mathbb{F}_q^n$, is then observed. We decode $\overline{x}'$ using the ECC-decoder and obtain $\overline{u}' = D(\overline{x}')$.

We claim that as long as the channel introduces no more then $t-r$ errors, i.e., $d(\overline{x}, \overline{x}') \leqslant t-r$, then we will successfully decode $\overline{u} = \overline{u}'$. Indeed, by the triangle inequality we have

$$d(\overline{c}, \overline{x}') \leqslant d(\overline{c}, \overline{x}) + d(\overline{x}, \overline{x}') \leqslant d(\overline{c}, Q(\overline{c})) + t - r \leqslant t,$$

where we used the fact that Q is an (n, r)-quantizer. Thus, $d(\overline{c}, \overline{x}') \leqslant t$ and therefore, by our assumptions on the encoding-decoding scheme, we conclude that $\overline{u} = D(\overline{c}) = D(\overline{x}')$, as desired.

By the definition of the covering radius of $\mathcal{B}_n(X)$, $r = R(\mathcal{B}_n(X))$, is the minimal number such that there exists an (n, r)-quantizer, and therefore, it bounds the error-correction capability of the QCC scheme. The coding-theoretic literature on the covering radius of error-correcting codes is quite extensive (e.g., see [12] and the many references within). However, as will become apparent, our setting is quite different since we are considering sequences of codes of finite block lengths, associated with constrained systems. As mentioned above, these constrained systems are mathematically formulated and studied using bi-infinite sequences. Thus, we shall borrow tools from ergodic theory to apply to the problem at hand.

III. The Covering Radius of a Constrained System

We begin with a definition of the covering radius of a set $B \subseteq \Sigma^n$ with respect to another set $A \subseteq \Sigma^n$ under the Hamming metric.

Definition 4: Let $A, C \subseteq \Sigma^n$, then the covering radius of C with respect to A is defined to be

$$R(C, A) \triangleq \min \left\{ r \in \mathbb{N} \cup \{0\} \ : \ A \subseteq \bigcup_{\overline{x} \in C} \mathsf{Ball}(r, \overline{x}) \right\} = \max_{\overline{y} \in A} \min_{\overline{x} \in C} d(\overline{x}, \overline{y}).$$

If $A = \Sigma^n$ then $R(C, A)$ is just the regular covering radius of C, and is denoted by $R(C)$.

For constrained systems $X, Y \subseteq \Sigma^{\mathbb{Z}}$ we define the asymptotic covering radius of X with respect to Y to be the asymptotic normalized covering radius of n-tuples from X with respect to n-tuples from Y.

Definition 5: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be shift spaces, then we define

$$R(X, Y) \triangleq \liminf_{n\to\infty} \frac{R(\mathcal{B}_n(X), \mathcal{B}_n(Y))}{n}, \tag{2}$$

where we remind that $\mathcal{B}_n(X)$ and $\mathcal{B}_n(Y)$ are the subwords of length n from X and Y respectively.

In a typical coding-theoretic framework, the covering radius is considered as a property of a single code in the Hamming space of a finite length n. A constrained system on the other hand may be associated with a sequence of codes, which are the sets of constrained words of fixed lengths. The covering radius of the constrained system, as defined above, is in fact the asymptotic value of the (normalized) covering radii of this corresponding sequence of codes.

We remark that throughout the most of this work we take Y to be the whole space $\Sigma^{\mathbb{Z}}$, however the results stated hold for general constrained systems. An immediate question that comes up when considering our definition of the covering radius, is whether the limit from (2) exists. We shall now show that the answer is yes when X or Y are primitive. For our proof, we consider the following simple generalization of Fekete's Lemma.

Lemma 6: Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers satisfying for all $m, n \in \mathbb{N}$

$$a_{n+m} \leqslant a_n + a_m + C, \tag{3}$$

for some constant $C \geqslant 0$. Then the sequence $(\frac{a_n}{n})_{n\in\mathbb{N}}$ converges and

$$\lim_{n\to\infty}\frac{a_n}{n}=\inf_n\frac{a_n+C}{n}.$$

Proof: Assume that $(a_n)_{n\in\mathbb{N}}$ is a sequence and $C\geqslant 0$ are such that (3) is satisfied. Consider the sequence $(b_n)_{n\in\mathbb{N}}$ defined by $b_n \triangleq a_n + C$. We note that b_n is subadditive as by the assumption

$$b_{n+m}=a_{n+m}+C\leqslant(a_n+a_m+C)+C=b_n+b_m.$$

Thus, by Fekete's Lemma [16] the limit $\lim_{n\to\infty} b_n/n$ exists and

$$\inf_n\frac{a_n+C}{n}=\inf_n\frac{b_n}{n}=\lim_{n\to\infty}\frac{a_n+C}{n}=\lim_{n\to\infty}\frac{a_n}{n}.$$

■

We remark that whenever $(a_n)_{n\in\mathbb{N}}$ is a non negative sequence, Lemma 6 also follows from a generalization of Fekete's Lemma to nearly subadditive functions, which are sequences satisfying the De Bruijn–Erdős condition: for all $m,n\in\mathbb{N}$

$$a_{n+m}\leqslant a_n+a_m+f_{n+m},$$

where $(f_k)_{k\in\mathbb{N}}$ is a sequence of numbers such that $\sum_{k=1}^{\infty}\frac{f_k}{k^2}<\infty$. In that case, it was proved in [13] that the sequence $(\frac{a_n}{n})_{n\in\mathbb{N}}$ converges.

Proposition 7: Assume that $X,Y\subseteq\Sigma^{\mathbb{Z}}$ are constrained systems. If X or Y are primitive, then the lim inf *in the definition of $R(X,Y)$ is actually a limit:*

$$R(X,Y)=\lim_{n\to\infty}\frac{R(\mathcal{B}_n(X),\mathcal{B}_n(Y))}{n}.$$

Proof: We begin with the case where X is primitive, and is presented by the primitive finite labeled graph $G=(V,E,L)$. Since G is primitive, there exists a sufficiently large N such that any two vertices are connected by a directed path of length N. This implies that for any two words $\overline{u},\overline{v}\in\mathcal{B}(X)$ there exists $\overline{w}\in\Sigma^N$ such that $\overline{u}\,\overline{w}\,\overline{v}\in\mathcal{B}(X)$. For $n\in\mathbb{N}$ let us denote $a_n=R(\mathcal{B}_n(X),\mathcal{B}_n(Y))$. We show that for all $n>N$ and $m\geqslant 1$, we have $a_{n+m}\leqslant a_n+a_m+N$. In order to show the desired inequality it is sufficient that we prove the following statement: given $\overline{y}\in\mathcal{B}_{n+m}(Y)$ there exists $\overline{x}\in\mathcal{B}_{n+m}(X)$ such that $d(\overline{x},\overline{y})\leqslant a_n+a_m+N$. From the definitions of a_m and a_{n-N} there exists $\overline{u}\in\mathcal{B}_m(X)$ and $\overline{v}\in\mathcal{B}_{n-N}(X)$ such that

$$d(\overline{u},\overline{y}_0^{m-1})\leqslant a_m\quad\text{and}\quad d(\overline{v},\overline{y}_{m+N}^{m+n-1})\leqslant a_{n-N}.$$

We also observe that $(a_k)_{k\in\mathbb{N}}$ is a non-decreasing sequence and therefore $a_{n-N}\leqslant a_n$. Indeed, let $k<k'$, it is sufficient to show that for any $\overline{y}\in\mathcal{B}_k(Y)$ there exists $\overline{x}\in\mathcal{B}_k(X)$ such that $d(\overline{x},\overline{y})\leqslant a_{k'}$. Since Y is a shift space, there exists $\overline{y}'\in\mathcal{B}_{k'}(Y)$ whose prefix of length k is $\overline{y}$. By the definition of $a_{k'}$, there exists $\overline{x}\in\mathcal{B}_{k'}(X)$ such that $d(\overline{x}',\overline{y}')\leqslant a_{k'}$. Let $\overline{x}$ be the prefix of length k of $\overline{x}'$, since X is a shift space, $\overline{x}\in\mathcal{B}_k(X)$, and clearly

$$d(\overline{x},\overline{y})\leqslant d(\overline{x}',\overline{y}')=a_k',$$

as desired.

Let $\overline{w}\in\Sigma^N$ be such that $\overline{x}=\overline{u}\,\overline{w}\,\overline{v}\in\mathcal{B}_{n+m}(X)$. From the properties of the Hamming metric

$$\begin{aligned}d(\overline{x},\overline{y})&=d(\overline{u},\overline{y}_0^{m-1})+d(\overline{w},\overline{y}_m^{m+N-1})+d(\overline{v},\overline{y}_{m+N}^{m+n-1})\\&\leqslant a_m+N+a_{n-N}\leqslant a_m+a_n+N.\end{aligned}$$

When $n\geqslant 1$ and $m>N$ a symmetric analysis follows. Since the remaining cases, i.e., both $m,n\leqslant N$, comprise of a finite number of cases, let $c=\max(N,\max\{a_{m+n}-a_n-a_m\ :\ n,m\leqslant N\})$. For all $m,n\in\mathbb{N}$ we have $a_{n+m}\leqslant a_n+a_m+c$, and therefore by Lemma 6, the sequence $(\frac{a_n}{n})_{n\in\mathbb{N}}$ converges, as desired.

We now turn to the the second case of the proposition, where Y is primitive. As in the first part, by the primitivity of Y, there exists an N such that for any word $\overline{u}\in\mathcal{B}(Y)$ there exists $\overline{u}'\in\Sigma^N$ such that $\overline{u}\,\overline{u}'\overline{u}\in\mathcal{B}(Y)$. Now let $\overline{u}\in\mathcal{B}_m(Y)$ be a deep hole, namely, a word such that $R(\mathcal{B}_m(X),\mathcal{B}_m(Y))=\min\{d(\overline{x},\overline{u})\ :\ \overline{x}\in\mathcal{B}_m(X)\}$. Denote $m'=m+N$. For $n\geqslant m'$, we consider the word $\overline{y}\in\mathcal{B}_n(Y)$ defined by

$$\overline{y}=(\overbrace{\overline{u},\overline{u}',\overline{u},\overline{u}',\ldots,\overline{u},\overline{u}'}^{\lfloor\frac{n}{m'}\rfloor\text{ times}},\overline{y}'),$$

where $\overline{y}'$ is an arbitrary suffix such that $\overline{y}\in\mathcal{B}_n(Y)$. We remark that it is possible to find $\overline{y}'$, for example by taking a prefix of $\overline{u}$. For any $\overline{x}\in\mathcal{B}_n(X)$ we have

$$\begin{aligned}d(\overline{x},\overline{y})&\geqslant\sum_{i=0}^{\lfloor\frac{n}{m'}\rfloor-1}d\Big(\overline{x}_{im'}^{(i+1)m'-1},\overline{y}_{im'}^{(i+1)m'-1}\Big)\\&=\sum_{i=0}^{\lfloor\frac{n}{m'}\rfloor-1}d\Big(\overline{x}_{im'}^{(i+1)m'-1},(\overline{u},\overline{u}')\Big)\\&\geqslant\sum_{i=0}^{\lfloor\frac{n}{m'}\rfloor-1}d\Big(\overline{x}_{im'}^{(i+1)m'-1-N},\overline{u}\Big)\\&\geqslant\left\lfloor\frac{n}{m'}\right\rfloor R(\mathcal{B}_m(X),\mathcal{B}_m(Y))\\&=\left\lfloor\frac{n}{m+N}\right\rfloor R(\mathcal{B}_m(X),\mathcal{B}_m(Y)).\end{aligned}$$

This proves that

$$R(\mathcal{B}_n(X),\mathcal{B}_n(Y))\geqslant\left\lfloor\frac{n}{m+N}\right\rfloor R(\mathcal{B}_m(X),\mathcal{B}_m(Y)).$$

Taking the limit we have

$$\begin{aligned}&\liminf_{n\to\infty}\frac{R(\mathcal{B}_n(X),\mathcal{B}_n(Y))}{n}\\&\quad\geqslant\liminf_{n\to\infty}\frac{\left\lfloor\frac{n}{m+N}\right\rfloor R(\mathcal{B}_m(X),\mathcal{B}_m(Y))}{n}\\&\quad=\frac{R(\mathcal{B}_m(X),\mathcal{B}_m(Y))}{m+N}.\end{aligned}\tag{4}$$

This holds for all $m\in\mathbb{N}$, which proves that

$$\begin{aligned}\liminf_{n\to\infty}\frac{R(\mathcal{B}_n(X),\mathcal{B}_n(Y))}{n}&\geqslant\limsup_{m\to\infty}\frac{R(\mathcal{B}_m(X),\mathcal{B}_m(Y))}{m+N}\\&=\limsup_{m\to\infty}\frac{R(\mathcal{B}_m(X),\mathcal{B}_m(Y))}{m},\end{aligned}$$

and the limit exists. ■

Remark 1: From the proof of Proposition 7 it follows that if $Y = \Sigma^{\mathbb{Z}}$ then

$$R(X,Y) = \lim_{n\to\infty} \frac{R(\mathcal{B}_n(X), \Sigma^n)}{n} = \sup_{n\in\mathbb{N}} \frac{R(\mathcal{B}_n(X), \Sigma^n)}{n}. \quad (5)$$

Indeed, since $Y = \Sigma^{\mathbb{Z}}$ is the full system, which is primitive, the limit exists. Furthermore, in that case the value N from the proof of Proposition 7 is in fact 0, as the concatenation of any two words is again a word in the full system. Plugging in $N = 0$ in (4) we obtain that for all $m \in \mathbb{N}$

$$\liminf_{n\to\infty} \frac{R(\mathcal{B}_n(X), \mathcal{B}_n(Y))}{n} \geqslant \frac{R(\mathcal{B}_m(X), \mathcal{B}_m(Y))}{m},$$

which proves the right-hand side of the equality (5).

We shall use the following as a running example throughout the paper.

Example 1: Consider the binary alphabet, $[2] \triangleq \{0,1\}$. Let $X_{0,k} \subseteq [2]^{\mathbb{Z}}$ be the $(0,k)$-RLL system, which comprises all the binary sequences that do not contain $k+1$ consecutive zeros. That is, $X_{0,k} = X_{\mathcal{F}}$, where $\mathcal{F} = \{\overline{0}^{k+1} = (0,\dots,0)\} \subseteq [2]^{k+1}$. We claim that

$$R(X_{0,k}, [2]^{\mathbb{Z}}) = \frac{1}{k+1}.$$

Indeed, by Remark 1

$$R(X_{0,k}, [2]^{\mathbb{Z}}) \geqslant \frac{R(\mathcal{B}_{k+1}(X_{0,k}), [2]^{k+1})}{k+1} = \frac{R([2]^{k+1} \setminus \{\overline{0}^{k+1}\}, [2]^{k+1})}{k+1} = \frac{1}{k+1}.$$

We now show that the obtained lower bound is tight. Let $\overline{y} \in [2]^n$ be any binary word. Consider $\overline{x}$ given by

$$\overline{x}_i \triangleq \begin{cases} \overline{y}_i & i \bmod (k+1) \neq 0, \\ 1 & i \bmod (k+1) = 0. \end{cases}$$

Clearly $\overline{x}$ does not contain any subword of $k+1$ consecutive zeros and therefore $\overline{x} \in \mathcal{B}_n(X_{0,k})$. Since $d(\overline{x},\overline{y}) \leqslant \lceil \frac{n}{k+1} \rceil$ we conclude that $\frac{1}{n} R(\mathcal{B}_n(X_{0,k}), [2]^n) \leqslant \frac{1}{n} \lceil \frac{n}{k+1} \rceil$, and by taking the limit,

$$R(X_{0,k}, [2]^{\mathbb{Z}}) = \lim_{n\to\infty} \frac{R(\mathcal{B}_n(X_{0,k}), [2]^n)}{n} \leqslant \frac{1}{k+1}.$$

Using a ball-covering argument, we provide a simple lower-bound on the covering radius in terms of the capacities of the systems. We recall that $H_q : [0,1] \to [0,1]$ denotes the q-ary entropy function defined by

$$H_q(x) \triangleq x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x),$$

and for continuity, $H_q(0) \triangleq 0$ as well as $H_q(1) \triangleq \log_q(q-1)$. We also use $H_q^{-1} : [0,1] \to [0, 1-\frac{1}{q}]$ to denote its inverse function.

Proposition 8: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be constrained systems with capacities $h(X) \leqslant h(Y)$, respectively, and let us denote $|\Sigma| = q$. Then

$$R(X,Y) \geqslant H_q^{-1}(h(Y) - h(X)).$$

Proof: If $R(X,Y) \geqslant 1 - \frac{1}{q}$ the claim follows immediately from the definition of H_q^{-1}. Thus, we assume that $R(X,Y) < 1 - \frac{1}{q}$. Let $V_{r,n,q}$ denote the size of a ball of radius r in Σ^n with respect to the Hamming metric (which is invariant to the choice of the center), and let us denote $\rho_n \triangleq \frac{1}{n} R(\mathcal{B}_n(X), \mathcal{B}_n(Y))$. By the union bound, for any $n \in \mathbb{N}$,

$$|\mathcal{B}_n(Y)| = \left| \mathcal{B}_n(Y) \cap \left(\bigcup_{\overline{x}\in\mathcal{B}_n(X)} \mathsf{Ball}(n\rho_n, \overline{x}) \right) \right| \leqslant |\mathcal{B}_n(X)| \cdot V_{\rho_n n, n, q}. \quad (6)$$

By a standard use of Stirling's approximation (e.g., see [21, Chapter 3]) it is well known that

$$V_{\rho n, n, q} = \sum_{i=0}^{\lfloor \rho n \rfloor} \binom{n}{i} (q-1)^i \leqslant \begin{cases} q^{n H_q(\rho)} & \rho \in [0, 1-\frac{1}{q}), \\ q^n & \rho \in [1-\frac{1}{q}, 1]. \end{cases}$$

Since the limit defining the capacity $h(X)$ exists, by (6) and the continuity of H_q we obtain

$$\begin{aligned} h(X) &= \lim_{n\to\infty} \frac{\log_q |\mathcal{B}_n(X)|}{n} = \liminf_{n\to\infty} \frac{\log_q |\mathcal{B}_n(X)|}{n} \\ &\geqslant \liminf_{n\to\infty} \left(\frac{\log_q |\mathcal{B}_n(Y)|}{n} - H_q(\rho_n) \right) \\ &= h(Y) - H_q(R(X,Y)). \end{aligned}$$

By rearranging and employing H_q^{-1} we conclude. ■

We remark that the case $h(Y) \leqslant h(X)$ includes the possibility that $Y \subseteq X$, in which case $R(X,Y) = 0$. So any general lower-bound on $R(X,Y)$ that depends only on $\delta = h(Y) - h(X)$ must vanish when $\delta \leqslant 0$.

Example 2: Fix $\Sigma = [q] \triangleq \{0,\dots,q-1\}$ and consider the repetition shift

$$X_{\mathrm{rep}} \triangleq \{(\dots,a,a,a,\dots) \ : \ a \in [q]\}.$$

Clearly, X_{rep} is the SFT defined by the set of forbidden patterns

$$\mathcal{F} = \{ab \ : \ a,b \in [q], a \neq b\} \subseteq [q]^2.$$

Since $h(X_{\mathrm{rep}}) = 0$, by Proposition 8

$$R(X_{\mathrm{rep}}, [q]^{\mathbb{Z}}) \geqslant H_q^{-1}(1-0) = 1 - \frac{1}{q}.$$

On the other hand, for any $n \in \mathbb{N}$ and for any $\overline{y} \in [q]^n$, it is clear that there exists at least one symbol $a \in [q]$ which appears in at least $\lceil \frac{n}{q} \rceil$ coordinates of $\overline{y}$, and in particular $d(\overline{y}, (a,\dots,a)) \leqslant \lfloor \frac{q-1}{q} n \rfloor$. This proves that

$$R(\mathcal{B}_n(X_{\mathrm{rep}}), [q]^n) \leqslant \left\lfloor \frac{q-1}{q} n \right\rfloor.$$

Taking the limit and combining with the lower bound

$$R(X_{\mathrm{rep}}, [q]^{\mathbb{Z}}) = 1 - \frac{1}{q},$$

and in particular, at this example, the lower bound of Proposition 8 is tight.

Remark 2: Example 2 shows that the covering radius of a union of two constrained systems can be strictly smaller *then the minimum of the covering radii. We note that* $X_{\mathrm{rep}} = \bigcup_{a\in[q]}\{\mathbf{x}_a\}$*, where* $\mathbf{x}_a$ *is the constant bi-infinite sequence of the symbol* a*. We note that* $\{\mathbf{x}_a\}$ *is a constrained system and that* $R(\{\mathbf{x}_a\},[q]^{\mathbb{Z}})=1$*. Thus*

$$\min_{a\in[q]} R(\{\mathbf{x}_a\},[q]^{\mathbb{Z}}) = 1 > 1-\frac{1}{q} = R(X_{\mathrm{rep}},[q]^{\mathbb{Z}}).$$

That is in contrast to the capacity of constrained systems, where the capacity of a finite union is equal to the maximum of the capacities.

Example 3: We recall the $(0,1)$*-RLL system,* $X_{0,1}$*, from Example 1, which is the system of words with no two consecutive zeros, and the binary repetition system,* X_{rep}*, from Example 2. As in the previous examples, we shall calculate* $R(X_{0,1},X_{\mathrm{rep}})$ *and* $R(X_{\mathrm{rep}},X_{0,1})$*. Since* $\mathcal{B}_n(X_{\mathrm{rep}})$ *contains only the all zero word,* $\overline{0}^n$*, and all one word,* $\overline{1}^n$*, and since* $\overline{1}^n \in \mathcal{B}_n(X_{0,1})$*, we have*

$$\begin{aligned} R(\mathcal{B}_n(X_{0,1}),\mathcal{B}_n(X_{\mathrm{rep}})) &= \min_{\overline{x}\in\mathcal{B}_n(X_{0,1})} d(\overline{0}^n,\overline{x}) \\ &= d(\overline{0}^n, 010101\dots) = \left\lfloor \frac{n}{2}\right\rfloor. \end{aligned}$$

We therefore conclude that

$$R(X_{0,1},X_{\mathrm{rep}}) = \lim_{n\to\infty}\frac{1}{n}R(\mathcal{B}_n(X_{0,1}),\mathcal{B}_n(X_{\mathrm{rep}})) = \frac{1}{2}.$$

For calculating $R(X_{\mathrm{rep}},X_{0,1})$*, the exact same argument as in Example 2 shows that*

$$R(\mathcal{B}_n(X_{\mathrm{rep}}),\mathcal{B}_n(X_{0,1})) \leqslant \left\lfloor \frac{n}{2}\right\rfloor.$$

For the lower bound, since the word $\overline{10}^n \triangleq (10101\dots)\in[2]^n$ *belongs to* $\mathcal{B}_n(X_{0,1})$*, we have*

$$\begin{aligned} R(\mathcal{B}_n(X_{\mathrm{rep}}),\mathcal{B}_n(X_{0,1})) &\geqslant \min\{d(\overline{0}^n,\overline{10}^n), d(\overline{1}^n,\overline{10}^n)\} \\ &= \left\lfloor \frac{n}{2}\right\rfloor. \end{aligned}$$

Combining the lower and upper bounds and taking the limit we obtain

$$R(X_{0,1},X_{\mathrm{rep}}) = R(X_{\mathrm{rep}},X_{0,1}) = \frac{1}{2}.$$

We note that unlike in the case where $Y=[2]^{\mathbb{Z}}$ *discussed in Example 2, the lower bound of Proposition 8 is not tight as the capacity of the* $(0,1)$*-RLL system is known to be* $\log_2((1+\sqrt{5})/2)$ *and therefore*

$$\begin{aligned} H_2^{-1}(h(X_{0,1})-h(X_{\mathrm{rep}})) &= H_2^{-1}(\log_2((1+\sqrt{5})/2)) \\ &\approx 0.186 < \frac{1}{2}. \end{aligned}$$

At this point we have reached a curious situation. For the sake of illustrating it, fix the binary alphabet $\Sigma=[2]$. If we consider $X_{0,1}$, the $(0,1)$-RLL system from Example 1, then its capacity is known to be $h(X_{0,1})=\log_2((1+\sqrt{5})/2)\approx 0.694$, and we have shown that its covering radius (with respect to $[2]^{\mathbb{Z}}$) is $R(X_{0,1},[2]^{\mathbb{Z}})=\frac{1}{2}$. However, in Example 2 we have seen that the binary repetition shift, X_{rep}, has the same covering radius $R(X_{\mathrm{rep}},[2]^{\mathbb{Z}})=\frac{1}{2}$, but zero capacity, $h(X_{\mathrm{rep}})=0$. From a coding perspective, even though $\mathcal{B}_n(X_{0,1})$ has exponentially more words than $\mathcal{B}_n(X_{\mathrm{rep}})$, the worst-case covering scenario, namely, a deep hole, is asymptotically within the same distance from the constrained code.

Apart from the mathematical curiosity, having $R(X_{0,1},[2]^{\mathbb{Z}}) = R(X_{\mathrm{rep}},[2]^{\mathbb{Z}}) = \frac{1}{2}$ hinders (in these two example cases) the possibility of correcting channel errors in the QCC scheme described in Section I. This is because the error-correcting code needs to correct more erroneous positions than $\frac{1}{2}$ of the code length, which is impossible to do with a non-vanishing rate. We are therefore motivated to seek a different version of the covering radius of a constrained system, which takes into account the rarity of deep holes.

As a final comment for this section, we would like to comment on the relation of $R(X,\Sigma^{\mathbb{Z}})$ to the QCC framework. Since we are interested in asymptotics, assume that the sequence of error-correcting codes in the QCC scheme is $(C_n)_{n\in\mathbb{N}}$, where C_n is of length n. The expression $R(X,\Sigma^{\mathbb{Z}}) = \lim_{n\to\infty}\frac{1}{n}R(\mathcal{B}_n(X),\Sigma^n)$ is an upper bound on the worse-case quantization error rate using a sequence of codes $(C_n)_{n\in\mathbb{N}}$, which is actually $\lim_{n\to\infty}\frac{1}{n}R(\mathcal{B}_n(X),C_n)$. The bound $R(X,\Sigma^{\mathbb{Z}})$ is pessimistic twice: once for allowing deep holes to determine the covering radius, and twice, for assuming they reside in C_n. Since $\lim_{n\to\infty}\frac{1}{n}R(\mathcal{B}_n(X),C_n)$ may be hard to compute and depends on the sequence of error-correcting codes, we may use $R(X,\Sigma^{\mathbb{Z}})$ as an upper bound on the worse-case quantization error, which is independent of the sequence of codes.

IV. The Essential Covering Radius

The covering radius that was studied in the previous section may be perhaps too pessimistic in the sense that it is determined by the worst-case quantization distance. In this section, we study a different definition of the covering radius, which we call the essential covering radius. Given $\varepsilon>0$, the ε-covering radius of a constraint system is, loosely speaking, the smallest r such that $(1-\varepsilon)$-fraction the words in the space can be quantized to the constraint system. In what follows, we further generalize this to a probabilistic definition of the covering radius.

We begin by stating some basic definitions and well-known results from ergodic theory. For any finite alphabet Σ, we consider $\Sigma^{\mathbb{Z}}$ as a measurable space, together with the Borel Σ-algebra induced by the product topology on $\Sigma^{\mathbb{Z}}$. Similarly, any subshift $Y\subseteq\Sigma^{\mathbb{Z}}$ is considered as a measurable space.

Definition 9 (Invariant and Ergodic Measures): Let $Y\subseteq\Sigma^{\mathbb{Z}}$ *be a subshift. A probability measure* μ *on* Y *is called* shift invariant *if* $\mu(T^{-1}B)=\mu(B)$ *for any measurable set* B*. A shift-invariant measure* μ *is further said to be* ergodic *if* $T^{-1}B=B$ *implies* $\mu(B)=0$ *or* $\mu(Y\setminus B)=0$*. The set of shift-invariant probability measures on* Y *is denoted by* $M(Y)$*, and the set of ergodic measures in* $M(Y)$ *is denoted by* $M_{\mathcal{E}}(Y)$*.*

For a measure $\mu\in M(Y)$ we denote by μ_n the marginal measure of μ on the coordinates $0,1\dots,n-1$, which is a probability measure on Σ^n. To avoid cumbersome notation, throughout this work we shall use $\mathbb{P}_\mu[A]$ in order to denote the measure $\mu(A)$, and $\mathbf{Y}$ for a random bi-infinite sequence on Y. Throughout this article we use bold upper-case letters for bi-infinite sequences of random variables, not to be confused with non-bold capital letters used to denote constrained systems. An important result that we use in our analysis is the following well-known ergodic theorem, which is a classical result in Ergodic Theory. The L_2 convergence in the ergodic theorem is due to von-Neumann and the almost-surely result is due to Birkhoff, both from 1931. A proof of this well-known classical theorem can be found in most standard introductory textbooks on ergodic theory, for instance [36, Chapter 3].

Theorem 10 (The Ergodic Theorem, [6], [31]): Let $Y \subseteq \Sigma^{\mathbb{Z}}$ be a shift space, $\mu \in M_{\mathcal{E}}(Y)$ be an ergodic measure, and $f \in L_1(\mu)$ be an integrable function. Then the sequence $(A_n)_{n\in\mathbb{N}}$ defined by

$$A_n = \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^{-i}$$

converges almost-surely, and in L_2, to $\int f \cdot d\mu$. In particular, for all $\delta > 0$

$$\lim_{n\to\infty} \mathbb{P}_\mu\left[\left|A_n - \int f \cdot d\mu\right| > \delta\right] = 0.$$

We are now ready to define the essential covering radius.

Definition 11: For any real $\varepsilon > 0$, two sets $A, C \subseteq \Sigma^n$, and η, a probability measure on A, we define

$$R_\varepsilon(C, A, \eta) \triangleq \min\left\{ r \in \mathbb{N}\cup\{0\} \ : \ \eta\left(A \cap \left(\bigcup_{\overline{x}\in C} \mathsf{Ball}(r, \overline{x})\right)\right) \geqslant 1-\varepsilon \right\}.$$

We remark that when η is the uniform measure on A, $R_\varepsilon(C, A, \eta)$ is the ε-covering radius of A, namely the smallest r such that at least $(1-\varepsilon)$-fraction of the words in C are at distance at most r from A, as desired.

*Definition 12: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be constrained systems, and $\mu \in M_{\mathcal{E}}(Y)$ be an ergodic measure. We define the ε-*covering radius *of X with respect to (Y, μ) by*

$$R_\varepsilon(X, Y, \mu) \triangleq \liminf_{n\to\infty} \frac{R_\varepsilon(\mathcal{B}_n(X), \mathcal{B}_n(Y), \mu_n)}{n},$$

and the essential covering radius *of X with respect to (Y, μ) by*

$$R_0(X, Y, \mu) \triangleq \lim_{\varepsilon\to 0} R_\varepsilon(X, Y, \mu).$$

We comment that the limit in the previous definition exists due to the monotonicity of $R_\varepsilon(X, Y, \mu)$ in ε. We also observe that, trivially, the essential covering radius is upper bounded by the (worst-case) covering radius, for every $\varepsilon > 0$

$$R_0(X, Y, \mu) \leqslant R_\varepsilon(X, Y, \mu) \leqslant R(X, Y). \tag{7}$$

We now review the examples of the repetition system (Example 2) and the $(0, k)$-RLL system (Example 1), considered in the previous section.

Proposition 13: Consider the q-ary repetition system $X_{\text{rep}} \subseteq [q]^{\mathbb{Z}}$ from Example 2, and assume $Y = [q]^{\mathbb{Z}}$ is equipped with the uniform Bernoulli i.i.d measure, denoted by μ^{u}. Then the essential covering radius is equal to the covering radius, i.e.,

$$R_0(X_{\text{rep}}, [q]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = R(X_{\text{rep}}, [q]^{\mathbb{Z}}) = 1 - \frac{1}{q}.$$

Proof: For a fixed $n \in \mathbb{N}$ and $i \in [q]$, let $Y_n^{(i)}$ be the random variable that counts the number of coordinates with the symbol i in a random (uniformly distributed) word in $[q]^n$. By the law of large numbers, for any $\delta > 0$

$$\lim_{n\to\infty} \mathbb{P}_{\mu_n^{\mathsf{u}}}\left[\bigcap_{i=0}^{q-1}\left\{\left|\frac{Y_n^{(i)}}{n} - \frac{1}{q}\right| < \delta\right\}\right] = 1. \tag{8}$$

We also note that for a word $\overline{y} \in [q]^n$ with each symbol $i \in [q]$ appearing in at least $(\frac{1}{q} - \delta)n$ coordinates, we have

$$\min_{\overline{x}\in\mathcal{B}_n(X_{\text{rep}})} d(\overline{x}, \overline{y}) \geqslant (q-1)\left(\frac{1}{q} - \delta\right)n. \tag{9}$$

For any $\varepsilon > 0$, from (8) it follows that for sufficiently large n, any set in $[q]^n$, of probability at least $1-\varepsilon$, contains a word such that any $i \in [q]$ appears in at least $(\frac{1}{q} - \delta)n$ coordinates. Combining this with (9) we conclude that for sufficiently large n

$$\frac{R_\varepsilon(\mathcal{B}_n(X_{\text{rep}}), [q]^n, \mu_n^{\mathsf{u}})}{n} \geqslant (q-1)\left(\frac{1}{q} - \delta\right).$$

It then follows that $R_\varepsilon(X_{\text{rep}}, [q]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant (q-1)(\frac{1}{q} - \delta)$, which is true for all $\delta > 0$, and therefore

$$R_0(X_{\text{rep}}, [q]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = \lim_{\varepsilon\to 0} R_\varepsilon(X_{\text{rep}}, [q]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant 1 - \frac{1}{q}.$$

The upper bound follows trivially from Example 2, and we have

$$R_0(X_{\text{rep}}, [q]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = R(X_{\text{rep}}, [q]^{\mathbb{Z}}) = 1 - \frac{1}{q}.$$

∎

Remark 3: Since (7) is true for all ergodic measures, we can lower bound the covering radius as follows,

$$R(X, Y) \geqslant \sup_{\mu\in M_{\mathcal{E}}(Y)} R_0(X, Y, \mu). \tag{10}$$

Proposition 13 shows that in the case where $X = X_{\text{rep}}$ is the repetition system and $Y = [q]^{\mathbb{Z}}$, (10) is tight, as equality holds for the uniform Bernoulli measure on Y.

We claim that (10) is tight also in the case where $X = X_{0,k}$, the binary $(0, k)$-RLL system described in Example 1. To see this, consider $\mu = \delta_{\mathbf{0}}$, the Dirac measure of the all-zero sequence $\mathbf{0} \in Y = [2]^{\mathbb{Z}}$. Since $\mathbf{0}$ is a periodic point with period 1 (with respect to the shift transformation), $\delta_{\mathbf{0}}$ is indeed an invariant measure, which is also ergodic, as it assigns a measure of 0 or 1 to any set. A simple calculation shows that for any $0 < \varepsilon < 1$ and $n \in \mathbb{N}$,

$$R_\varepsilon(\mathcal{B}_n(X_{0,k}), [2]^n, \mu_n) = \min_{\overline{x}\in\mathcal{B}_n(X_{0,k})} d(\overline{x}, \overline{0}^n) = \left\lfloor \frac{n}{k+1} \right\rfloor,$$

which implies

$$R_0(X_{0,k}, [2]^{\mathbb{Z}}, \mu) = \lim_{\varepsilon\to 0} R_\varepsilon(X_{0,k}, [2]^{\mathbb{Z}}, \mu) = \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{R_\varepsilon(\mathcal{B}_n(X_{0,k}), [2]^n, \mu_n)}{n}$$

$$= \frac{1}{k+1} = R(X,Y).$$

We conjecture that the bound (10) is tight in general, and leave it as a direction for future work.

As we have seen, the repetition system, whose capacity is zero, has the same covering radius and essential covering radius. The $(0,k)$-RLL system has positive capacity. While its covering radius is $\frac{1}{k+1}$, the following theorem shows its essential covering radius decays exponentially fast with k, in stark contrast to the repetition system.

Theorem 14: Let $X_{0,k} \subseteq [2]^{\mathbb{Z}}$ be the $(0,k)$-RLL system from Example 1, and let $Y = [2]^{\mathbb{Z}}$ be equipped with the uniform Bernoulli i.i.d measure μ^{u}. Then

$$R_0(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = \frac{1}{2(2^{k+1}-1)}.$$

Proof: For a word $\overline{y} \in [2]^n$ and an integer $i \geqslant k+1$, let us denote by $S_i(\overline{y})$ the number of appearances of the pattern $1\overline{0}^i 1 \in [2]^{i+2}$ in $\overline{y}$. We also denote by $M(\overline{y})$ the number of coordinates $j \in [n]$ that are not part of a pattern of the form $1\overline{0}^i 1$ in $\overline{y}$, for any $i \geqslant 1$. The key observation is the following inequality, asserting that for all $\ell \geqslant k+1$ we have

$$\sum_{i=k+1}^{\ell} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\overline{y}) \leqslant \min_{\overline{x} \in \mathcal{B}_n(X_{0,k})} d(\overline{x}, \overline{y}) \leqslant M(\overline{y}) + \sum_{i=k+1}^{\infty} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\overline{y}). \quad (11)$$

Indeed, we note that for any $\overline{x} \in \mathcal{B}_n(X_{0,k})$ and for any instance of the pattern $\overline{y}_m^{m+i+1} = 1\overline{0}^i 1$, we have that $\overline{x}_{m+1}^{m+i}$ and $\overline{y}_{m+1}^{m+i}$ must differ in at least $\lfloor \frac{i}{k+1} \rfloor$ places since $\overline{x}$ does not contain $k+1$ consecutive zeros. We further observe that patterns of the form $1\overline{0}^i 1$ in $\overline{y}$ do not overlap in zeros. Thus we obtain,

$$d(\overline{x}, \overline{y}) \geqslant \sum_{i=k+1}^{\infty} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\overline{y}) \geqslant \sum_{i=k+1}^{\ell} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\overline{y}),$$

and the lower-bound follows by taking the minimum over all $\overline{x} \in \mathcal{B}_n(X_{0,k})$. On the other hand, in order to prove the upper bound, it suffices to construct a word $\overline{x} \in \mathcal{B}_n(X_{0,k})$ satisfying

$$d(\overline{x}, \overline{y}) \leqslant M(\overline{y}) + \sum_{i=k+1}^{\infty} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\overline{y}). \quad (12)$$

We construct $\overline{x}$ as follows: for any $i \geqslant k+1$ and any instance of the pattern $1\overline{0}^i 1$ in $\overline{y}_m^{m+i+1}$, we set

$$\overline{x}_m^{m+i+1} = 1 \overbrace{\overline{0}^k 1 \overline{0}^k 1 \cdots \overline{0}^k 1}^{\lfloor \frac{i}{k+1} \rfloor (k+1)} \overbrace{0 \cdots 0}^{i \bmod (k+1)} 1.$$

In the remaining coordinates we define $\overline{x}$ to be the same as $\overline{y}$ except the coordinates counted by $M(\overline{y})$ which we set to 1. From the construction of $\overline{x}$, the longest run of zeros it contains is at most k, which implies $\overline{x} \in \mathcal{B}_n(X_{0,k})$. Thus, $\overline{x}$ satisfies (12) as desired.

For any $\ell \geqslant k+1$ (including $\ell = \infty$) we define $f_\ell : [2]^{\mathbb{Z}} \to \mathbb{R}$ to be

$$f_\ell(\mathbf{y}) = \begin{cases} \left\lfloor \frac{i}{k+1} \right\rfloor & \mathbf{y}_0^{i+1} = 1\overline{0}^i 1 \text{ for } k+1 \leqslant i \leqslant \ell, \\ 0 & \text{otherwise,} \end{cases}$$

and we denote

$$E_\ell \triangleq \int f_\ell \cdot d\mu^{\mathsf{u}}.$$

A simple calculation shows that for any finite $\ell \in \mathbb{N}$

$$E_\ell = \sum_{i=k+1}^{\ell} \left\lfloor \frac{i}{k+1} \right\rfloor \frac{1}{2^{i+2}},$$

and by the monotone convergence theorem $E_\ell \to E_\infty$ as $\ell \to \infty$, since f_ℓ (monotonically) converges to f_∞ pointwise.

Let $\overline{Y}_n$ be a random word of length n distributed according to μ_n^{u}, and let $\mathbf{Y}$ be a random sequence distributed according to μ^{u}. We note that for any $\ell \leqslant n-2$,

$$\sum_{j=0}^{n-\ell-2} f_\ell \circ T^j(\mathbf{Y}) = \sum_{i=0}^{\ell} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\mathbf{Y}_0^{n-1}).$$

Combining with (11), for a fixed ℓ and $\delta > 0$, and for all sufficiently large n,

$$\begin{aligned} &\mathbb{P}_{\mu_n^{\mathsf{u}}}\left[\min_{\overline{x}_n \in \mathcal{B}_n(X_{0,k})} \frac{d(\overline{x}_n, \overline{Y}_n)}{n} > E_\ell - \delta \right] \\ &= \mathbb{P}_{\mu^{\mathsf{u}}}\left[\min_{\overline{x}_n \in \mathcal{B}_n(X_{0,k})} \frac{d(\overline{x}_n, \mathbf{Y}_0^{n-1})}{n} > E_\ell - \delta \right] \\ &\geqslant \mathbb{P}_{\mu^{\mathsf{u}}}\left[\frac{1}{n} \sum_{i=k+1}^{\ell} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\mathbf{Y}_0^{n-1}) > E_\ell - \delta \right] \\ &= \mathbb{P}_{\mu^{\mathsf{u}}}\left[\frac{1}{n} \sum_{j=0}^{n-\ell-2} f_\ell \circ T^j(\mathbf{Y}) > E_\ell - \delta \right] \xrightarrow[n\to\infty]{} 1, \end{aligned}$$

where the convergence to 1 is due to the ergodic theorem (see Theorem 10). This proves that for any $\varepsilon \in (0,1)$, $\delta > 0$, and $\ell \geqslant k+1$, for sufficiently large n

$$\frac{R_\varepsilon(\mathcal{B}_n(X_{0,k}), \mathcal{B}_n(Y), \mu^{\mathsf{u}})}{n} > E_\ell - \delta,$$

and therefore,

$$R_\varepsilon(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant \lim_{\substack{\ell \to \infty \\ \delta \to 0}} E_\ell - \delta = E_\infty.$$

For the upper bound, we use a similar technique. We first note that the sequence of random variables $(\frac{1}{n} M(\mathbf{Y}_0^{n-1}))_{n \in \mathbb{N}}$ converges in probability to 0. We note that for any $N \in \mathbb{N}$, and a word $\overline{y}$ of length n, if $M(\overline{y}) > 2N$ then $\overline{y}_0^{N-1} = \overline{0}$ or $\overline{y}_{n-N}^{n-1} = \overline{0}$. Thus for any $\delta > 0$

$$\begin{aligned} &\mathbb{P}_{\mu^{\mathsf{u}}}\left[\left| \frac{1}{n} M(\mathbf{Y}_0^{n-1}) - 0 \right| > 2\delta \right] \\ &\leqslant \mathbb{P}_{\mu^{\mathsf{u}}}\left[\mathbf{Y}_0^{\lfloor \delta n \rfloor} = \overline{0} \right] + \mathbb{P}_{\mu^{\mathsf{u}}}\left[\mathbf{Y}_{n-\lfloor \delta n \rfloor}^{n-1} = \overline{0} \right] \\ &= \frac{2}{2^{\lfloor \delta n \rfloor}} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

We similarly note that for any n,

$$\sum_{j=0}^{n-1} f_\infty \circ T^j(\mathbf{Y}) \geqslant \sum_{i=0}^{\infty} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\mathbf{Y}_0^{n-1}).$$

Combining with (11), for all $\delta > 0$

$$\begin{aligned}
&\mathbb{P}_{\mu_n^{\mathsf{u}}}\left[\min_{\overline{x}_n \in \mathcal{B}_n(X_{0,k})} \frac{d(\overline{x}_n, \overline{Y}_n)}{n} < E_\infty + \delta\right] \\
&\geqslant \mathbb{P}_{\mu^{\mathsf{u}}}\left[\frac{1}{n}\left(M(\mathbf{Y}_0^{n-1}) + \sum_{i=k+1}^{\infty} \left\lfloor \frac{i}{k+1} \right\rfloor S_i(\mathbf{Y}_0^{n-1})\right) < E_\infty + \delta\right] \\
&\geqslant \mathbb{P}_{\mu^{\mathsf{u}}}\left[\frac{1}{n}\left(M(\mathbf{Y}_0^{n-1}) + \sum_{j=0}^{n-1} f_\infty \circ T^j(\mathbf{Y})\right) < E_\infty + \delta\right] \xrightarrow[n\to\infty]{} 1,
\end{aligned}$$

where again, the convergence to 1 follows from the ergodic theorem and from the convergence in probability of $\frac{1}{n}M(\mathbf{Y}_0^{n-1})$ to 0. As before, this proves that $R_\varepsilon(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant E_\infty$ for all $\varepsilon > 0$, and by the lower bound $R_\varepsilon(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = E_\infty$. In particular,

$$R_0(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = \lim_{\varepsilon\to 0} R_\varepsilon(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = E_\infty.$$

In order to complete the proof it only remains to compute E_∞:

$$\begin{aligned}
E_\infty &= \sum_{i=k+1}^{\infty} \left\lfloor \frac{i}{k+1} \right\rfloor \frac{1}{2^{i+2}} = \sum_{j=1}^{\infty} j \sum_{i=j(k+1)}^{(j+1)(k+1)-1} \frac{1}{2^{i+2}} \\
&= \frac{2^{k+1}-1}{2^{k+2}} \sum_{j=1}^{\infty} \frac{j}{2^{j(k+1)}} = \frac{2^{k+1}-1}{2^{k+2}} \cdot \frac{2^{k+1}}{(2^{k+1}-1)^2} \\
&= \frac{1}{2(2^{k+1}-1)}.
\end{aligned}$$

■

It is desirable to have alternative expressions for the essential covering radius, which could assist in calculating or estimating its value. Inspired by tools used in the proof of Theorem 14, we give an equivalent ergodic-theoretic characterization.

Definition 15: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be shift spaces, we consider $X \times Y$ as shift space, with the left shift acting as $T(\mathbf{x}, \mathbf{y}) = (T\mathbf{x}, T\mathbf{y})$. For an ergodic measure $\mu \in M_{\mathcal{E}}(Y)$, an extension of μ over $X \times Y$ is a shift-invariant measure ν on the product space $X \times Y$ whose Y-marginal is μ. Namely, ν satisfies that for any measurable $A \subseteq Y$

$$\nu(X \times A) = \mu(A).$$

An extension on $X \times Y$ is said to be ergodic if it is an ergodic measure with respect to the shift transformation on the product space. We let $M(X, Y, \mu)$ denote the set of all extensions of μ, and $M_{\mathcal{E}}(X, Y, \mu)$ denote the set of all ergodic extensions in $M(X, Y, \mu)$. We comment that $M(X, Y, \mu)$ and $M_{\mathcal{E}}(X, Y, \mu)$ are non-empty for any measure μ. Indeed, taking the independent coupling of any invariant measure η on X with the measure μ gives an invariant extension in $M(X, Y, \mu)$. The existence of an ergodic extension is a shown inside the proof of Proposition 18.

In the following proposition, we provide an upper bound on the essential covering radius, that holds with no further assumptions on X and Y.

Proposition 16: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be shift spaces, and $\mu \in M_{\mathcal{E}}(Y)$. Then

$$R_0(X, Y, \mu) \leqslant \inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \ : \ \nu \in M_{\mathcal{E}}(X, Y, \mu)\},$$

where $\mathbf{X}_0, \mathbf{Y}_0$ are the random variables defined by the projections of two random sequences $\mathbf{X}$ and $\mathbf{Y}$ on the 0 coordinate.

Proof: Let ν be an extension in $M_{\mathcal{E}}(X, Y, \mu)$, and let $0 < \varepsilon < 1$ and $\delta > 0$ be arbitrarily small numbers. It is sufficient to prove that

$$R_\varepsilon(X, Y, \mu) \leqslant \mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] + \delta.$$

We consider the function $f : X \times Y \to \{0, 1\}$, defined to be the indicator function of the event $\{\mathbf{X}_0 \neq \mathbf{Y}_0\}$. Clearly

$$\int f \cdot d\nu = \mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0].$$

By Theorem 10, for sufficiently large n,

$$\mathbb{P}_\nu\left[\left|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^n - \int f \cdot d\nu\right| > \delta\right] < \varepsilon. \tag{13}$$

We also observe that

$$\sum_{i=0}^{n-1} f \circ T^n = d\big(\mathbf{X}_0^{n-1}, \mathbf{Y}_0^{n-1}\big), \tag{14}$$

where d is the Hamming distance. Let us denote

$$s_\delta \triangleq \mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] + \delta$$

$$B_{s_\delta}(\mathcal{B}_n(X)) \triangleq \mathcal{B}_n(Y) \cap \left(\bigcup_{\overline{x} \in \mathcal{B}_n(X)} \mathsf{Ball}(n \cdot s_\delta, \overline{x})\right).$$

Combining (13) and (14) with the law of total probability we obtain

$$\begin{aligned}
&1 - \varepsilon \\
&< \mathbb{P}_\nu\big[d\big(\mathbf{X}_0^{n-1}, \mathbf{Y}_0^{n-1}\big) \leqslant n \cdot s_\delta\big] \\
&= \sum_{\overline{y} \in \mathcal{B}_n(Y)} \mathbb{P}_\nu\left[d\big(\mathbf{X}_0^{n-1}, \mathbf{Y}_0^{n-1}\big) \leqslant n \cdot s_\delta \;\middle|\; \mathbf{Y}_0^{n-1} = \overline{y}\right] \cdot \mathbb{P}_\nu\big[\mathbf{Y}_0^{n-1} = \overline{y}\big] \\
&= \sum_{\overline{y} \in B_{s_\delta}(\mathcal{B}_n(X))} \mathbb{P}_\nu\left[d\big(\mathbf{X}_0^{n-1}, \mathbf{Y}_0^{n-1}\big) \leqslant n \cdot s_\delta \;\middle|\; \mathbf{Y}_0^{n-1} = \overline{y}\right] \cdot \mathbb{P}_\nu\big[\mathbf{Y}_0^{n-1} = \overline{y}\big] \\
&\leqslant \sum_{\overline{y} \in B_{s_\delta}(\mathcal{B}_n(X))} \mathbb{P}_\nu\big[\mathbf{Y}_0^{n-1} = \overline{y}\big] \\
&= \sum_{\overline{y} \in B_{s_\delta}(\mathcal{B}_n(X))} \mathbb{P}_\mu\big[\mathbf{Y}_0^{n-1} = \overline{y}\big]
\end{aligned}$$

$$= \mathbb{P}_{\mu_n}\left[\mathcal{B}_n(Y) \cap \left(\bigcup_{\overline{x}\in\mathcal{B}_n(X)} \mathsf{Ball}(n\cdot s_\delta, \overline{x})\right)\right].$$

This shows that

$$R_\varepsilon(X,Y,\mu) \leqslant s_\delta = \mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] + \delta,$$

and therefore completes the proof. ■

The following proposition shows that the minimization problem from the upper bound of Proposition 16 over the set of ergodic extensions $M_{\mathcal{E}}(X,Y,\mu)$ is in fact equivalent to a minimization over the set of all invariant extensions $M(X,Y,\mu)$. In order to show that, we shall require the ergodic decomposition theorem. Let Z be a compact metric space equipped with the Borel σ-algebra and $T : Z \to Z$ be a continuous function. We consider the space of T-invariant measures on Z, $M(Z)$, as a measurable space with the Σ-algebra induced by the weak-$*$ topology.

Theorem 17 (Ergodic Decomposition, [14, Theorem 4.8]): Let Z be a compact metric space, $T : X \to X$ be a continuous map, and $\mu \in M(Z)$ be an invariant measure. Then there exists a unique probability measure P_μ on $M(Z)$, supported on the set of ergodic measures $M_{\mathcal{E}}(Z)$, such that for any measurable set $E \subseteq Z$,

$$\mu(E) = \int_{M_{\mathcal{E}}(Z)} \nu(E)dP_\mu(\nu).$$

Proposition 18: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$, and let $\mu \in M_{\mathcal{E}}(Y)$ be a shift-invariant ergodic measure. Then

$$\begin{aligned}\inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \;:\; \nu \in M_{\mathcal{E}}(X,Y,\mu)\}\\ = \inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \;:\; \nu \in M(X,Y,\mu)\}.\end{aligned}$$

Proof: Let us denote

$$m \triangleq \inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \;:\; \nu \in M_{\mathcal{E}}(X,Y,\mu)\}.$$

The inequality

$$m \geqslant \inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \;:\; \nu \in M(X,Y,\mu)\}$$

is trivial as $M_{\mathcal{E}}(X,Y,\mu) \subseteq M(X,Y,\mu)$. For the other direction, we are required to show that for all $\eta \in M(X,Y,\mu)$

$$\mathbb{P}_\eta[\mathbf{X}_0 \neq \mathbf{Y}_0] \geqslant m.$$

Let P_η be the ergodic decomposition of η. Namely $\eta = \int_{M_{\mathcal{E}}(X\times Y)} \nu \cdot dP_\eta(\nu)$, and the set of ergodic measures has full dP_η-measure. We start by showing that P_η is supported on $M_{\mathcal{E}}(X,Y,\mu)$, which also proves that $M_{\mathcal{E}}(X,Y,\mu)$ is non-empty. Assume to the contrary that

$$P_\eta[\{\nu \in M_{\mathcal{E}}(X\times Y) \;:\; \nu_Y \neq \mu\}] > 0,$$

where ν_Y is the projection (marginal) of ν on Y. Since the Borel Σ-algebra on Y is countably generated, there exists a measurable set $E \subseteq Y$ and $n \in \mathbb{N}$ such that at least one of the sets A_+ and A_- defined by

$$A_+ \triangleq \left\{\nu \in M_{\mathcal{E}}(X\times Y) \;:\; \nu_Y(E) \geqslant \mu(E) + \frac{1}{n}\right\},$$
$$A_- \triangleq \left\{\nu \in M_{\mathcal{E}}(X\times Y) \;:\; \nu_Y(E) \leqslant \mu(E) - \frac{1}{n}\right\},$$

has a positive P_η-measure. Without loss of generality, assume $P_\eta(A_+) > 0$. We define a new measure η' by

$$\eta'(A) = \frac{1}{P_\eta(A_+)} \int_{A_+} \nu(A) \cdot dP_\eta(\nu).$$

We observe that η' is a shift-invariant measure (as an integral over shift-invariant measures). We also note that $\eta'_Y \ll \eta_Y$ (that is, η'_Y is absolutely continuous with respect to η_Y) as any null set with respect to η is obviously a null set with respect to η'. It is well known that any invariant probability measure which is absolutely continuous with respect to an invariant ergodic probability measure must be equal to it (e.g. see [37, Remark 1, p.153]). We recall that $\eta_Y = \mu$ is ergodic, and therefore $\eta_Y = \eta'_Y$. This is a contradiction as from the definition of A_+ and η' we have

$$\begin{aligned}\mu(E) &= \eta_Y(E) = \eta'_Y(E) = \eta'(X\times E)\\ &= \frac{1}{P_\eta(A_+)} \int_{A_+} \nu(X\times E)\cdot dP_\eta(\nu)\\ &\geqslant \frac{1}{P_\eta(A_+)} \int_{A_+} \left(\mu(E) + \frac{1}{n}\right) dP_\eta(\nu) = \mu(E) + \frac{1}{n}.\end{aligned}$$

The claim now follows, since from the definition of m we have

$$\begin{aligned}\mathbb{P}_\eta[\mathbf{X}_0 \neq \mathbf{Y}_0] &= \int_{M_{\mathcal{E}}(X\times Y)} \mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] dP_\eta(\nu)\\ &= \int_{M_{\mathcal{E}}(X,Y,\mu)} \mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] dP_\eta(\nu)\\ &\geqslant \int_{M_{\mathcal{E}}(X,Y,\mu)} m \cdot dP_\eta(\nu) = m.\end{aligned}$$

■

We are now ready to prove the main result of the section: the upper bound given in Proposition 16 is in fact tight, and it provides an exact characterization of the essential covering radius by a minimization problem over invariant extensions.

Theorem 19: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be constrained systems, and let $\mu \in M_{\mathcal{E}}(Y)$ be an ergodic measure. Then

$$\begin{aligned}R_0(X,Y,\mu) &= \inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \;:\; \nu \in M_{\mathcal{E}}(X,Y,\mu)\}\\ &= \inf\{\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \;:\; \nu \in M(X,Y,\mu)\}.\end{aligned}$$

Proof: By Proposition 18 and Proposition 16, it is sufficient to prove that for any $\delta > 0$, there exists an invariant extension $\nu \in M(X,Y,\mu)$ with

$$\mathbb{P}_\nu[\mathbf{X}_0 \neq \mathbf{Y}_0] \leqslant R_0(X,Y,\mu) + \delta. \tag{15}$$

We shall prove the existence of such an extension using the compactness of the simplex of probability measures $X \times Y$ with respect to the weak-$*$ topology. We fix $\delta > 0$ and find $\varepsilon < \delta/4$ in $(0,1)$ such that

$$R_\varepsilon(X,Y,\mu) \leqslant R_0(X,Y,\mu) + \frac{\delta}{4}. \tag{16}$$

For any fixed word $\overline{y} \in \mathcal{B}_n(Y)$ we fix an arbitrary $\overline{x}(\overline{y}) \in \mathcal{B}_n(X)$ which is closest to $\overline{y}$ among the words in $\mathcal{B}_n(X)$.

From the definition of $R_\varepsilon(X,Y,\mu)$, there exists a sequence of distinct integers $(n_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty}\frac{R_\varepsilon(\mathcal{B}_{n_k}(X),\mathcal{B}_{n_k}(Y),\mu_{n_k})}{n_k}=R_\varepsilon(X,Y,\mu).$$

From the definition of $R_\varepsilon(\mathcal{B}_{n_k}(X),\mathcal{B}_{n_k}(Y),\mu_{n_k})$ we have that for sufficiently large k

$$\mathbb{P}_\mu\left[\frac{1}{n_k}d\big(\overline{x}(\mathbf{Y}_0^{n_k-1}),\mathbf{Y}_0^{n_k-1}\big)<R_\varepsilon(X,Y,\mu)+\frac{\delta}{4}\right]>1-\varepsilon. \tag{17}$$

For a fixed k, we define a map $f_k:Y\to X$ as follows: for any $\overline{x}\in\mathcal{B}_{n_k}(X)$ we fix some $\mathbf{x}\in X$ such that $\mathbf{x}_0^{n_k-1}=\overline{x}$. We then define $f_k(\mathbf{y})=\mathbf{x}$, where $\mathbf{x}$ is the sequence in X corresponding to $\overline{x}(\mathbf{y}_0^{n_k-1})$. Clearly f_k is measurable since $f_k(\mathbf{y})_m$ depends on finitely many coordinates of $\mathbf{y}$. We now consider the measure ν'_k on $X\times Y$, defined as the pushforward of $(f_k,\mathrm{Id}):Y\to X\times Y$. We also define ν_k to be the measure obtained by averaging of ν'_k along the action of the shift, which is given by

$$\nu=\frac{1}{n_k}\sum_{i=0}^{n_k-1}T_*^i\nu'_k,$$

where $T_*^i\nu'_k$ is the pushforward of ν'_k with the ith-shift, defined by $T_*^i\nu'_k(A)=\nu'_k(T^{-i}A)$. We note that the Y-marginal of ν'_k is μ since (f_k,Id) is the identity on the Y-coordinate. Since μ is shift invariant it follows that the Y-marginal of ν_k is also μ.

By the compactness of the set of probability measures on $X\times Y$ with respect to the weak-$*$ topology, there exists a convergent subsequence $(\nu_{k_l})_{l\in\mathbb{N}}$, which by abuse of notation we denote by $(\nu_k)_k$. Let ν denote the weak-$*$ limit of $(\nu_k)_k$. Since the projection of a measure to its marginal is continuous with respect to the weak-$*$ topology, the Y-marginal of ν is μ. We shall now show that ν is indeed an invariant measure satisfying (15).

For the invariance of ν, it is sufficient to show that for any continuous function on $X\times Y$ we have

$$\int f\cdot d\nu=\int f\circ T\cdot d\nu.$$

Indeed,

$$\begin{aligned}&\int f\cdot d\nu-\int f\circ T\cdot d\nu\\&=\lim_{k\to\infty}\int f\cdot d\nu_k-\lim_{k\to\infty}\int f\circ T\cdot d\nu_k\\&=\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=0}^{n_k-1}\left(\int f\cdot dT_*^i\nu'_k-\int f\circ T\cdot dT_*^i\nu'_k\right)\\&=\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=0}^{n_k-1}\left(\int f\circ T^i\cdot d\nu'_k-\int f\circ T^{i+1}\cdot d\nu'_k\right)\\&=\lim_{k\to\infty}\frac{1}{n_k}\left(\int f\cdot d\nu'_k-\int f\circ T^{n_k}\cdot d\nu'_k\right)=0,\end{aligned}$$

where the convergence to 0 follows since f is bounded (as a continuous function on a compact space).

It now remains to show that ν satisfies (15). From the definitions of ν, ν_k, and ν'_k, for all k we have

$$\begin{aligned}\mathbb{P}_\nu[\mathbf{X}_0\neq\mathbf{Y}_0]&=\lim_{k\to\infty}\mathbb{P}_{\nu_k}[\mathbf{X}_0\neq\mathbf{Y}_0]\\&=\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=0}^{n_k-1}\mathbb{P}_{\nu'_k}\left[T^{-i}\{\mathbf{X}_0\neq\mathbf{Y}_0\}\right]\\&=\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=0}^{n_k-1}\mathbb{P}_{\nu'_k}[\mathbf{X}_i\neq\mathbf{Y}_i]\\&=\lim_{k\to\infty}\mathbb{E}_{\nu'_k}\left[\frac{1}{n_k}\sum_{i=0}^{n_k-1}I_{\mathbf{X}_i\neq\mathbf{Y}_i}\right]\\&=\lim_{k\to\infty}\mathbb{E}_{\nu'_k}\left[\frac{1}{n_k}d\big(\mathbf{Y}_0^{n_k-1},\mathbf{X}_0^{n_k-1}\big)\right]\\&=\lim_{k\to\infty}\mathbb{E}_\mu\left[\frac{1}{n_k}d\big(\mathbf{Y}_0^{n_k-1},\overline{x}(\mathbf{Y}_0^{n_k-1})\big)\right].\end{aligned}$$

We define E_δ to be the event that $d(\mathbf{Y}_0^{n_k-1},\overline{x}(\mathbf{Y}_0^{n_k-1}))\geqslant n_k(R_\varepsilon(X,Y,\mu)+\delta/4)$. By (17), for sufficiently large k,

$$\begin{aligned}&\mathbb{E}_\mu\left[\frac{1}{n_k}d\big(\mathbf{Y}_0^{n_k-1},\overline{x}(\mathbf{Y}_0^{n_k-1})\big)\right]\\&\leqslant\mathbb{E}_\mu\left[\frac{1}{n_k}d\big(\mathbf{Y}_0^{n_k-1},\overline{x}(\mathbf{Y}_0^{n_k-1})\big)\cdot I_{E_\delta}\right]+\mathbb{E}_\mu[1-I_{E_\delta}]\\&\leqslant R_\varepsilon(X,Y,\mu)+\frac{\delta}{4}+(1-\mathbb{P}_\mu[E_\delta])\\&\leqslant R_\varepsilon(X,Y,\mu)+\frac{\delta}{4}+\varepsilon\leqslant R_\varepsilon(X,Y,\mu)+\frac{\delta}{2}.\end{aligned}$$

Combining the above inequality with (16) we conclude that

$$\mathbb{P}_\nu[\mathbf{X}_0\neq\mathbf{Y}_0]=\lim_{k\to\infty}\mathbb{P}_{\nu_k}[\mathbf{X}_0\neq\mathbf{Y}_0]\leqslant R_0(X,Y,\mu)+\frac{3\delta}{4},$$

as desired. ■

In the following example, we explicitly describe a sequence of extensions in $M_{\mathcal{E}}(X,Y,\mu)$ which approximates the essential covering radius of the $(0,k)$-RLL system from Example 1 with respect to the full-shift (equipped with the uniform Bernoulli measure).

Example 4: Let $X_{0,k}\subseteq[2]^{\mathbb{Z}}$ denote the $(0,k)$-RLL shift as in Example 1. Let $\overline{y}\in[2]^n$ be a finite binary word. We define $c(\overline{y})$ to be the length of longest zero suffix of $\overline{y}$, formally given by

$$c(\overline{y})\triangleq\max\Big\{i\ :\ \overline{y}=\overline{y}_0^{n-i-1}\overline{0}^i\Big\}.$$

We fix $N\in\mathbb{N}$ and consider the map $f^{(N)}:[2]^{\mathbb{Z}}\to X_{0,k}$ defined by

$$f^{(N)}(\mathbf{y})_m=\begin{cases}1 & c\Big(\mathbf{y}_{m-(N(k+1)-1)}^{m-1}\Big)\equiv k\pmod{k+1},\\ \mathbf{y}_m & \textit{otherwise.}\end{cases}$$

Clearly, $\mathrm{Im}(f)\subseteq X_{0,k}$ since no run of $k+1$ zeroes may appear in $f^{(N)}(\mathbf{y})$. We note that the map $(f^{(N)},\mathrm{Id}):[2]^{\mathbb{Z}}\to X_{0,k}\times[2]^{\mathbb{Z}}$ is a sliding-block-code function (i.e., a function such that the value in each coordinate is determined by a finite block of adjacent coordinates), and therefore it is measurable and commutes with the shift transformation. Let μ^{u} be the uniform measure over $[2]^{\mathbb{Z}}$, and let ν_N be its pushforward measure on $X_{0,k}\times[2]^{\mathbb{Z}}$ using $f^{(N)}$. Clearly ν_N is an invariant *measure, which is also ergodic (as a factor of an ergodic measure). Therefore, $\nu_N \in M_{\mathcal{E}}(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}})$. We note that*

$$\begin{aligned}
&\mathbb{P}_{\nu_N}[\mathbf{X}_0 \neq \mathbf{Y}_0] \\
&= \mathbb{P}_{\mu^{\mathsf{u}}}\left[c\left(\mathbf{Y}_{-(N(k+1)-1)}^{-1}\right) \equiv k \pmod{k+1} \text{ and } \mathbf{Y}_0 = 0\right] \\
&= \sum_{i=0}^{N-1} \mathbb{P}_{\mu^{\mathsf{u}}}\left[c\left(\mathbf{Y}_{-(N(k+1)-1)}^{-1}\right) = i(k+1)+k \text{ and } \mathbf{Y}_0 = 0\right] \\
&= \mathbb{P}_{\mu^{\mathsf{u}}}\left[\mathbf{Y}_{-(N(k+1)-1)}^{0} = \overline{0}^{N(k+1)}\right] \\
&\quad + \sum_{i=1}^{N-1} \mathbb{P}_{\mu^{\mathsf{u}}}\left[\mathbf{Y}_{-i(k+1)}^{0} = 1\overline{0}^{i(k+1)}\right] \\
&= \frac{1}{2^{N(k+1)}} + \frac{1}{2}\sum_{i=1}^{N-1} \frac{1}{2^{i(k+1)}}.
\end{aligned}$$

Taking $N \to \infty$ we obtain

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{P}_{\nu_N}[\mathbf{X}_0 \neq \mathbf{Y}_0] &= \lim_{N\to\infty} \frac{1}{2^{N(k+1)}} + \frac{1}{2}\sum_{i=1}^{N-1} \frac{1}{2^{i(k+1)}} \\
&= \frac{1}{2(2^{k+1}-1)} = R_0(X_{0,k}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}).
\end{aligned}$$

To conclude this section we briefly discuss the essential covering radius in the context of the QCC scheme. Loosely speaking, asymptotically, all but a vanishing fraction of Σ^n may be quantized to $\mathcal{B}_n(X)$ by changing an $R_0(X, \Sigma^{\mathbb{Z}}, \mu^{\mathsf{u}})$-fraction of the positions. This fraction may be significantly lower than the worst-case fraction $R(X, \Sigma^{Z})$. In a finite-length setting, at least a $(1-\varepsilon)$-fraction of Σ^n may be quantized to $\mathcal{B}_n(X)$ by changing at most $r_\varepsilon = R_\varepsilon(\mathcal{B}_n(X), \Sigma^n, \mu_n^{\mathsf{u}})$ positions. A small obstacle we need to overcome is the fact that in the QCC scheme we do not quantize any word from Σ^n, but rather only codewords of the error-correcting code C. The ε-fraction of words from Σ^n that are a long distance from $\mathcal{B}_n(X)$ may disproportionately reside in C. However, if we further assume that C is a linear error-correcting code, by a simple averaging argument there exists at least one coset of the code, C', such that the fraction of codewords whose distance to the language of X is at most r_ε. This means that there exists $C'' \subseteq C'$ with $|C''| \geqslant (1-\varepsilon)|C'|$ such that $R(B_n(X), C'') \leqslant r_\varepsilon$.

V. Upper Bounds On The Essential Covering Radius

The goal of this section is to establish general upper bounds on the essential covering radius. While Theorem 19 gives an exact expression for the essential covering radius, the minimization problem involved is hard to solve. In Example 4, we found a sequence of good extensions which approximates the essential covering radius. In general, by Theorem 19, such a sequence of extensions provides a sequence of upper-bounds on the essential covering radius. In this section we shall present two different approaches for constructing extensions for general constrained systems, thus providing upper bounds on the essential covering radius. The first approach, using Markov chains, provides an upper bound which is efficiently computable as the solution of a linear-programming problem. An alternative method for constructing extensions is by sliding-block-code functions. In that case, we prove that if X is primitive, the essential covering radius can be approximated by increasing the block size in such functions.

A. Markov Chains

We consider the scenario where X and Y are constrained systems generated by labeled graphs $G_X = (V_X, E_X, L_X)$ and $G_Y = (V_Y, E_Y, L_Y)$ respectively. Throughout this section, we assume that G_X and G_Y contain no parallel edges with the same label. An edge $u \to v$ shall be denoted by the ordered pair $e = (u, v)$, and we say its source is $\sigma(e) = u$ and its target is $\tau(e) = v$. We focus on the case where the measure $\mu \in M_{\mathcal{E}}(Y)$ is generated by some Markov chain on the graph G_Y. We remark that the case of $Y = \Sigma^{\mathbb{Z}}$ and $\mu = \mu^{\mathsf{u}}$ is the uniform Bernoulli measure, falls into that category. We begin with some definitions and basic results from the theory of Markov chains on finite graphs.

Definition 20: Let $G = (V, E)$ be a finite directed graph. A stationary Markov chain on G is a pair (π, Q), where π is a probability measure on V and Q is a function from V to the space of probability measures on E that sends $v \in V$ to a probability measure $Q(\cdot|v)$ on E such that for every $v \in V$,

$$\sum_{\substack{e\in E\\ \sigma(e)=v}} Q(e|v) = 1,$$

and so that for every $v \in V$ we have:

$$\pi(v) = \sum_{\substack{e\in E\\ \tau(e)=v}} \pi(\sigma(e)) Q(e|\sigma(e)).$$

Note that for any Markov chain on (π, Q) on $G = (V, E)$, $Q(e|v) > 0$ implies that $\sigma(e) = v$ so we can conveniently write $Q(e)$ as an abbreviation for $Q(e|\sigma(e))$. In the case where G is a simple graph (i.e., without parallel edges), for any edge $e = (u, v) \in E$ we use the notation $Q(v|u)$ for $Q(e)$. Also, when G is a simple graph, Q may be identified with a $|V| \times |V|$ stochastic matrix (often called the transition matrix), for which π is a left eigenvector with eigenvalue 1.

There is a one-to-one correspondence between Markov chains on $G = (V, E)$ and probability measures on E that satisfy the condition

$$\sum_{\substack{e\in E\\ \sigma(e)=v}} P(e) = \sum_{\substack{e\in E\\ \tau(e)=v}} P(e).$$

Indeed, such a probability measure P corresponds to a stationary Markov chain (π, Q), where

$$\pi(v) \triangleq \sum_{\substack{e\in E\\ \sigma(e)=v}} P(e) = \sum_{\substack{e\in E\\ \tau(e)=v}} P(e),$$

and

$$Q(e|v) \triangleq \frac{P(e)}{\pi(v)}.$$

By abuse of notation, we denote $P = (\pi, Q)$. We assume the Markov chain does not contain degenerate vertices, i.e.,

$\pi(v) > 0$ for all $v \in V$. Any stationary Markov chain P induces an invariant measure $\hat{P}$ on the space of bi-infinite paths on G by

$$\hat{P}([(e_0, e_1, \ldots, e_{n-1})]) = P(e_0) \prod_{i=1}^{n-1} Q(e_i).$$

for any cylinder set $[(e_0, \ldots, e_{n-1})]$ corresponding to a finite path $(e_0, \ldots, e_{n-1})$, which by Definition 1 is the set

$$[(e_0, \ldots, e_{n-1})] = \left\{ \mathbf{e} \in E^{\mathbb{Z}} \; : \; \mathbf{e}_0^{n-1} = (e_0, \ldots, e_{n-1}) \right\}.$$

We call $\hat{P}$ the stationary Markov process on G, induced by P. In the case where G is a simple graph (i.e., without parallel edges), for any edge $e = (u, v) \in E$ we use the notation $Q(v|u)$ for $Q(e)$.

If $G = G_Y$ generates the constrained system Y by the labeling function L_Y, then $P = (\pi, Q)$ induces an invariant probability measure on Y, which is the pushforward measure of $\hat{P}$ via the labeling function, i.e., for a cylinder set $[\overline{y}]$,

$$\mu_P([\overline{y}]) = \sum_{\substack{\overline{e} \text{ path in } G \\ L(\overline{e}) = \overline{y}}} \pi(\sigma(e_0)) \prod_{i=0}^{|\overline{y}|-1} Q(e_i).$$

We note that $\mathbf{Y}$ is a hidden Markov process with respect to μ_P, and refer to the measure μ_P as above as *the hidden Markov measure* induced by P via the labeling function L.

Assume that $X, Y \subseteq \Sigma^{\mathbb{Z}}$ are irreducible constrained systems given by labeled graphs G_X and G_Y respectively, and assume that $\mu = \mu_{P_Y} \in M_{\mathcal{E}}(Y)$ is a measure on Y, induced by P_Y, a stationary Markov Chain on G_Y. We consider the strong product graph of G_X and G_Y given by $G_{X\times Y} = (V_{X\times Y}, E_{X\times Y}, (L_X, L_Y))$ where $V_{X\times Y} \triangleq V_X \times V_Y$, $E_{X\times Y} \triangleq E_X \times E_Y$ with $\sigma(e_x, e_y) = (\sigma(e_x), \sigma(e_y))$, $\tau(e_x, e_y) = (\tau(e_x), \tau(e_y))$ and labeling function $L_{X\times Y}$ given by:

$$L_{X\times Y}(e_x, e_y) = (L_X(e_x), L_Y(e_y)).$$

We note that a stationary Markov chain P on $G_{X\times Y}$ naturally defines a stationary Markov process $\hat{P}$ on $G_{X\times Y}$, which induces the hidden Markov measure ν_P on $X \times Y$ by the labeling function $L_{X\times Y}$. Since each edge $e \in E_{X\times Y}$ is composed of a pair $(e_x, e_y) \in E_X \times E_Y$, $\hat{P}$ may be considered as a measure on the space of pairs of bi-infinite paths where the first is a path on G_X and the second is a path on G_Y. We denote the marginal measure of $\hat{P}$ on the space of bi-infinite paths on G_Y by $(\hat{P})_Y$

For our purpose, we are interested in stationary Markov chains on $G_{X\times Y}$ such that the Y-marginal of the induced measure is μ_{P_Y}. A sufficient condition is the following: considering $\hat{P}$ as a stationary Markov process on $G_{X\times Y}$, the G_Y-marginal measure of $\hat{P}$ is the Markov process $\hat{P}_Y$. We are therefore interested in the following question: given a stationary Markov chain P on $G_{X\times Y}$, when is $\hat{P}$ an extension of the measure $\hat{P}_Y$? An obvious necessary condition is that the G_Y-marginal of pairs in the distribution of $\hat{P}$ equals the pairs distribution of $\hat{P}_Y$. This condition may equivalently be described by

$$P_Y(e) = \sum_{\substack{e' \in E_{X\times Y} \\ e'_y = e}} P(e'), \quad \text{for all } e \in E_Y. \tag{18}$$

It is tempting to speculate that the condition given in (18) is also sufficient, however, it is not the case. The marginal of a Markov measure, in general, is not necessarily a Markov measure, but rather a hidden Markov measure. In fact, there exists an example for stationary Markov chains P and P_Y satisfying (18) such that there is no invariant measure on the space of bi-infinite paths on $G_{X\times Y}$ which has the same pairs distribution as $\hat{P}$ and an G_Y-marginal that equals to the Markov process $\hat{P}_Y$. In the following lemma, we propose a sufficient condition under which the G_Y-marginal of $\hat{P}$ equals $\hat{P}_Y$. For ease of notation, we state and prove the claim for simple graphs. However, the claim is true without the assumption of graph simplicity, and the proof generalizes immediately from simple graphs to general labeled graphs. The corresponding condition for the general (not necessarily simple) case is given in (22).

Lemma 21: Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be finite simple directed graphs and let $G_1 \times G_2 = G = (V, E)$ denote strong graph product of G_1 and G_2 (as defined above). For given stationary Markov chains $P = (\pi, Q)$ and $P_1 = (\pi_1, Q_1)$ on G and G_1 respectively, the G_1-marginal of the Markov process $\hat{P}$ is the Markov process $\hat{P}_1$ (on G_1) if the following conditions hold:

- *For all $e \in E_1$*
$$P_1(e) = \sum_{\substack{e' \in E \\ e'_1 = e}} P(e') \tag{19}$$
- *For all $u_0 \in V_2$ and $v_0, v_1 \in V_1$ we have*
$$\begin{aligned} Q(v_1|v_0, u_0) &\triangleq \sum_{u_1 \in V_2} Q((v_1, u_1)|(v_0, u_0)) \\ &= Q_1(v_1|v_0). \end{aligned} \tag{20}$$

Proof: The proof is a straightforward calculation. Let us denote by $(\hat{P})_1$ the marginal measure of $\hat{P}$ on G_1. By the law of total probability, for any path $\overline{e} = (e_0, \ldots, e_{n-1}) = ((v_0, v_1), (v_1, v_2), \ldots, (v_{n-1}, v_n))$ on G_1,

$$\begin{aligned} (\hat{P})_1([\overline{e}]) &= \sum_{u_0, \ldots, u_n \in V_2} P((v_0, u_0), (v_1, u_1)) \\ &\quad \cdot \prod_{i=1}^{n-1} Q((v_{i+1}, u_{i+1})|(v_i, u_i)) \\ &= \sum_{u_0, u_1 \in V_2} P((v_0, u_0), (v_1, u_1)) \\ &\quad \cdot \prod_{i=1}^{n-1} \left(\sum_{u_{i+1} \in V_2} Q((v_{i+1}, u_{i+1})|(v_i, u_i)) \right) \\ &\overset{(a)}{=} \sum_{u_0, u_1 \in V_2} P((v_0, u_0), (v_1, u_1)) \prod_{i=1}^{n-1} Q_1(v_{i+1}|v_i) \\ &\overset{(b)}{=} P_1(v_0, v_1) \prod_{i=1}^{n-1} Q_1(v_{i+1}|v_i) = \hat{P}_1([\overline{e}]), \end{aligned}$$

where (a) and (b) follow from the assumptions (19) and (20) respectively. ∎

We are now ready to give an upper bound on $R_0(X,Y,\mu)$, formulated as an optimization problem over stationary Markov chains. Let P_Y and P be stationary Markov chains on G_Y and $G_{X\times Y}$ respectively such that P satisfies the conditions of Lemma 21 with respect to P_Y. Let us denote the measure on $X\times Y$ induced by P by ν_P, and denote the measure on Y induced by P_Y by μ. By Lemma 21 we have $\nu_P\in M(X,Y,\mu)$ and therefore by Theorem 19 we have

$$R_0(X,Y,\mu)\leqslant \mathbb{P}_{\nu_P}[\mathbf{X}_0\neq\mathbf{Y}_0]=\sum_{\substack{e=E_{X\times Y}\\ L_X(e)\neq L_Y(e)}}P(e).$$

We note that if we consider the Markov chain P as an element in the simplex contained in $[0,1]^{E_X\times E_Y}$, we have that $\nu_P[\mathbf{X}_0\neq\mathbf{Y}_0]$ is a linear function of P. In addition, we note the condition (19) is linear in P as a constraint defined by a sum over the elements of P. In the case of simple graphs a straightforward calculation shows that (20) is also a linear condition, since $Q(v_1^y|v_0^x,v_0^y)=Q(v_1^y|v_0^y)$ if and only if

$$\sum_{v_1^x\in V_X}P((v_1^x,v_1^y),(v_0^x,v_0^y)) = Q_Y(v_1^y|v_0^y)\cdot\sum_{\substack{v^x\in V_X\\ v^y\in V_Y}}P((v^x,v^y),(v_0^x,v_0^y)) \tag{21}$$

We observe that (21) may be equivalently formulated as a condition on edges, under which the conclusion of Lemma 21 is true in the general case. That is, the conclusion of of Lemma 21 is true if (19) is satisfied and for all $e\in E_{X\times Y}$

$$\sum_{\substack{e'\in E_{X\times Y}\\ e'_y=e_y\\ \sigma(e'_x)=\sigma(e_x)}}P(e')=Q_Y(e_y)\sum_{\substack{e'\in E_{X\times Y}\\ \sigma(e'_y)=\sigma(e_y)\\ \sigma(e'_x)=\sigma(e_x)}}P(e') \tag{22}$$

We therefore obtain an upper bound by minimizing $\mathbb{P}_{\nu_P}[\mathbf{X}_0\neq\mathbf{Y}_0]$ over all stationary Markov chains on $G_{X\times Y}$ satisfying (19) and (22), which turns out to be a linear-programming problem.

Theorem 22: Let $X,Y\subseteq\Sigma^{\mathbb{Z}}$ be shift spaces defined by the labeled graphs G_X and G_Y respectively, and let P_Y be a stationary Markov chain on Y that induces the measure μ. Then

$$R_0(X,Y,\mu)\leqslant \mathrm{MB}(G_X,G_Y,P_Y),$$

where $\mathrm{MB}(G_X,G_Y,P_Y)$ is the solution to the following linear-programming problem:

$$\begin{aligned}
&\underset{P\in\mathbb{R}^{E_{X\times Y}}}{\text{minimize}} && \sum_{\substack{e\in E_{X\times Y}\\ L_X(e)\neq L_Y(e)}}P(e)\\
&\text{subject to} && P(e)\geqslant 0, && \forall e\in E_{X\times Y},\\
& && \sum_{e\in E_{X\times Y}}P(e)=1,\\
& && \sum_{\substack{e'\in E_{X\times Y}\\ e'_y=e}}P(e')=P_Y(e), && \forall e\in E_Y,\\
& && \sum_{\substack{e\in E_{X\times Y}\\ \sigma(e)=v}}P(e)=\sum_{\substack{e\in E_{X\times Y}\\ \tau(e)=v}}P(e), && \forall v\in V_{X\times Y},\\
& && \sum_{\substack{e'\in E_{X\times Y}\\ e'_y=e_y\\ \sigma(e'_x)=\sigma(e_x)}}P(e')=Q_Y(e_y)\sum_{\substack{e'\in E_{X\times Y}\\ \sigma(e'_y)=\sigma(e_y)\\ \sigma(e'_x)=\sigma(e_x)}}P(e') && \forall e\in E_{X\times Y}.
\end{aligned}$$

Example 5: Consider $X_{0,k}$, the $(0,k)$-RLL shift from Example 1. Take $Y=[2]^{\mathbb{Z}}$ and let μ^{u} be the uniform Bernoulli measure on Y. We consider the labeled graphs G_X and G_Y shown in Figure 2, generating X and Y respectively. We take P_Y to be the uniform measure on E_Y (inducing the measure μ^{u}). The product graph, $G_{X\times Y}$ (shown in Figure 3), is therefore a "doubled" version of the graph G_X.

We consider the Markov measure P, defined by the edge probabilities given in Figure 3. For an appropriate choice of α, P is indeed a stationary Markov chain satisfying (19) and (22). First, in order to get a probability measure on edges we require

$$\begin{aligned}1&=\sum_{e\in G_{X\times Y}}P(e)=2\alpha(2^k+2^{k-1}+\cdots+2+1)\\ &=2\alpha(2^{k+1}-1),\end{aligned}$$

which implies that $\alpha=\frac{1}{2(2^{k+1}-1)}$. We observe that for the j-th state in $V_{X\times Y}$,

$$\sum_{\sigma(e)=j}P(e)=\sum_{\tau(e)=j}P(e)=\alpha\cdot 2^{k-j},$$

which implies that P is indeed stationary. We also observe that for any edge with $L_Y(e)=0$ there is a corresponding edge e' with $L_Y(e')=1$ such that $P(e)=P(e')$. This shows that the marginal of P on G_Y is indeed $P_Y=\mu^{\mathsf{u}}$. We further observe that for any edge $e\in E_{X\times Y}$, if $\sigma(e_x)=j\in V_X$ we have

$$\begin{aligned}\sum_{\substack{e'\in E_{X\times Y}\\ e'_y=e_y\\ \sigma(e'_x)=j}}P(e')&=\alpha 2^{k-j-1}=\frac{1}{2}\alpha 2^{k-j}\\ &=Q_Y(e_y)\sum_{\substack{e'\in E_{X\times Y}\\ \sigma(e'_y)=\sigma(e_y)\\ \sigma(e'_x)=j}}P(e').\end{aligned}$$

We now compute:

$$\sum_{\substack{e\in G_{X\times Y}\\ L_X(e)\neq L_Y(e)}}P(e)=\alpha=\frac{1}{2(2^{k+1}-1)}=R_0(X_{0,k},[2]^{\mathbb{Z}},\mu^{\mathsf{u}}).$$

Thus, by Theorem 14 and Theorem 22, P attains the minimal value for the linear-programming problem $\mathrm{MB}(G_X,G_Y,P_Y)$, and in particular in this case, the upper bound from Theorem 22 is tight.

Example 6: Let $X=X_{d,\infty}$ be the (d,∞)-RLL system, defined by the constraint of having a run of at least d zeroes between any two consecutive ones. Equivalently, $X_{d,\infty}$ is defined by G_X presented in Figure 4. Let $Y=[2]^{\mathbb{Z}}$ and μ^{u} be as in Example 1. The product graph $G_{X\times Y}$ is shown in Figure 5. We consider the Markov measure P, defined by the

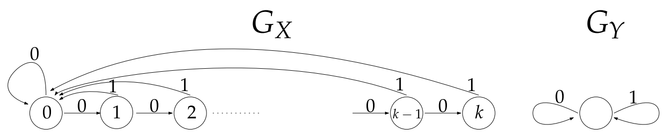

Fig. 2. Labeled graphs generating the shift spaces $X = X_{0,k}$ (the $(0,k)$-RLL shift), and $Y = [2]^{\mathbb{Z}}$.

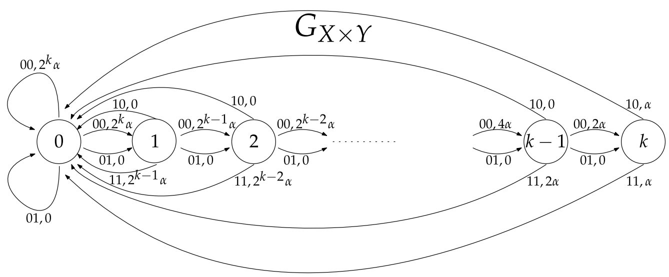

Fig. 3. The product graph $G_{X\times Y}$ for the graphs G_X and G_Y from Figure 2. Each edge is given a two-bit label, xy, corresponding to the label x from G_X and the label y from G_Y. A stationary Markov chain achieving the bound $MB(G_X, G_Y, P_Y)$ is shown by writing $P(e)$ after the label on each edge.

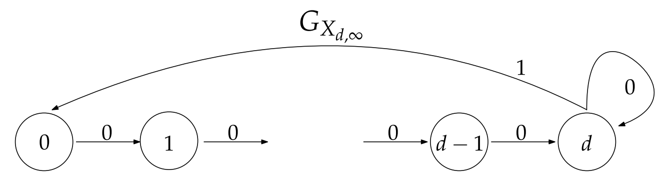

Fig. 4. A labeled graph generating the constrained system $X = X_{d,\infty}$ (the (d,∞)-RLL shift).

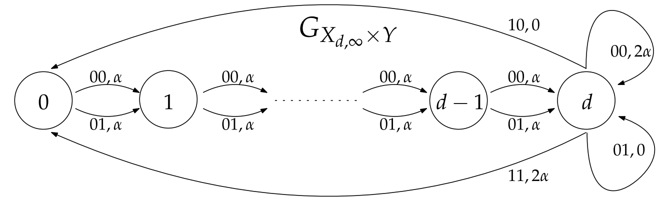

Fig. 5. The product graph $G_{X\times Y}$ for the graphs $G_{X_{d,\infty}}$ and G_Y from Figure 4 and Figure 2 respectively. Each edge is given a two-bit label, xy, corresponding to the label x from $G_{X_{d,\infty}}$ and the label y from G_Y. A stationary Markov chain achieving the bound $MB(G_X, G_Y, P_Y)$ is shown by writing $P(e)$ after the label on each edge.

edge probabilities given in Figure 5. For $\alpha = (2(d+2))^{-1}$ we have

$$1 = \sum_{e\in G_{X\times Y}} P(e) = 2\alpha(d+2),$$

and so P is indeed a stationary Markov chain satisfying the conditions of Theorem 22. We now compute the upper-bound:

$$R_0(X_{d,\infty}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \leqslant \sum_{\substack{e\in G_{X\times Y}\\ L_X(e)\neq L_Y(e)}} P(e) = \alpha\cdot d = \frac{d}{2(d+2)}. \tag{23}$$

We note that when $d = 1$, the system $X_{1,\infty}$ is isomorphic to $X_{0,1}$, by complementing all the bits. Thus

$$R_0(X_{1,\infty}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = R_0(X_{0,1}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) = \frac{1}{6},$$

and in particular, the bound (23) is tight.

For a lower bound, we claim that

$$R_0(X_{d,\infty}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant \frac{1}{2} - \frac{1}{d+1}.$$

The proof is straight-forward from the definition of the essential covering radius. We note that for $d = 1$ the bound is meaningless, and therefore we assume that $d \geqslant 2$. We start by a simple observation. Let $\mathrm{wt}(\overline{w}) = d(\overline{w}, \overline{0})$ denote the number of non-zero coordinates of a binary word $\overline{w}$. We note that for any $\overline{x}, \overline{y} \in [2]^n$,

$$d(\overline{x}, \overline{y}) \geqslant |\mathrm{wt}(\overline{y}) - \mathrm{wt}(\overline{x})| \geqslant \mathrm{wt}(\overline{y}) - \mathrm{wt}(\overline{x}).$$

From the definition of the $X_{d,\infty}$ system, for any any $\overline{x} \in \mathcal{B}_n(X)$, there must be at least d zeroes between any two consecutive ones. Hence, for any $\overline{x} \in \mathcal{B}_n(X)$ we have

$$\mathrm{wt}(\overline{x}) \leqslant \left\lceil \frac{n}{d+1} \right\rceil \leqslant \frac{n}{d+1} + 1.$$

Thus, for any word $\overline{y} \in [2]^n$, we have

$$\min_{\overline{x}\in\mathcal{B}_n(X)} d(\overline{x}, \overline{y}) \geqslant \mathrm{wt}(\overline{y}) - \frac{n}{d+1} - 1. \tag{24}$$

Let $\mathbf{Y}$ be a random bi-infinite sequence generated by the distribution μ^{u}. That is, $(\mathbf{Y}_k)_{k\in\mathbb{Z}}$ are i.i.d $\mathrm{Ber}(\frac{1}{2})$ random variables. Applying the law of large numbers, we have

$$\frac{1}{n}\,\mathrm{wt}\big(\mathbf{Y}_0^{n-1}\big) = \frac{1}{n}\sum_{i=0}^{n-1} \mathbf{Y}_i \xrightarrow[n\to\infty]{\mathbb{P}_{\mu^{\mathsf{u}}}} \mathbb{E}_{\mu^{\mathsf{u}}}[\mathbf{Y}_0] = \frac{1}{2}.$$

Namely, the normalized weights of random words converge in $\mathbb{P}_{\mu^{\mathsf{u}}}$ probability to $\frac{1}{2}$. Combining this with (24), we obtain that for every $\delta > 0$

$$\begin{aligned} &\mathbb{P}_{\mu_n^{\mathsf{u}}}\left[\min_{\overline{x}\in\mathcal{B}_n(X_{d,\infty})} \frac{d(\overline{x}, \mathbf{Y}_0^{n-1})}{n} > \frac{1}{2} - \frac{1}{d+1} - \frac{1}{n} - \delta\right] \\ &\quad\geqslant \mathbb{P}_{\mu^{\mathsf{u}}}\left[\frac{1}{n}\,\mathrm{wt}(\mathbf{Y}_0^{n-1}) > \frac{1}{2} - \delta\right] \xrightarrow[n\to\infty]{} 1. \end{aligned}$$

This proves that for any $\varepsilon \in (0,1)$, $\delta > 0$, for sufficiently large n

$$\frac{R_\varepsilon(\mathcal{B}_n(X_{d,\infty}), \mathcal{B}_n(Y), \mu^{\mathsf{u}})}{n} > \frac{1}{2} - \frac{1}{d+1} - \delta,$$

and therefore,

$$R_\varepsilon(X_{d,\infty}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant \lim_{\delta\to 0} \frac{1}{2} - \frac{1}{d+1} - \delta = \frac{1}{2} - \frac{1}{d+1}.$$

Taking $\varepsilon \to 0$, we get

$$R_0(X_{d,\infty}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \geqslant \frac{1}{2} - \frac{1}{d+1}.$$

Combining the lower and upper bounds we get,

$$\frac{1}{2} - \frac{1}{d+1} \leqslant R_0(X_{d,\infty}, [2]^{\mathbb{Z}}, \mu^{\mathsf{u}}) \leqslant \frac{1}{2} - \frac{1}{d+2}.$$

B. Sliding Block Codes

We now present an alternative approach for constructing extensions by using sliding-block-codes functions. We begin by revising Example 4, where we found a sequence of extensions approximating the essential covering radius of $X_{0,k}$ with respect to $[2]^{\mathbb{Z}}$ with μ^{u}, (the uniform i.i.d measure). The main idea in the construction of these extensions was the following:

for $X, Y \subseteq \Sigma^{\mathbb{Z}}$ and $\mu \in M_{\mathcal{E}}(Y)$, given a measurable function $g : Y \to X$ which commutes with the shift transformation, the map $(g, \mathrm{Id}) : Y \to X \times Y$ defines an extension ν_g in $M(X, Y, \mu)$ by the pushforward of μ via (g, Id). That is

$$\nu_g(A_X \times A_Y) \triangleq \mu\big(A_Y \cap g^{-1}(A_X)\big).$$

We call such a function g a *stationary coding function* from Y to X. By Proposition 16, for any such stationary coding function g,

$$R_0(X, Y, \mu) \leqslant \mathbb{P}_{\nu_g}[\mathbf{X}_0 \neq \mathbf{Y}_0] = \mathbb{P}_{\mu}[g(\mathbf{Y})_0 \neq \mathbf{Y}_0],$$

Namely, any such measurable function g that commutes with the shift provides an upper bound on the essential covering radius.

In Example 4, we used sliding-block-code functions (functions defined by a local rule) as our measurable functions that commute with the shift. Sliding-block-code functions are of particular interest to us since they provide a rich family of functions, easily described by a local rule. The properties and constructions of sliding-block codes have been extensively studied in the literature (for example, see [1], [3], [4]). The goal of this section is to explicitly describe the bound obtained from a sliding-block-code function, and to give sufficient conditions under which the essential covering radius can be approximated using extensions constructed by sliding-block codes.

Definition 23: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be shift spaces. A function $\hat{f} : Y \to X$ is called a sliding-block code *if there exist $N \in \mathbb{N}$ and a function $f : \mathcal{B}_{2N+1}(Y) \to \Sigma$ such that for all $\mathbf{y} \in Y$ and all $i \in \mathbb{Z}$,*

$$\hat{f}(\mathbf{y})_i = f(\mathbf{y}_{i-N}^{i+N}).$$

In that case, $\hat{f}$ is said to be a sliding-block code of block length N.

Let $\hat{f} : Y \to X$ be a sliding-block-code function defined by a local function $f : \mathcal{B}_{2N+1}(Y) \to \Sigma$, and let $\mu \in M_{\mathcal{E}}(Y)$ be an ergodic measure. We denote the extension obtained from $\hat{f}$ by ν_f. The quantity $\mathbb{P}_{\nu_f}[\mathbf{X}_0 \neq \mathbf{Y}_0]$ is now easily computable:

$$\mathbb{P}_{\nu_f}[\mathbf{X}_0 \neq \mathbf{Y}_0] = \mathbb{P}_{\mu}[\hat{f}(\mathbf{Y})_0 \neq \mathbf{Y}_0] = \mathbb{P}_{\mu}\big[f(\mathbf{Y}_{-N}^{N}) \neq \mathbf{Y}_0\big] = \sum_{\substack{\overline{y} \in \mathcal{B}_{2N+1}(Y) \\ f(\overline{y}) \neq \overline{y}_N}} \mu([\overline{y}]).$$

We recall the notion of aperiodicity for ergodic measures:

Definition 24: An ergodic measure $\mu \in M_{\mathcal{E}}(Y)$ is called aperiodic *if the measure of periodic points is 0. Namely,*

$$\mu(\{\mathbf{y} \in Y \ : \ \exists n \in \mathbb{N} \text{ such that } T^n\mathbf{y} = \mathbf{y}\}) = 0.$$

We now state the main result of this section: if X is a primitive constrained system and μ is aperiodic then the essential covering radius $R_0(X, Y, \mu)$ may be approximated by extensions constructed by sliding-block codes, as the block length is increased. We note that the uniform i.i.d measure is aperiodic, meaning that the conclusion of Theorem 25 is true in the case of chief interest where $(Y, \mu) = (\Sigma^{\mathbb{Z}}, \mu^{\mathsf{u}})$.

Theorem 25: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be constrained systems such that X is primitive and $\mu \in M_{\mathcal{E}}(Y)$ is an aperiodic ergodic measure. Then for any $\varepsilon > 0$ there exists a sufficiently large N and a sliding-block-code function $\hat{f}$ of length N such that

$$\mathbb{P}_{\mu}[g(\mathbf{Y})_0 \neq \mathbf{Y}_0] - \varepsilon \leqslant R_0(X, Y, \mu) \leqslant \mathbb{P}_{\mu}[g(\mathbf{Y})_0 \neq \mathbf{Y}_0].$$

The proof of Theorem 25 has two major components. The first part shows that it is possible to approximate the essential covering radius by extensions obtained from stationary coding functions. In the second step, we use the fact that sliding-block-code functions are dense in the space of stationary coding functions to obtain the main result.

For the first component of the proof, we a version of a standard result in ergodic-theory known as Alpern's Lemma [2]. The precise version we use appears in [10].

Lemma 26 ([10, Theorem 1]): Let $Y \subseteq \Sigma^{\mathbb{Z}}$ be a constrained system, $\mu \in M_{\mathcal{E}}(Y)$ be an aperiodic ergodic measure, $n_1, \ldots, n_k$ be integers whose greatest common divisor is 1, and $q_1 \ldots, q_k$ positive numbers such that $\sum_{i=1}^{k} q_i n_i = 1$. For any finite measurable partition $\mathcal{P}$ of Y there exist measurable sets $Q_1, \ldots, Q_k$ such that:

1) *The set*

$$\mathcal{P}' = \big\{T^{-i}(Q_j) \ : \ 1 \leqslant j \leqslant k, 0 \leqslant i \leqslant n_j - 1\big\}$$

is a partition of Y.

2) *For all $1 \leqslant j \leqslant k$, $\mu(Q_j) = q_j$.*
3) *For all $1 \leqslant j \leqslant k$, Q_j is independent of the partition $\mathcal{P}$ with respect to μ (namely, for any $A \in \mathcal{P}$ $\mu(A \cap Q_j) = \mu(A)\mu(Q_j)$).*

Proposition 27: Let $X, Y \subseteq \Sigma^{\mathbb{Z}}$ be constrained systems, X be primitive, and $\mu \in M_{\mathcal{E}}(Y)$ be an aperiodic ergodic measure. Then

$$R_0(X, Y, \mu) = \inf \big\{\mathbb{P}_{\mu}[g(\mathbf{Y})_0 \neq \mathbf{Y}_0] \ : \ g : Y \to X \text{ is a stationary coding function}\big\}.$$

Proof: We start by outlining the proof strategy. By Theorem 19, $\mathbb{P}_{\mu}[g(\mathbf{Y})_0 \neq \mathbf{Y}_0] \geqslant R_0(X, Y, \mu)$ for every stationary coding function $g : Y \to X$. It is therefore sufficient to prove the other opposite inequality by showing that for any $\varepsilon > 0$ there exists a stationary coding function $g : Y \to X$ such that the probability of the event $\{g(\mathbf{Y})_0 \neq \mathbf{Y}_0\}$ is at most $R_0(X, Y, \mu) + \varepsilon$. We construct such a stationary coding function using "block coding": Using Lemma 26, we partition an infinite sequence from Y into blocks of two types, one large and the other of a fixed length. The large blocks are common while the other type occurs only rarely. The large blocks we code (with high probability) to a close counterpart from the language of X, such that the whole bi-infinite sequence ends in X. This is possible since X is primitive. We show that using this kind of block coding, the probability of the event $\{g(\mathbf{Y})_0 \neq \mathbf{Y}_0\}$ is dominated by $R_0(X, Y)$. The construction procedure of our stationary block function is demonstrated in Figure 6.

We now start with the proof. Let X be defined by the primitive labeled graph $G = (V, E, L)$. We fix an arbitrarily small $\varepsilon \in (0, 1)$. We also fix an arbitrary vertex $v_0 \in V$. Since G is primitive, there exists a number p such that for any two vertices $v_1, v_2 \in V$, there exists a directed path of length p from v_1 to v_2. For any $v_1, v_2 \in V$ we fix such a directed path of length p, which we denote by $\Gamma(v_1, v_2)$. We also define $L(\Gamma) \triangleq (L(e_1), \ldots, L(e_m)) \in \mathcal{B}_m(X)$ to be the word obtained by reading the labels on Γ.

In order to simplify the notation, for any $M' \in \mathbb{N}$ we denote $r_\varepsilon(M') \triangleq R_\varepsilon(\mathcal{B}_{M'}(X), \mathcal{B}_{M'}(Y), \mu_{M'})$. For any $M' \in \mathbb{N}$ we let $\delta_{M'} \geqslant 0$ be such that

$$\frac{r_\varepsilon(M')}{M'} = R_\varepsilon(X, Y, \mu) + \delta_{M'}.$$

By the definition of $R_\varepsilon(X, Y, \mu)$, there exists arbitrarily large M' such that $\delta_{M'}$ is sufficiently small. For a fixed M', we let $N \triangleq \lceil \frac{M'}{p} \rceil$, and we easily observe that

$$r_\varepsilon(Np+1) \leqslant r_\varepsilon(M') + p.$$

Thus,

$$\begin{aligned} \frac{r_\varepsilon(Np+1)}{Np+1} &\leqslant \frac{r_\varepsilon(M')+p}{Np+1} \\ &= \frac{M'}{Np+1} \cdot \frac{r_\varepsilon(M')}{M'} + \frac{p}{Np+1} \\ &\leqslant R_\varepsilon(X, Y, \mu) + \delta_{M'} + \frac{1}{N}. \end{aligned}$$

Since $N \to \infty$ as $M' \to \infty$, by the above inequality, replacing M' by $M \triangleq Np + 1$, we conclude that one may always find $M \in \mathbb{N}$ arbitrarily large such that δ_M is arbitrarily small and $\gcd(M, p) = 1$.

We now fix such $M = Np + 1 \in \mathbb{N}$. From the definition of $r_\varepsilon(n)$, there exist a map $\varphi' : \mathcal{B}_M(Y) \to E^M$ and a set $S \subseteq \mathcal{B}_M(Y)$ of μ_M-measure at least $1 - \varepsilon$, such that for all $\overline{y} \in S$, $\varphi'(\overline{y})$ is a directed path on G and

$$d(\overline{y}, L(\varphi'(\overline{y}))) \leqslant r_\varepsilon(M) \leqslant M \cdot (R_\varepsilon(X, Y, \mu) + \delta_M). \quad (25)$$

We call the words in S *good words*. For every $\overline{y} \in \mathcal{B}_M(Y)$ we now replace the length p prefix and suffix of $\varphi'(\overline{y})$ by paths of length p starting and ending in v_0 respectively. We denote the new path by $\varphi(\overline{y})$. That is, for $\overline{y}$ such that $\varphi'(\overline{y}) = (e_0, \ldots, e_{M-1})$ we define

$$\begin{aligned} \varphi(\overline{y}) \triangleq \big(&\Gamma(v_0, \sigma(e_{p-1})), e_p, \\ &\ldots, e_{M-p-1}, \Gamma(\sigma(e_{M-p}), v_0)\big) \in E^M. \end{aligned}$$

We denote $\tilde{\delta}_M \triangleq \delta_M + 2p/M$. By (25)

$$\begin{aligned} d(\overline{y}, L(\varphi(\overline{y}))) &\leqslant d(\overline{y}, L(\varphi'(\overline{y}))) + 2p \\ &\leqslant M \cdot \left(R_\varepsilon(X, Y, \mu) + \delta_M + \frac{2p}{M} \right) \quad (26) \\ &= M \cdot \left(R_\varepsilon(X, Y, \mu) + \tilde{\delta}_M \right). \end{aligned}$$

We recall that the measure μ is aperiodic and ergodic, and we note that the positive integers $n_1 = p$ and $n_2 = M$ are co-prime. Defining $q_1 = \frac{1}{p \cdot M}$ and $q_2 = \frac{1}{M} \cdot (1 - \frac{1}{M})$, we clearly have $q_1 n_1 + q_2 n_2 = 1$. We consider the partition defined by the M first coordinates of $\mathbf{Y}$. That is

$$\mathcal{P}' \triangleq \{[\overline{y}] \ : \ \overline{y} \in \mathcal{B}_M(Y)\},$$

where we recall the definition of a cylinder from (1). By Lemma 26 there exist measurable sets $Q_1, Q_2 \subseteq Y$ such that the set

$$\mathcal{P} = \left\{ T^{-\ell}(Q_j) \ : \ 0 \leqslant \ell \leqslant n_j - 1, \quad j = 1, 2 \right\}$$

is a partition of Y, as well as for $j = 1, 2$ we have $\mu(Q_j) = q_j$ and Q_j is independent of $\mathcal{P}'$. We shall use the partition $\mathcal{P}$ in order to divide a bi-infinite sequence to blocks.

For a fixed $\mathbf{y} \in Y$, and $m \in \mathbb{Z}$, we say that $\mathbf{y}$ admits a block of length n_j, $j = 1, 2$, in the coordinates $m, m+1, \ldots, m + n_j - 1$ if $T^m(\mathbf{y}) \in Q_j$. We enumerate the blocks composing $\mathbf{y}$ and denote them by $(B_k(\mathbf{y}) = (m_k(\mathbf{y}), l_k(\mathbf{y})))_{k \in \mathbb{Z}}$, where $B_0(\mathbf{y})$ is the block containing the 0 coordinate, m_k is the starting point of the k-th block and $l_k \in \{n_1, n_2\}$ is its length. By the construction of the partition $\mathcal{P}$, each coordinate in $\mathbb{Z}$ belongs to exactly one block. Therefore, in order to define a function $Y \to X$ it is sufficient to define the values that it takes in each and every block.

Given a $\mathbf{y} \in Y$ and the corresponding sequence of blocks $(B_k(\mathbf{y}) = (m_k(\mathbf{y}), l_k(\mathbf{y})))_{k \in \mathbb{Z}}$ we define f_M as follows:

- We define f_M on blocks of length $n_1 = p$ to be the labels on the self loop $\Gamma(v_0, v_0) \triangleq \Gamma_0$.
- We define f_M on blocks of length $n_2 = M$ to be the labels $\varphi(\overline{y})$, where $\overline{y} \in \mathcal{B}_M(Y)$ is the word that appears in the n_2-length block.

That is

$$f_M(\mathbf{y})_{m_k}^{m_k + l_k - 1} = \begin{cases} L(\Gamma_0) & l_k = n_1 = p, \\ L\big(\varphi(\mathbf{y}_{m_k}^{m_k + M - 1})\big) & l_k = n_2 = M. \end{cases}$$

By the construction of φ and f_M, for any given block, the corresponding path begins and ends in v_0, which implies that $f_M(\mathbf{y})$ corresponds to a bi-infinite path in G, and therefore $\mathrm{Im}(f_M) \subseteq X$. The function f_M is also measurable as $\mathcal{P}$ is a measurable partition. We note that for any $\mathbf{y} \in Y$, from the definition of the block partition, the block partition of $T(\mathbf{y})$ is a shifted version of the block partition of $\mathbf{y}$. Thus, f_M also commutes with the shift, which makes it a stationary coding function.

We now turn to prove that $\mathbb{P}_\mu[f_M(\mathbf{Y})_0 \neq \mathbf{Y}_0]$ indeed approximates $R_0(X, Y, \mu)$ for sufficiently large M. Let $\mathbf{X}$ and $\mathbf{Y}$ denote the random bi-infinite sequences in $\Sigma^{\mathbb{Z}}$ generated with respect to μ, and let $I_{\{f_M(\mathbf{Y})_0 \neq \mathbf{Y}_0\}}$ be the indicator function of the event $\{f_M(\mathbf{Y})_0 \neq \mathbf{Y}_0\}$. Since μ is an invariant measure, for all $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}_\mu[f_M(\mathbf{Y})_0 \neq \mathbf{Y}_0] &= \int I_{\{f_M(\mathbf{Y})_0 \neq \mathbf{Y}_0\}} \cdot d\mu \\ &= \int \frac{1}{n} \sum_{k=0}^{n-1} I_{\{f_M(\mathbf{Y}) \neq \mathbf{Y}_0\}} \circ T^k \cdot d\mu \\ &= \mathbb{E}_\mu \left[\frac{1}{n} d\big(f_M(\mathbf{Y})_0^{n-1}, \mathbf{Y}_0^{n-1}\big) \right]. \end{aligned}$$

Let C_n be the random variable that counts the number of good blocks in coordinates $0, 1, \ldots, n-1$. These are blocks of length $n_2 = M$ contained in $\mathbf{Y}_0^{n-1}$ which contain a good word, i.e., a word in the set S. Formally, for a sequence $\mathbf{y} \in Y$, with corresponding blocks $(B_k(\mathbf{y}) = (m_k(\mathbf{y}), l_k(\mathbf{y}))_{k \in \mathbb{Z}}$

$$\begin{aligned} C_n(\mathbf{y}) \triangleq \Big| \Big\{ &k \in \mathbb{Z} \ : \ l_k = n_2, \\ &[m_k, m_k + n_2 - 1] \subseteq [0, n-1] \text{ and } \mathbf{y}_{m_k}^{m_k + n_2 - 1} \in S \Big\} \Big|. \end{aligned}$$

By the construction of f_M, φ, S, and by (26), the number of coordinates inside a single good block in which $f_M(\mathbf{Y})$ and

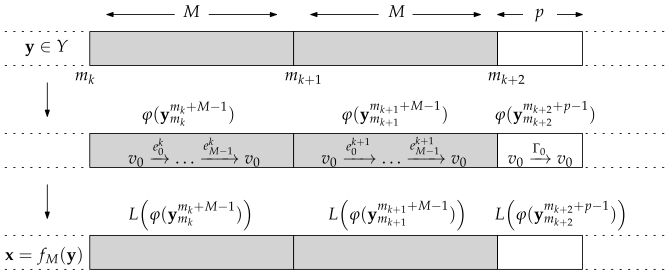

Fig. 6. A demonstration of the coding procedure described in the proof of Proposition 27.

$\mathbf{Y}$ do not agree is upper bounded by $M(R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M)$. Thus, we have

$$\begin{aligned} d\big(f_M(\mathbf{Y})_0^{n-1},\mathbf{Y}_0^{n-1}\big) &\leqslant C_n(\mathbf{Y}_0^{n-1})\cdot(M(R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M)) \\ &\quad+\big(n-C_n(\mathbf{Y}_0^{n-1})M\big) \\ &= C_n(\mathbf{Y}_0^{n-1})\cdot M(R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M-1)+n \end{aligned}$$

and therefore,

$$\begin{aligned} &\mathbb{E}_\mu\big[d\big(f_M(\mathbf{Y})^{n-1},\mathbf{Y}_0^{n-1}\big)\big] \\ &\leqslant \mathbb{E}_\mu[C_n(\mathbf{Y}_0^{n-1})]\cdot M(R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M-1)+n. \quad (27) \end{aligned}$$

Let A be the event that a good block starts at the 0 coordinate. That is,

$$A = Q_2\cap\big\{\mathbf{Y}_0^{M-1}\in S\big\}.$$

We note that we can write $C_n(\mathbf{Y}_0^{n-1})=\sum_{k=0}^{n-M}I_A\circ T^k$, where I_A is the indicator function of the event A. Thus, since μ is shift invariant, and since Q_2 is independent of the first M coordinates, we have

$$\begin{aligned} \mathbb{E}_\mu[C_n(\mathbf{Y}_0^{n-1})] &= (n-M)\mathbb{P}_\mu[A] \\ &= (n-M)\mathbb{P}_\mu\big[Q_2\cap\big\{\mathbf{Y}_0^{M-1}\in S\big\}\big] \\ &= (n-M)\mathbb{P}_\mu[Q_2]\mathbb{P}_\mu\big[\mathbf{Y}_0^{M-1}\in S\big] \\ &\geqslant (n-M)\left(\frac{1}{M}\left(1-\frac{1}{M}\right)\right)(1-\varepsilon). \end{aligned}$$

Assume without loss of generality that δ_M and $2p/M$ are sufficiently small such that $R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M-1<0$. Thus, by combining the lower bound on $\mathbb{E}_\mu[C_n(\mathbf{Y}_0^{n-1})]$ with (27) we obtain

$$\begin{aligned} \mathbb{P}_\mu[f_M(\mathbf{Y})_0\neq\mathbf{Y}_0] &= \mathbb{E}_\mu\left[\frac{1}{n}d\big(f_M(\mathbf{Y})_0^{n-1},\mathbf{Y}_0^{n-1}\big)\right] \\ &\leqslant \frac{n-M}{n}\cdot\frac{1}{M}\left(1-\frac{1}{M}\right)(1-\varepsilon) \\ &\quad\cdot M(R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M-1)+1. \end{aligned}$$

Taking the limit as $n\to\infty$ we get

$$\begin{aligned} &\mathbb{P}_\mu[f_M(\mathbf{Y})_0\neq\mathbf{Y}_0] \\ &\leqslant \left(1-\frac{1}{M}\right)(1-\varepsilon)(R_\varepsilon(X,Y,\mu)+\tilde{\delta}_M-1)+1. \end{aligned}$$

We now conclude as for ε' choosing ε such that $\varepsilon\leqslant\varepsilon'/4$, $R_\varepsilon(X,Y)\leqslant R_0(X,Y)-\varepsilon'/4$ and M sufficiently large such that $\tilde{\delta}_M, 1/M\leqslant\varepsilon'/4$ we obtain

$$\mathbb{P}_\mu[f_M(\mathbf{Y})_0\neq\mathbf{Y}_0]\leqslant R_0(X,Y,\mu)+\varepsilon',$$

which concludes the proof. ■

We now have the first component of the proof of Theorem 25. The second part of the proof requires the well-known fact that stationary coding functions may be approximated by sliding-block-codes.

Lemma 28 ([19, Theorem 3.1]): Let $X,Y\subseteq\Sigma^{\mathbb{Z}}$ be shift spaces, $\mu\in M(Y)$, and $g:Y\to X$ be a stationary coding function. Then for any $\varepsilon>0$, there exists a sliding-block-code function $\hat{f}:Y\to X$ such that

$$\mathbb{P}_\mu\Big[g(\mathbf{Y})\neq\hat{f}(\mathbf{Y})\Big]<\varepsilon.$$

We are now ready to prove Theorem 25.

Proof of Theorem 25: Fix $\varepsilon>0$. By Proposition 27, there exists a stationary coding function $g:Y\to X$ such that

$$\mathbb{P}_\mu[g(\mathbf{Y})_0\neq\mathbf{Y}_0]<R_0(X,Y,\mu)+\frac{1}{2}\varepsilon.$$

By Lemma 28, there exists a sliding-block-code function $\hat{f}:Y\to X$ and a set E with $\mu(E)>1-\frac{1}{2}\varepsilon$ such that g coincides with $\hat{f}$ on E. We now have,

$$\begin{aligned} \mathbb{P}_\mu[\mathbf{Y}_0\neq\hat{f}(\mathbf{Y})_0] &= \mathbb{P}_\mu[\{\mathbf{Y}_0\neq\hat{f}(\mathbf{Y})_0\}\cap E] \\ &\quad+\mathbb{P}_\mu[\{\mathbf{Y}_0\neq\hat{f}(\mathbf{Y})_0\}\cap E^C] \\ &\leqslant \mathbb{P}_\mu[g(\mathbf{Y})_0\neq\mathbf{Y}_0]+\mathbb{P}_\mu[E^C] \\ &< R_0(X,Y,\mu)+\frac{1}{2}\varepsilon+\frac{1}{2}\varepsilon \\ &= R_0(X,Y,\mu)+\varepsilon. \end{aligned}$$

The upper bound $R_0(X,Y,\mu)\leqslant\mathbb{P}_\mu[f(\mathbf{Y})_0\neq\mathbf{Y}_0]$ is immediate by Proposition 16.

VI. Conclusion

In this work, we introduced the Quantized-Constraint Concatenation (QCC) scheme, providing a new general framework for implementing error correction in constrained systems. We have shown that by embedding codewords of an error-correcting code in a constrained system by way of quantization, it is possible to correct $\Theta(n)$ errors (with respect to the code's length n). We discovered that the asymptotic error-correction capabilities of our method for a given constrained system are determined by a new fundamental parameter of the constrained system – its covering radius – which bounds the amount of noise caused by the quantization process. Unlike previous methods, such as concatenation and reverse concatenation, the embedding into the constrained system is not reversible, hence the term quantization.

We presented two different notions for the covering radius of a constrained system, one combinatorial and the other probabilistic. While the combinatorial notion takes into account the worst-case scenario (deep holes), in the probabilistic approach, the essential covering radius ignores rare cases and therefore allows a smaller covering radius. We have studied the properties of the essential and combinatorial covering radii, and provided general lower and upper bounds.

While the covering radius of constrained systems is of independent intellectual merit, let us put our results in the context of the QCC scheme. Consider $X_{0,k}$, the $(0,k)$-RLL system described in Example 1. Using the coding scheme presented in [18, Theorem 1], it is possible to correct up to $O(\sqrt{n})$ errors. However, using QCC with the combinatorial covering radius (which in that case is $\frac{1}{k+1}$), since there exist error-correcting codes with non-vanishing rate capable of correcting up to $(\frac{1}{4}-\delta)n$ errors (for every $\delta>0$), we obtain codes with non-vanishing rate capable of correcting up to $(\frac{1}{4}-\frac{1}{k+1}-\delta)n$ channel errors. On the other hand, we may use the essential covering radius of $X_{0,k}$ to bound the *probable* quantization noise. In that case, since

$$R_{\frac{1}{2}}(X_{0,k},[2]^{\mathbb{Z}},\mu^{\mathsf{u}}) \leqslant R_0(X_{0,k},[2]^{\mathbb{Z}},\mu^{\mathsf{u}}) = \frac{1}{2(2^{k+1}-1)},$$

using the QCC, it is possible to find error-correcting codes such that by removing at most half of the codewords (which asymptotically does not affect the rate), it is possible to improve our error correction capability to $(\frac{1}{4}-\frac{1}{2(2^{k+1}-1)}-\delta)n$ channel errors.

While QCC can operate in the regime where the number of errors is $\Theta(n)$, and other methods cannot, it may be inferior in the regime where the number errors vanishes with respect to the block length. Consider a constrained system X with capacity $\mathcal{C}$ over Σ. In the coding scheme obtained in [18] (called segmented reverse concatenation, or in short SRC), the optimal possible rate for correcting $t=O(\sqrt{n})$ errors is approximately $\mathcal{C}(1-\frac{\log(t)}{n}+o(\frac{1}{n}))$, where n is the block length. Namely, the asymptotic optimal rate using SRC approaches the capacity of the system. On the other hand, as before, in order to correct t errors using QCC, a code correcting $R_0(X)n+t$ errors is required. Thus, the optimal asymptotic rate of QCC is upper bounded by the maximal rate of an error-correcting code capable of correcting $R_0(X)n+t = R_0(X)n+O(\sqrt{n})$ errors. Using the GV bound, the best known asymptotic rate of codes correcting $R_0(X)n$ errors is $1-h_{|\Sigma|}(2R_0(X))$, which for certain parameters might be strictly smaller than $\mathcal{C}$. For example, if $X=X_{0,2}$ is the $(0,2)-$RLL system, the maximal asymptotic rate obtained by QCC is $1-h_2(2/14)\approx 0.408$, while the capacity of the system is $\mathcal{C}\approx 0.8791$.

Previous lower bounds on the possible rates for error-correcting constrained codes have been established in previous work [23], [28], via somewhat non-constructive methods. A certain advantage of our scheme is its simplicity and constructive nature. We also remark that it was suggested in [35] that an error-correcting scheme for constrained systems capable of correcting $\Theta(n)$ errors with a non-vanishing rate may also be obtained by using concatenation with an inner constrained code and outer error-correcting code. Such frameworks however, have yet to be studied in general in the context of constrained systems, and may be of interest for future research.

REFERENCES

[1] R. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes—An application of symbolic dynamics to information theory," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 5–22, Jan. 1983.

[2] S. Alpern, "Generic properties of measure preserving homeomorphisms," in *Ergodic Theory*. Berlin, Germany: Springer, 1979, pp. 16–27.

[3] J. Ashley and B. Marcus, "Canonical encoders for sliding block decoders," *SIAM J. Discrete Math.*, vol. 8, no. 4, pp. 555–605, Nov. 1995.

[4] J. J. Ashley, R. Karabed, and P. H. Siegel, "Complexity and sliding-block decodability," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1925–1947, Nov. 1996.

[5] K. G. Benerjee and A. Banerjee, "On homopolymers and secondary structures avoiding, reversible, reversible-complement and GC-balanced DNA codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 204–209.

[6] G. D. Birkhoff, "Proof of the ergodic theorem," *Proc. Nat. Acad. Sci. USA*, vol. 17, no. 12, pp. 656–660, Dec. 1931.

[7] W. Bliss, "Circuitry for performing error correction calculations on baseband encoded data to eliminate error propagation," *IBM Tech. Discl. Bul.*, vol. 23, pp. 4633–4634, Mar. 1981.

[8] K. Cai, Y. M. Chee, R. Gabrys, H. M. Kiah, and T. T. Nguyen, "Correcting a single indel/edit for DNA-based data storage: Linear-time encoders and order-optimality," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3438–3451, Jun. 2021.

[9] K. Cai, H. M. Kiah, M. Motani, and T. T. Nguyen, "Coding for segmented edits with local weight constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1694–1699.

[10] J. T. Campbell, R. McCutcheon, and A. Windsor, "Independence and alpern multitowers," *Dyn. Syst.*, vol. 34, no. 2, pp. 239–243, Apr. 2019.

[11] Y. M. Chee, H. M. Kiah, and H. Wei, "Efficient and explicit balanced primer codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5344–5357, Sep. 2020.

[12] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Amsterdam, The Netherlands: North Holland, 1997.

[13] N. G. de Bruijn and P. Erdös, "Some linear and some quadratic recursion formulas. I," *Indagationes Mathematicae*, vol. 54, no. 5, pp. 374–382, 1951.

[14] M. Einsiedler and T. Ward, *Ergodic Theory: With a View Towards Number Theory*. London, U.K.: Springer Verlag, 2011.

[15] J. L. Fan and A. R. Calderbank, "A modified concatenated coding scheme, with applications to magnetic data storage," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1565–1574, Jul. 1998.

[16] M. Fekete, "On the distribution of the roots of certain algebraic equations with integer coefficients," *Math. J.*, vol. 17, no. 1, pp. 228–249, 1923.

[17] H. C. Ferreira, K. A. S. Abdel-Ghaffar, L. Cheng, T. G. Swart, and K. Ouahada, "Moment balancing templates: Constructions to add insertion/deletion correction capability to error correcting or constrained codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3494–3500, Aug. 2009.

[18] R. Gabrys, P. H. Siegel, and E. Yaakobi, "Segmented reverse concatenation: A new approach to constrained ECC," in *Proc. Int. Symp. Inf. Theory Its Appl. (ISITA)*, Oct. 2020, pp. 254–258.

[19] R. Gray, "Sliding-block source coding," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 4, pp. 357–368, Jul. 1975.

[20] U. Gupta, H. M. Kiah, A. Vardy, and H. Yao, "Polar codes with balanced codewords," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 700–705.

[21] V. Guruswami, A. Rudra, and M. Sudan, *Essential Coding Theory*, 2022. [Online]. Available: https://cse.buffalo.edu/faculty/atri/courses/codingtheory/book/web-coding-book.pdf

[22] K. A. S. Immink, "A practical method for approaching the channel capacity of constrained channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1389–1399, Mar. 1997.

[23] V. D. Kolesnik and V. Y. Krachkovsky, "Generating functions and lower bounds on rates for limited error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 778–788, May 1991.

[24] X. Li, M. Chen, and H. Wu, "Multiple errors correction for position-limited DNA sequences with GC balance and no homopolymer for DNA-based data storage," *Briefings Bioinf.*, vol. 24, no. 1, Jan. 2023, Art. no. bbac484.

[25] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding Coding*. Cambridge, U.K.: Cambridge Univ. Press, 2021.

[26] X. Lu and S. Kim, "Design of nonbinary error correction codes with a maximum run-length constraint to correct a single insertion or deletion error for DNA storage," *IEEE Access*, vol. 9, pp. 135354–135363, 2021.
[27] M. Mansuripur, "Enumerative modulation coding with arbitrary constraints and postmodulation error correction coding for data storage systems," *Proc. SPIE*, vol. 1499, pp. 72–86, Jul. 1991.
[28] B. H. Marcus and R. M. Roth, "Improved Gilbert–Varshamov bound for constrained systems," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1213–1221, Jul. 1992.
[29] B. H. Marcus, R. M. Roth, and P. H. Siegel, *An Introduction to Coding for Constrained Systems* (Lecture Notes), Oct. 2001. [Online]. Available: https://personal.math.ubc.ca/Handbook/index.html
[30] A. Mazumdar, R. M. Roth, and P. O. Vontobel, "On linear balancing sets," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 2699–2703.
[31] J. V. Neumann, "Proof of the quasi-ergodic hypothesis," *Proc. Nat. Acad. Sci. USA*, vol. 18, no. 1, pp. 70–82, Jan. 1932.
[32] T. T. Nguyen, K. Cai, K. A. S. Immink, and H. M. Kiah, "Constrained coding with error control for DNA-based data storage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 694–699.
[33] T. T. Nguyen, K. Cai, K. A. S. Immink, and H. M. Kiah, "Capacity-approaching constrained codes with error correction for DNA-based data storage," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5602–5613, Aug. 2021.
[34] W. H. Press, J. A. Hawkins, S. K. Jones, J. M. Schaub, and I. J. Finkelstein, "HEDGES error-correcting code for DNA storage corrects indels and allows sequence constraints," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 31, pp. 18489–18496, Aug. 2020.
[35] R. M. Roth, private communication, 2023.
[36] D. J. Rudolph, *Fundamentals of Measurable Dynamics: Ergodic Theory on Lebesgue Spaces*. New York, NY, U.K.: Oxford Univ. Press, 1990.
[37] P. Walters, *An Introduction to Ergodic Theory*, vol. 79. New York, NY, USA: Springer, 2000.
[38] J. H. Weber, J. A. M. de Groot, and C. J. van Leeuwen, "On single-error-detecting codes for DNA-based data storage," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 41–44, Jan. 2021.

Dor Elimelech (Student Member, IEEE) received the first B.Sc. degree in mathematics and the second B.Sc. and M.Sc. (summa cum laude) degrees in electrical engineering from the Ben-Gurion University of the Negev, Israel, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include coding theory, probability, and dynamical systems.

Tom Meyerovitch received the B.Sc. degree in mathematics and computer science and the M.Sc. and Ph.D. degrees in mathematics from Tel Aviv University, Israel, in 1999, 2004, and 2010, respectively. From 2010 to 2012, he held a post-doctoral position with The University of British Columbia, and the Pacific Institute for Mathematical Studies, Vancouver, Canada. Since 2012, he has been on the Faculty of the Department of Mathematics, Ben-Gurion University of the Negev, Be'er Sheva, Israel. He studies various mathematical aspects of dynamical systems, in particular symbolic dynamics.

Moshe Schwartz (Fellow, IEEE) received the B.A. (summa cum laude), M.Sc., and Ph.D. degrees from the Computer Science Department, Technion–Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004, respectively.

He is a Professor with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel (on a leave of absence), and the Department of Electrical and Computer Engineering, McMaster University, Canada. He was a Fulbright Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, University of California San Diego; and a Post-Doctoral Researcher with the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012–2014, he was a Visiting Scientist with the Massachusetts Institute of Technology (MIT). His research interests include algebraic coding, combinatorial structures, and digital sequences.

Prof. Schwartz received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage and the 2020 NVMW Persistent Impact Prize. He served as an Associate Editor for Coding Techniques and Coding Theory for IEEE TRANSACTIONS ON INFORMATION THEORY, from 2014 to 2021. Since 2021, he has been serving as an Area Editor for Coding and Decoding for IEEE TRANSACTIONS ON INFORMATION THEORY. He has been an Editorial Board Member of the *Journal of Combinatorial Theory Series A*, since 2021.