Semiconstrained Systems

Ohad Elishco, *Student Member, IEEE*, Tom Meyerovich, and Moshe Schwartz, *Senior Member, IEEE*

***Abstract*— When transmitting information over a noisy channel, two approaches, dating back to Shannon's work, are common: assuming the channel errors are independent of the transmitted content and devising an error-correcting code or assuming the errors are data dependent and devising a constrained-coding scheme that eliminates all offending data patterns. In this paper, we analyze a middle road, which we call a semiconstrained system. In such a system, which is an extension of the channel with the cost constraints model, we do not eliminate the error-causing sequences entirely, but rather restrict the frequency in which they appear. We address several key issues in this paper. The first is proving closed-form bounds on the capacity, which allow us to bound the asymptotics of the capacity. In particular, we bound the rate at which the capacity of the semiconstrained $(0, k)$-RLL tends to 1 as k grows. The second key issue is devising efficient encoding and decoding procedures that asymptotically achieve capacity with vanishing error. Finally, we consider delicate issues involving the continuity of the capacity and a relaxation of the definition of semiconstrained systems.**

***Index Terms*— Constrained coding, capacity, large deviations, encoder construction.**

I. Introduction

ONE OF the most fundamental problems in coding and information theory is that of transmitting a message over a noisy channel and attempting to recover it at the receiving end. This is either when the transmission is over a distance (a communications system), or over time (a storage system). Two common approaches to deal with this problem were already described in Shannon's work [31]. The first approach uses an error-correcting code to combat the errors introduced by the channel. The theory of error-correcting codes has been studied extensively, and a myriad of code constructions are known for a wide variety of channels (for example, see [19], [21], [25], [26], and the many references therein). The second approach asserts that the channel introduces errors in the data stream only in response to certain patterns, such as offending substrings. It follows that removing the offending substrings from the stream entirely will render the channel noiseless. Schemes of this sort have been called constrained systems or constrained codes, and they have also been extensively studied and used (for example, see [8], [20], and the references therein).

Both approaches are not free of cost. Error-correcting codes incur a rate penalty, depending on the specific code used, and bounded by the channel error model that is assumed. Constrained codes also impose a rate penalty that is bounded by the capacity of the constrained system.

The two approaches, one based on error-correcting codes and one based on constrained codes, may be viewed as two extremes: while the first assumes the errors are data independent, the second assumes the errors are entirely data dependent. Since in the real world the situation may not be either of the extremes, existing solutions may over-pay in rate.

The goal of this paper is to define and study semiconstrained systems and their properties, as well as suggest encoding and decoding procedures.

Arguably, the most famous constrained system is the (d, k)-RLL system, which contains only binary strings with at least d 0's between adjacent 1's, and no $k+1$ consecutive 0's (see [8] for uses of this system). In particular, $(0, k)$-RLL is defined by the removal of a single offending substring, namely, it contains only binary strings with no occurrence of $k+1$ consecutive zeros, denoted 0^{k+1}. Informally, a semiconstrained $(0, k)$-RLL system has an additional parameter, $p \in [0, 1]$, a real number. A binary string is in the system if the frequency that the offending pattern 0^{k+1} occurs does not exceed p. When $p = 1$ this degenerates into a totally unconstrained system that contains all binary strings, whereas when $p = 0$ this is nothing but the usual constrained system, which we call a *fully-constrained system* for emphasis.

The capacity of the semiconstrained $(0, k)$-RLL system is known using the methods of [23]. The expression involves an optimization problem that does not lend itself to finding other properties of the system, such as the rate at which the capacity converges to 1 as k grows. This rate of convergence is known when the system is fully constrained [28]. Additionally, the capacity is known only in the one-dimensional case, whereas the general bounds may be extended to the multi-dimensional case as well.

The first main contribution of this paper is establishing analytic lower and upper bounds on the capacity of semiconstrained $(0, k)$-RLL. These bounds are then used to derive the rate at which the capacity of these systems converges to 1 as k grows, up to a small constant multiplicative factor. The bounds extend previous techniques from [28] as well

Manuscript received July 7, 2015; accepted December 24, 2015. Date of publication February 18, 2016; date of current version March 16, 2016. This work was supported in part by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) within REA through the Israel Science Foundation under Grant 333598, and in part by the Israel Science Foundation under Grant 626/14. This paper was presented at the 2015 Information Theory and Applications Workshop and the 2015 IEEE International Symposium on Information Theory.

O. Elishco and M. Schwartz are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel (e-mail: ohadeli@post.bgu.ac.il; schwartz@ee.bgu.ac.il).

T. Meyerovich is with the Department of Mathematics, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel (e-mail: mtom@math.bgu.ac.il).

Communicated by A. G. Dimakis, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2016.2531722

as employ large-deviations theory. These bounds are also extended to the multi-dimensional case.

This paper is not motivated or limited solely by the case of a single offending substring. We can define multiple offending substrings, each equipped with its own limited empirical frequency. Indeed, with a proper set of semiconstraints, variants such as DC-free RLL are possible (see [16] and references therein). Another motivating example is the system of strings over $\mathbb{Z}_q$ where the offending substring is $q-1, 0, q-1$. In the case of multi-level flash memory cells, inter-cell interference is at its maximum when three adjacent cells are at the highest, lowest, and then highest charge levels possible [3]. By adjusting the amount of such substrings we can mitigate the noise caused by inter-cell interference. Further restrictions, such as the requirement for constant-weight strings (see the recent [13], [24]) correspond to a semiconstrained system. A similar case appears in [29], where the frequency of 101 in a binary code is suppressed, to mitigate the appearance of a "ghost pulse" in optical communications. A more general analysis in the same setting appears in [30] where the probabilities of bit triplets are skewed and the redundancy of coding for such systems is studied. Another motivation comes from reconstructing DNA sequences using partial information about their subsequences [15]. A part of the solution suggested in [15] uses the distribution of different substrings of length ℓ and their growth rate as ℓ grows. This can be formulated as a problem of calculating the capacity of a semiconstrained system.

Coding schemes for some of these systems exist. They are mostly ad-hoc and tailored for each specific case, as in [13], [16], and [29]. A more general constant-to-constant coding scheme exists [14] for a channel with cost constraints model. However, it is not optimal in terms of rate, and it addresses only scalar cost functions, which is a slightly different model than the semiconstrained systems we study. In the scalar cost function, each symbol receive a weight and the goal is to limit the expected weight of the codewords. In the semiconstrained systems the goal is to control the appearance of each problematic pattern independently. Both models can be viewed as a private case of a general model in which we would like to control the problematic patterns with respect to some functions on those patterns.

The second main contribution of this paper is a general, explicit, constant bit-rate to constant bit-rate encoding and decoding scheme. This coding scheme is based on the theory of large deviations, and it asymptotically achieves capacity, with a vanishing failure probability as the block length grows. To that end, we also define and study a relaxation of semiconstrained systems, allowing us to address the issue of the existence of the limit in the definition of the capacity, as well as the continuity of the capacity.

We would like to highlight some of the main differences between this paper and previous works. In [11] and [14] the capacity of channels with cost constraints is investigated. Such channels define a scalar cost function that is applied to each sliding-window k-tuple in the transmission. The admissible sequences are those whose average cost per symbol is less than some given scalar constraint. In our paper, however, we investigate sequences with a cost function which can control *separately* the appearance of any unwanted word (not necessarily of the same length).

The more general framework we study is similar to that of [1], [22], [23]. In [22] some embedding theorems and results concerning the entropy of a weight-per-symbol shift of finite type (SFT) are presented. A weight-per-symbol SFT is a graph representation of an SFT with a weight function which assigns to each edge a weight in $\mathbb{R}^d$. In [1], some large-deviation theorems are proved for empirical types of Markov chains that are constrained to thin sets. A thin set is a set whose convex hull has a strictly lower dimension (which means it has an empty interior topologically). We also mention [23], in which an improved Gilbert-Varshamov bound for fully-constrained systems is found. Thus, [23] studies certain semiconstrained systems as means to an altogether different end. Using these works, the exact capacity of semiconstrained systems, as defined in this paper, may be calculated. However, key issues we address are not covered by these papers, including the rate of convergence of the capacity, the existence of the limit in the capacity definition, and continuity of the capacity.

Lastly, in [14], coding for channels with cost constraints is investigated. The main focus is given to functions with a scalar cost on symbols, whereas the model we study in this paper is different. The proposed constant-to-constant coding scheme of [14] is based on the state-splitting algorithm and does not achieve the capacity in general.

As a final introductory note, semiconstrained systems are not to be confused with weakly-constrained systems [7]. Unlike the model we study, the weakly-constrained scheme comprises of an encoder function from strings of length n to strings of length $n + r$, and a prescribed error probability, p. The encoder must, on a $1 - p$ fraction of the possible unconstrained input strings, produce fully-constrained output strings. For the remaining p fraction of the input strings, the encoder may produce totally unconstrained output strings.

The paper is organized as follows. In Section II we give the basic definitions and the notation used throughout the paper. We also cite some previous work and derive some elementary consequences. In Section IV we introduce a relaxation called weak semiconstrained systems, and study issues involving existence of the limit in the capacity definition as well as continuity of the capacity. In Section III we present an upper and a lower bound on the capacity of the $(0, k, p)$-RLL semiconstrained system, as well as bound the capacity's rate of convergence as k grows. Section V is devoted to devising an encoding and decoding scheme for weak semiconstrained systems. We conclude in Section VI with a summary of the results.

II. Preliminaries

Let Σ be a finite alphabet and let Σ^* denote the set of all the finite sequences over Σ. The elements of Σ^* are called *words* (or *strings*). The *length* of a word $\boldsymbol{\omega} \in \Sigma^*$ is denoted by $|\boldsymbol{\omega}|$. Assuming $\boldsymbol{\omega} = \omega_0\omega_1 \dots \omega_{\ell-1}$, with $\omega_i \in \Sigma$, a *subword* (or *substrings*) is a string of the form $\omega_i\omega_{i+1} \dots \omega_{i+m-1}$, where $0 \leqslant i \leqslant i+m \leqslant \ell$. For convenience, we define

$$\boldsymbol{\omega}_{i,m} = \omega_i \omega_{i+1} \dots \omega_{i+m-1},$$

i.e., $\boldsymbol{\omega}_{i,m}$ denotes the substring of $\boldsymbol{\omega}$ which is of length m and is starting at the ith position.

Given two words, $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Sigma^*$, their concatenation is denoted by $\boldsymbol{\omega}\boldsymbol{\omega}'$. Repeated concatenation is denoted using a superscript, i.e., for any natural $m \in \mathbb{N}$, $\boldsymbol{\omega}^m$ denotes

$$\boldsymbol{\omega}^m = \boldsymbol{\omega}\boldsymbol{\omega} \dots \boldsymbol{\omega},$$

where m copies of $\boldsymbol{\omega}$ are concatenated. As an example,

$$1(10)^3 01^2 0^3 = 1101010011000.$$

The following definition will be used later when defining semiconstrained systems.

For any two words $\boldsymbol{\tau}, \boldsymbol{\omega} \in \Sigma^*$, let $\mathrm{fr}(\boldsymbol{\tau}, \boldsymbol{\omega})$ denote the frequency of $\boldsymbol{\tau}$ as a subword of $\boldsymbol{\omega}$, i.e.,

$$\mathrm{fr}(\boldsymbol{\tau}, \boldsymbol{\omega}) = \frac{1}{|\boldsymbol{\omega}| - |\boldsymbol{\tau}| + 1} \sum_{i=0}^{|\boldsymbol{\omega}|-|\boldsymbol{\tau}|} [\boldsymbol{\omega}_{i,|\boldsymbol{\tau}|} = \boldsymbol{\tau}]. \quad (1)$$

Here, $[A]$ denotes the Iverson bracket, having a value of 1 if A is true, and 0 otherwise. If $|\boldsymbol{\tau}| > |\boldsymbol{\omega}|$ then we define $\mathrm{fr}(\boldsymbol{\tau}, \boldsymbol{\omega}) = 0$.

We are now ready to define semiconstrained systems.

Definition 1: Let $\mathcal{F} \subseteq \Sigma^*$ be a finite set of words, and let $P \in [0,1]^{\mathcal{F}}$ be a function from $\mathcal{F}$ to the real interval $[0, 1]$. *A* semiconstrained system (SCS)*, $X(\mathcal{F}, P)$, is the following set of words,*

$$X(\mathcal{F}, P) = \left\{\boldsymbol{\omega} \in \Sigma^* : \forall \boldsymbol{\phi} \in \mathcal{F}, \mathrm{fr}(\boldsymbol{\phi}, \boldsymbol{\omega}) \leqslant P(\boldsymbol{\phi})\right\}.$$

When $\mathcal{F}$ and P are understood from the context, we may omit them and just write X. We also define the set of words of length exactly n in $X(\mathcal{F}, P)$ as

$$\mathcal{B}_n(\mathcal{F}, P) = X(\mathcal{F}, P) \cap \Sigma^n.$$

Example 2: We recall that the (d,k)-RLL constrained system contains exactly the binary words with at least d 0's between adjacent 1's, and no $k+1$ 0's in a row. Using our notation, after setting $\Sigma = \{0, 1\}$, the (d,k)-RLL constrained system is a semiconstrained system $X(\mathcal{F}, P)$, where

$$\mathcal{F} = \left\{11, 101, 10^2 1, \dots, 10^{d-1}1, 0^{k+1}\right\},$$

and $P(\boldsymbol{\phi}) = 0$ for all $\boldsymbol{\phi} \in \mathcal{F}$. □

Another example is the constant-weight ICI-free codes which are described in [13].

Example 3: The constant-weight ICI-free codes comprised of a codebook that satisfies the following two constraints:

1) *For code length n and some fixed r, $r \in [0, 1]$, every codeword has constant Hamming weight of rn.*
2) *The subsequence* 101 *does not appear in any codeword.*

We can rewrite this two constraints as a SCS system as follows. Let $\mathcal{F} = \{0, 1, 101\}$ and $P = (r, 1-r, 0)$ such that $P(0) = r$, $P(1) = 1-r$, $P(101) = 0$. □

An important object of interest is the capacity of an SCS. We define it as follows.

Definition 4: Let $X(\mathcal{F}, P)$ be an SCS. The capacity of $X(\mathcal{F}, P)$, which is denoted by $\mathsf{cap}(\mathcal{F}, P)$, is defined as

$$\mathsf{cap}(\mathcal{F}, P) = \limsup_{n\to\infty} \frac{1}{n} \log_2 |\mathcal{B}_n(\mathcal{F}, P)|.$$

If we had a closed-form expression for $|\mathcal{B}_n(\mathcal{F}, P)|$, we could calculate the capacity of $(\mathcal{F}, P)$. As in [28], we translate the combinatorial counting problem with a probability-bounding problem. Assume p_n denotes the probability that a random string from Σ^n, which is chosen with uniform distribution, is in $\mathcal{B}_n(\mathcal{F}, P)$. Then,

$$|\mathcal{B}_n(\mathcal{F}, P)| = p_n \cdot |\Sigma|^n,$$

and then

$$\mathsf{cap}(\mathcal{F}, P) = \log_2 |\Sigma| + \limsup_{n\to\infty} \frac{1}{n} \log_2 p_n. \quad (2)$$

In certain cases, a different definition of semiconstrained systems is helpful. The new definition has a cyclic nature. In a similar manner to (1), for any two words $\boldsymbol{\tau}, \boldsymbol{\omega} \in \Sigma^*$, let $\mathrm{fr}^{\mathrm{cyc}}(\boldsymbol{\tau}, \boldsymbol{\omega})$ denote the cyclic frequency of $\boldsymbol{\tau}$ as a subword of $\boldsymbol{\omega}$, i.e.,

$$\mathrm{fr}^{\mathrm{cyc}}(\boldsymbol{\tau}, \boldsymbol{\omega}) = \frac{1}{|\boldsymbol{\omega}|} \sum_{i=0}^{|\boldsymbol{\omega}|-1} [\boldsymbol{\omega}_{i,|\boldsymbol{\tau}|} = \boldsymbol{\tau}],$$

where this time, $\boldsymbol{\omega}_{i,m} = \omega_i \omega_{i+1} \dots \omega_{i+m-1}$, and the indices are taken modulo $|\boldsymbol{\omega}|$. We extend the definitions of $X(\mathcal{F}, P)$ and $\mathcal{B}_n(\mathcal{F}, P)$ to $X^{\mathrm{cyc}}(\mathcal{F}, P)$ and $\mathcal{B}_n^{\mathrm{cyc}}(\mathcal{F}, P)$ in the natural way, by replacing fr with $\mathrm{fr}^{\mathrm{cyc}}$.

We now give a brief overview of some basic definitions and known results in large deviations theory that we use in this paper. Let $\overline{\Gamma}$ denote the closure of a set Γ, and let Γ° denote its interior. Let $\mathcal{X}$ be some Polish space equipped with the Borel sigma algebra.

Definition 5: A rate function *I is a mapping $I : \mathcal{X} \to [0, \infty]$ such that for all $\alpha \in [0, \infty)$, the level set $\phi_I(\alpha) = \{x \in \mathcal{X} : I(x) \leqslant \alpha\}$ is a closed subset of $\mathcal{X}$.*

Definition 6: Let $\{\mu_n\}$ be a sequence of probability measures. We say that $\{\mu_n\}$ satisfies the large-deviation principle (LDP) with a rate function I, if for every Borel set $\Gamma \subseteq \mathcal{X}$,

$$\begin{aligned} -\inf_{x\in\Gamma^\circ} I(x) &\leqslant \liminf_{n\to\infty} \frac{1}{n} \log_2 \mu_n(\Gamma) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{n} \log_2 \mu_n(\Gamma) \leqslant -\inf_{x\in\overline{\Gamma}} I(x). \end{aligned}$$

Definition 7: A good rate function $I : \mathcal{X} \to [0, \infty]$ is a rate function for which all the level sets are compact subsets of $\mathcal{X}$.

Lemma 8 ([5, Lemma 4.1.5]): Let $\mathcal{X}$ be a polish space and let $\mathcal{Y} \subseteq \mathcal{X}$ be a G_δ-subset of $\mathcal{X}$ (countable intersection of open sets in $\mathcal{X}$), equipped with the induced topology. Let $\{\mu_n\}$ be finite measures on $\mathcal{X}$ such that $\mu_n(\mathcal{X} \setminus \mathcal{Y}) = 0$ for all $n \geqslant 1$ and let I be a good rate function on $\mathcal{X}$ such that $I(x) = \infty$ for all $x \in \mathcal{X} \setminus \mathcal{Y}$. Let $\mu_n|_{\mathcal{Y}}$ and $I|_{\mathcal{Y}}$ denote the restriction of μ_n and I to $\mathcal{Y}$, respectively. Then, $I|_{\mathcal{Y}}$ is a good rate function on $\mathcal{Y}$ and the following statements are equivalent.

1) *The sequence $\{\mu_n\}$ satisfies the LDP with rate function I.*

2) *The sequence* $\{\mu_n|_{\mathcal{Y}}\}$ *satisfies the LDP with rate function* $I|_{\mathcal{Y}}$.

Let $M_1(\Sigma)$ denote the space of all probability measures on some finite alphabet Σ.

Definition 9: Let $\overline{\boldsymbol{Y}} = Y_0, Y_1, \dots$ *be a sequence over some alphabet* Σ*, and let* $\boldsymbol{y} \in \Sigma^*$*. We denote by* $\mathrm{fr}_n^{\overline{\boldsymbol{Y}}}(\boldsymbol{y})$ *the empirical occurrence frequency of the word* $\boldsymbol{y}$ *in the first n places of* $\overline{\boldsymbol{Y}}$*, i.e.*

$$\mathrm{fr}_n^{\overline{\boldsymbol{Y}}}(\boldsymbol{y}) = \mathrm{fr}\left(\boldsymbol{y}, \overline{\boldsymbol{Y}}_{0,n+|\boldsymbol{y}|-1}\right) = \frac{1}{n}\sum_{i=0}^{n-1}\left[\overline{\boldsymbol{Y}}_{i,|\boldsymbol{y}|} = \boldsymbol{y}\right].$$

We denote by $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}} \in M_1(\Sigma^k)$ *the vector of empirical distribution of* Σ^k *in* $\overline{\boldsymbol{Y}}$*, i.e., for a k-tuple* $\boldsymbol{y} \in \Sigma^k$*, the coordinate that corresponds to* $\boldsymbol{y}$ *in* $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$ *is* $\mathrm{fr}_n^{\overline{\boldsymbol{Y}}}(\boldsymbol{y})$.

Suppose $\overline{\boldsymbol{Y}} = Y_0, Y_1, \dots$ are Σ-valued i.i.d. random variables, with $q(\sigma)$ denoting the probability that $Y_i = \sigma$, for all i. We assume that $q(\sigma) > 0$ for all $\sigma \in \Sigma$. We denote by $q(\sigma_0, \sigma_1, \dots, \sigma_{k-1})$ the probability of the sequence $\sigma_0, \sigma_1, \dots, \sigma_{k-1}$. The following theorem connects the empirical distribution with the large-deviation principle.

Theorem 10 ([5, Sec. 3.1]): Let $\nu \in \mathcal{X} = M_1(\Sigma^k)$*, and let* $\overline{\boldsymbol{Y}} = Y_0, Y_1, \dots$ *be* Σ*-valued i.i.d. random variables, with* $q(\sigma) > 0$ *denoting the probability that* $Y_i = \sigma$ *for* $\sigma \in \Sigma$*. For every Borel set,* $\Gamma \subseteq \mathcal{X}$*, define*

$$\mu_n(\Gamma) = \Pr\left[\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}} \in \Gamma\right].$$

Let us denote by $\nu_1 \in M_1(\Sigma^{k-1})$ *the marginal of* ν *obtained by projecting onto the first* $k-1$ *coordinates,*

$$\nu_1(\sigma_0, \dots, \sigma_{k-2}) = \sum_{\sigma\in\Sigma} \nu(\sigma_0, \dots, \sigma_{k-2}, \sigma).$$

Then the rate function, $I : \mathcal{X} \to [0, \infty]$*, governing the LDP of the empirical distribution* $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$ *with respect to* Γ *is,*

$$I(\nu) = \begin{cases} \sum\limits_{\boldsymbol{\sigma}\in\Sigma^k} \nu(\boldsymbol{\sigma}) \log_2 \frac{\nu(\boldsymbol{\sigma})}{\nu_1(\boldsymbol{\sigma}_{0,k-1})q(\sigma_{k-1})} & \nu \text{ is shift invariant,} \\ \infty & \text{otherwise,} \end{cases}$$

where $\nu \in \mathcal{X} = M_1(\Sigma^k)$ *is shift invariant if*

$$\sum_{\sigma\in\Sigma} \nu(\sigma, \sigma_1, \sigma_2, \dots, \sigma_{k-1}) = \sum_{\sigma\in\Sigma} \nu(\sigma_1, \sigma_2, \dots, \sigma_{k-1}, \sigma).$$

In the context of an $X(\mathcal{F}, P)$ SCS with $\mathcal{F} \subseteq \Sigma^k$, i.e., all the offending words are of equal length, the set $\Gamma \subseteq M_1(\Sigma^k)$ takes on the following intuitive form,

$$\Gamma = \left\{(p_{\boldsymbol{\phi}})_{\boldsymbol{\phi}\in\Sigma^k} \in M_1(\Sigma^k) : \forall \boldsymbol{\phi} \in \mathcal{F}, p_{\boldsymbol{\phi}} \leqslant P(\boldsymbol{\phi})\right\}.$$

In other words, Γ contains all the vectors indexed by the elements of Σ^k, such that each entry is a real number from $[0, 1]$, the entries sum to 1, and each entry corresponding to an offending word $\boldsymbol{\phi} \in \mathcal{F}$ does not exceed $P(\boldsymbol{\phi})$.

Note that if I is continuous and $\Gamma \subseteq \mathcal{X}$ is such that $\overline{\Gamma} = \overline{\Gamma^\circ}$, then $\inf_{x\in\overline{\Gamma}} I(x) = \inf_{x\in\Gamma^\circ} I(x)$. In that case the limit of Definition 6 exists, giving $\lim_{n\to\infty} \frac{1}{n}\log_2 \mu_n(\Gamma) = -\inf_{x\in\Gamma} I(x)$.

An important observation is the following. Assume $\Sigma = \{1, 2, \dots, |\Sigma|\}$. Let us denote the vector of probabilities for the symbols from Σ by $\boldsymbol{q} = (q(1), q(2), \dots, q(|\Sigma|))$. For a probability measure $\mu \in M_1(\Sigma^{k-1})$ we define the probability measure $\mu \otimes \boldsymbol{q} \in M_1(\Sigma^k)$ as follows. The measure $\mu \otimes \boldsymbol{q}$ may be viewed as a vector indexed by Σ^k, whose value at the coordinate $i_1, i_2, \dots, i_k$ is

$$(\mu \otimes \boldsymbol{q})(i_1, \dots i_k) = \mu(i_1, \dots, i_{k-1}) \cdot q(i_k).$$

We now note that the rate function on the set of shift-invariant measures, that governs the LDP of $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$ as defined in Theorem 10, can be written as

$$I(\nu) = H(\nu|\nu_1 \otimes \boldsymbol{q}),$$

where $H(\cdot|\cdot)$ is the relative-entropy function. Since the relative entropy is nonnegative and convex, and the set of shift-invariant measures is closed and convex, we reach the following corollary.

Corollary 11: The rate function governing the LDP of $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$ *defined in Theorem 10 is nonnegative and convex on the set of shift-invariant measures* $\nu \in M_1(\Sigma^k)$.

Finally, the following corollary shows cases in which the constraints in P are redundant.

Corollary 12: Let $\mathcal{F} \subseteq \Sigma^k$*. If* $P(\boldsymbol{\phi}) \geqslant |\Sigma|^{-k}$ *for all* $\boldsymbol{\phi} \in \mathcal{F}$*, then* $\mathsf{cap}(\mathcal{F}, P) = \log_2|\Sigma|$.

Proof: Assume $\overline{\boldsymbol{Y}} = Y_0, Y_1, \dots$ is a sequence of i.i.d. random variables distributed uniformly (each symbol with probability $|\Sigma|^{-1}$). Let I be the rate function governing the LDP of $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$ as defined in Theorem 10. Consider the shift-invariant measure $\nu \in M_1(\Sigma^k)$, $\nu(\boldsymbol{\sigma}) = |\Sigma|^{-k}$, for all $\boldsymbol{\sigma} \in \Sigma^k$.

Note that we obtain that $I(\nu) = 0$ and by Corollary 11, ν minimizes the rate function. Let $\Gamma \subseteq M_1(\Sigma^k)$ be the set associated with the constraint, then

$$\begin{aligned} \mathsf{cap}(\mathcal{F}, P) &= \lim_{n\to\infty} \frac{1}{n}\log_2\left(|\Sigma|^n \Pr[\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}} \in \Gamma]\right) \\ &= \log_2|\Sigma| + \lim_{n\to\infty} \frac{1}{n}\log_2\left(\Pr[\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}} \in \Gamma]\right) \\ &= \log_2|\Sigma| - I(\nu) = \log_2|\Sigma|. \end{aligned}$$

■

III. Bounds on the Capacity of $(0, k, p)$-RLL SCS and Rate of Convergence

In this section we obtain closed-form bounds on the capacity that allow us to analyze the asymptotics of the capacity of semiconstrained systems, and prove bounds on the capacity in the multi-dimensional case. We focus on the family of semiconstrained $(0, k)$-RLL, since it is defined by a single offending string of 0^{k+1}. While the results are specific to this family, we note that the method may be extended to some more general semiconstrained systems.

For reasons that will become apparent later, we conveniently invert the bits of the system, and define the $(0, k)$-RLL constrained system as the set of all finite binary strings not containing the 1^{k+1} substring. We therefore consider the semiconstrained system $X(\mathcal{F}, P)$ defined by

$$\mathcal{F} = \left\{1^{k+1}\right\}, \quad P(1^{k+1}) = p,$$

for some real constant $p \in [0, 1]$. We call this semiconstrained system the *$(0, k, p)$-RLL SCS*, and throughout this section we denote its capacity by $C_{k,p}$. Thus, $C_{k,0}$ denotes the capacity of the fully-constrained $(0, k)$-RLL system.

In the case of fully-constrained $(0, k)$-RLL, the asymptotics of the capacity, as k tends to infinity, are well known. It was mentioned in [12], that

$$1 - C_{k,0} = \frac{\log_2 e}{4 \cdot 2^k}(1 + o(1))$$

together with bounds on the capacity of two dimensional $(0, k)$-RLL. The rate was later extended in [28] to the multi-dimensional case, where it was shown that

$$1 - C_{k,0}^{(D)} = \frac{D \log_2 e}{4 \cdot 2^k}(1 + o(1)),$$

where $C_{k,0}^{(D)}$ denotes the capacity of the fully-constrained D-dimensional $(0, k)$-RLL.

Our analysis will proceed by proving a lower bound and an upper bound on the capacity of $(0, k, p)$-RLL SCS, and then analyzing it when $k \to \infty$. We note that due to Corollary 12 the interesting region is $p \leqslant \frac{1}{2^{k+1}}$ or else the capacity is exactly 1.

A. An Upper Bound on the Capacity of (0, k, p)-RLL SCS

In order to obtain an upper bound on the capacity of $(0, k, p)$-RLL SCS we employ a bound by Janson [9]. Consider an index set, Q, and a set $\{J_i\}_{i \in Q}$ of independent random indicator variables. Let $\mathcal{A}$ be a family of subsets of Q, namely, $\mathcal{A} \subseteq 2^Q$. We define the random variables

$$S = \sum_{A \in \mathcal{A}} I_A, \quad I_A = \prod_{i \in A} J_i.$$

Moreover, we define

$$p_A = E[I_A], \quad \lambda = E[S] = \sum_A p_A, \ \delta = \frac{1}{\lambda} \sum_{A \sim B} E[I_A I_B],$$

where, for $A, B \in \mathcal{A}$, we write $A \sim B$ if $A \cap B \neq \emptyset$ and $A \neq B$.

Theorem 13 ([9, Th. 1]): If η is an integer such that $0 \leqslant \eta \leqslant \lambda$, then

$$\Pr[S \leqslant \eta] \leqslant \left(\sqrt{2\pi(\eta + 1)} \frac{\lambda^\eta}{\eta!} e^{-\lambda} \right)^{\frac{1}{1+\delta}}.$$

Lemma 14: For the $(0, k, p)$-RLL SCS, let

$$\mathcal{F} = \left\{ 1^{k+1} \right\}, \quad P(1^{k+1}) = p \in [0, 1].$$

Then

$$\mathsf{cap}(X^{\text{cyc}}(\mathcal{F}, P)) = \mathsf{cap}(X(\mathcal{F}, P)).$$

Proof: For any $\boldsymbol{\omega} \in \mathcal{B}_n^{\text{cyc}}$, $n \geqslant k + 1$, by definition, $\boldsymbol{\omega}\boldsymbol{\omega}_{0,k} \in \mathcal{B}_{n+k}$. Thus, $|\mathcal{B}_n^{\text{cyc}}| \leqslant |\mathcal{B}_{n+k}|$ for all $n \geqslant k + 1$. In the other direction, we note that for any $\boldsymbol{\omega} \in \mathcal{B}_{n-1}$, one can easily verify that $\boldsymbol{\omega}0 \in \mathcal{B}_n^{\text{cyc}}$, for all $n \geqslant k + 2$. It follows that, for all $n \geqslant k + 2$,

$$|\mathcal{B}_{n-1}| \leqslant \left|\mathcal{B}_n^{\text{cyc}}\right| \leqslant |\mathcal{B}_{n+k}|.$$

Taking the appropriate limits required by the definition of the capacity, we prove the claim. ■

Before stating the upper bound on the capacity of $(0, k, p)$-RLL SCS, we explain briefly how Theorem 13 is going to be used. We conveniently set the index set of a string of length n to be $Q = \{0, 1, \ldots, n - 1\}$. For $(0, k, p)$-RLL SCS we define the family of subsets of Q,

$$\mathcal{A} = \{\{i, i + 1, \ldots, i + k\} : 0 \leqslant i < n\},$$

where the coordinates are taken modulo n. Setting $\eta = pn$, we have that $\Pr[S \leqslant pn]$ is the probability that a sequence of length n obeys the cyclic $(0, k, p)$-RLL SCS, for some integer $0 \leqslant pn \leqslant \lambda$. For this reason, Corollary 12 implies that in case $\eta \geqslant \lambda$ the capacity is $\log_2 2 = 1$.

Theorem 15: For $0 < p \leqslant \frac{1}{2^{k+1}}$, the capacity of the $(0, k, p)$-RLL SCS is bounded by

$$C_{k,p} \leqslant 1 - \frac{1}{3 - 2^{-k+1}} \left(\frac{\log_2 e}{2^{k+1}} + p(k + 1) - p \log_2 \frac{e}{p} \right).$$

Proof: Assume a sequence of n bits are randomly chosen i.i.d. Bernoulli(1/2). It follows that

$$\lambda = E[S] = \sum_A p_A = \frac{n}{2^{k+1}}.$$

For each $A \in \mathcal{A}$ there are exactly $2k$ sets, $B_i \in \mathcal{A}$, $i = 0, 1, \ldots, 2k - 1$, such that $A \sim B_i$. If $|A \cap B_i| = t$ then $E[I_A I_{B_i}] = \frac{1}{2^{2(k+1)-t}}$. Hence,

$$\begin{aligned} \delta &= \frac{1}{\lambda} \sum_{A \in \mathcal{A}} \sum_{B \sim A} E[I_A I_B] = \frac{1}{\lambda} \sum_{A \in \mathcal{A}} \sum_{t=1}^{k} 2 \frac{1}{2^{2(k+1)-t}} \\ &= \frac{1}{\lambda} \sum_{A \in \mathcal{A}} \frac{2^k - 1}{2^{2k}} = 2 - \frac{1}{2^{k-1}}. \end{aligned}$$

Applying Theorem 13 yields

$$\begin{aligned} C_{k,p} &= \limsup_{n \to \infty} \frac{1}{n} \log_2 |B_n| \\ &= \limsup_{n \to \infty} \frac{1}{n} \log_2 2^n \Pr[S \leqslant pn] \\ &= 1 + \limsup_{n \to \infty} \frac{1}{n} \log_2 \Pr[S \leqslant \lfloor pn \rfloor] \\ &\overset{(a)}{\leqslant} 1 + \lim_{n \to \infty} \frac{1}{n} \log_2 \left(\sqrt{2\pi(\lfloor pn \rfloor + 1)} \frac{\lambda^{\lfloor pn \rfloor}}{\lfloor pn \rfloor!} e^{-\lambda} \right)^{\frac{1}{1+\delta}} \\ &= 1 + \lim_{n \to \infty} \frac{1}{(3 - \frac{1}{2^{k-1}})n} \left(-\lambda \log_2 e + \log_2 \frac{\lambda^{\lfloor pn \rfloor}}{\lfloor pn \rfloor!} \right) \\ &= 1 - \frac{\log_2 e}{(3 - \frac{1}{2^{k-1}})2^{k+1}} \\ &\quad + \lim_{n \to \infty} \frac{1}{(3 - \frac{1}{2^{k-1}})n} \log_2 \frac{n^{\lfloor pn \rfloor}}{2^{\lfloor pn \rfloor (k+1)} (\lfloor pn \rfloor !)} \\ &\overset{(b)}{=} 1 - \frac{\log_2 e}{(3 - \frac{1}{2^{k-1}})2^{k+1}} - \frac{p(k + 1)}{(3 - \frac{1}{2^{k-1}})} \\ &\quad + \frac{p}{(3 - \frac{1}{2^{k-1}})} \log_2 \frac{e}{p}, \end{aligned}$$

where (a) follows from Theorem 13 and from the existence of the limit, and (b) follows from Stirling's approximation. Using Lemma 14 we complete the proof. ∎

The same method can be applied for the D-dimensional $(0, k, p)$-RLL SCS, extending the results of [28]. We briefly define the extension of SCS to the multi-dimensional case, and only sketch the proof since it is similar to that of [28].

Define $[n] = \{0, 1, \ldots, n-1\}$, and let $\boldsymbol{e}_j$ be the jth standard unit vector, containing all 0's, except the jth position which is 1. Assume Σ is a finite alphabet, and $\boldsymbol{\omega} \in \Sigma^{[n]^D}$ is an $n \times \cdots \times n$ D-dimensional array over Σ. A substring of $\boldsymbol{\omega}$ is defined as

$$\boldsymbol{\omega}_{\boldsymbol{i},m,\boldsymbol{e}_j} = \omega_{\boldsymbol{i}}\omega_{\boldsymbol{i}+\boldsymbol{e}_j}\ldots\omega_{\boldsymbol{i}+(m-1)\boldsymbol{e}_j},$$

where $\boldsymbol{i} \in [n]^D$ is a D-dimensional index. We note that $\boldsymbol{\omega}_{\boldsymbol{i},m,\boldsymbol{e}_j}$ is a one-dimensional string of length m. We naturally extend $\mathrm{fr}^{\mathrm{cyc}}$ to the D-dimensional case in the following manner: For $\boldsymbol{\omega} \in \Sigma^{[n]^D}$ and $\tau \in \Sigma^*$, the frequency of τ as a cyclic substring of $\boldsymbol{\omega}$ is defined as

$$\mathrm{fr}^{\mathrm{cyc}}(\boldsymbol{\tau}, \boldsymbol{\omega}) = \frac{1}{n^D}\sum_{j=0}^{D-1}\sum_{\boldsymbol{i}\in[n]^D}[\boldsymbol{\omega}_{\boldsymbol{i},|\boldsymbol{\tau}|,\boldsymbol{e}_j} = \boldsymbol{\tau}],$$

where indices are taken modulo n appropriately.

The D-dimensional cyclic $(0, k, p)$-RLL SCS is defined as

$$\mathcal{B}_n^{\mathrm{cyc},(D)}(1^{k+1}, p) = \left\{\boldsymbol{\omega} \in \{0,1\}^{[n]^D} : \mathrm{fr}^{\mathrm{cyc}}(1^{k+1}, \boldsymbol{\omega}) \leqslant p\right\},$$
$$X^{\mathrm{cyc},(D)}(1^{k+1}, p) = \bigcup_n \mathcal{B}_n^{\mathrm{cyc},(D)}(1^{k+1}, p).$$

Its capacity is defined as

$$C_{k,p}^{(D)} = \limsup_{n\to\infty}\frac{1}{n^D}\log_2\left|\mathcal{B}_n^{\mathrm{cyc},(D)}(1^{k+1}, p)\right|.$$

We obtain the following upper bound on $C_{k,p}^{(D)}$:

Theorem 16: The capacity of the D-dimensional $(0, k, p)$-RLL SCS is bounded by the following.

$$C_{k,p}^{(D)} \leqslant 1 - \frac{D\frac{\log_2(e)}{2^{k+1}} + p(k+1) - p\log_2\frac{De}{p}}{3 - 2^{-k+1} + 2^{-k}(D-1)(k+1)^2}.$$

(Sketch of Proof): We use Janson's method, a direct calculation of the expected number of appearances of a sequence of $k+1$ ones, together with direct calculations of the value of δ. We obtain that

$$\lambda = \frac{Dn^D}{2^{k+1}},$$
$$\delta = \frac{1}{\lambda}\sum_{A\in\mathcal{A}}\sum_{B\sim A}E[I_A I_B] = 2 - \frac{1}{2^{k-1}} + \frac{(D-1)(k+1)^2}{2^k}.$$

Placing λ and δ in Theorem 13 yields the wanted result. ∎

We return to the one-dimensional case. The upper bound of Theorem 15 converges to 1 as k grows. We now find the rate of this convergence. To that end we prove a stronger upper bound on the capacity, that does not have as nice a form as Theorem 15 in the finite case, but does have a nice asymptotic form. Note that p must be a function of k since $p \leqslant \frac{1}{2^{k+1}}$.

Theorem 17: For $p = p(k)$, assume $c = \lim_{k\to\infty}\frac{p}{2^{-(k+1)}}$ where $c \in [0, 1]$, and let

$$b_{\mathrm{L}} = \begin{cases} \frac{3-\sqrt{1+8c}}{4}\log_2 e - c\log_2\left(\frac{1+4c+\sqrt{1+8c}}{8c}\right) & c > 0, \\ \frac{1}{2}\log_2 e & c = 0. \end{cases}$$

Then,

$$1 - C_{k,p} \geqslant \frac{b_{\mathrm{L}}}{2^{k+1}}(1 + o(1)),$$

where $o(1)$ denotes a function a of k tending to 0 as $k \to \infty$.

Proof: Let $S'_A = I_A + \sum_{B\sim A} I_B$ and hence, given $I_A = 1$, $S'_A \in \{1, 2, \ldots, 2k+1\}$. We denote $\psi(t) = E[e^{-tS}]$. For all $t \geqslant 0$, it was shown in [9] that

$$-\frac{d}{dt}\ln\psi(t) = \frac{1}{\psi(t)}\sum_{A\in\mathcal{A}}E[I_A e^{-tS}]$$
$$\geqslant \sum_{A\in\mathcal{A}}p_A E[e^{-tS'_A}|I_A = 1]. \quad (3)$$

While [9] bounded $z = E[e^{-tS'_A}|I_A = 1]$, we proceed by calculating it explicitly. Due to symmetry, z does not depend on A or n (for large enough n). Thus, (3) becomes

$$-\frac{d}{dt}\ln\psi(t) \geqslant z\sum_{A\in\mathcal{A}}p_A = \lambda z. \quad (4)$$

However, z does depend on k and t, which we will sometime emphasize by writing $z(k, t)$.

We assume that the length of the sequence is at least $3k$, and recall that we may consider sequences cyclically. A tedious calculation gives

$$\Pr[S'_A = \ell|I_A = 1] = \begin{cases} \frac{\ell}{2^{\ell+1}} & 1 < \ell \leqslant k \\ \frac{2k+4-\ell}{2^{\ell+1}} & k < \ell \leqslant 2k \\ \frac{4}{2^{\ell+1}} & \ell = 2k+1. \end{cases}$$

Thus,

$$z = \sum_{j=1}^{k}\frac{j}{2^{j+1}}e^{-tj} + \sum_{j=k+1}^{2k}\frac{2k+4-j}{2^{j+1}}e^{-tj} + \frac{4}{2^{2k+2}}e^{-t(2k+1)}$$
$$= \frac{2^{-2k}e^{-(1+2k)t}\left(e^t - 2^k e^{(k+1)t} - 1\right)^2}{(1-2e^t)^2}.$$

Since $\psi(0) = 1$, (4) implies

$$-\ln\psi(t) \geqslant \int_0^t \lambda z(k, u)\mathrm{d}u.$$

For $b' \geqslant 0$ and $t \geqslant 0$ we have

$$e^{-tb'\lambda}\Pr[S \leqslant b'\lambda] \leqslant E[e^{-tS}],$$

it follows that

$$\ln\Pr[S \leqslant b'\lambda] \leqslant -\lambda\int_0^t z(k, u)\mathrm{d}u + tb'\lambda.$$

Recall that in our setting we consider the value $\ln\Pr[S \leqslant pn]$ where p is the constraint and n is the length of the sequence. Therefore, since $\lambda = \frac{n}{2^{k+1}}$, we set $b' = p2^{k+1}$ to get $b'\lambda = pn$.

For any k, the following upper bound on the capacity holds for any $t \geqslant 0$,

$$C_{k,p} = 1 + \limsup_{n\to\infty} \frac{1}{n} \log_2 \Pr[S \leqslant pn] \leqslant 1 - \frac{\log_2 e}{2^{k+1}} \int_0^t z(k,u)\mathrm{d}u + tp\log_2 e,$$

where we note that the change of logarithm base introduces a factor of $\log_2 e$. Thus,

$$1 - C_{k,p} \geqslant \frac{\log_2 e}{2^{k+1}} \int_0^t z(k,u)\mathrm{d}u - tp\log_2 e. \quad (5)$$

Note that for any $t \geqslant 0$, by Lebesgue's dominated convergence we obtain that

$$\lim_{k\to\infty} \int_0^t z(k,u)\mathrm{d}u = \int_0^t \lim_{k\to\infty} z(k,u)\mathrm{d}u = \int_0^t \frac{e^u}{(1-2e^u)^2}\mathrm{d}u.$$

It follows that for any fixed $b > 0$, multiplying the right-hand side of (5) by $2^{k+1}/b$ gives,

$$\lim_{k\to\infty} \frac{2^{k+1}\log_2 e}{b}\left(\frac{1}{2^{k+1}}\int_0^t z(k,u)\mathrm{d}u - tp\right) = \frac{\log_2 e}{b}\left(\int_0^t \frac{e^u}{(1-2e^u)^2}\mathrm{d}u - tc\right) = \frac{\log_2 e}{2b}\left(1 + \frac{1}{1-2e^t} - 2tc\right).$$

Thus, to get a bound of the claimed form, this expression must equal 1, i.e.,

$$b = \frac{1}{2}\left(1 + \frac{1}{1-2e^t} - 2tc\right)\log_2 e.$$

Since we are lower-bounding $1 - C_{k,p}$, we would like to maximize this expression by choosing an appropriate value of t. When $c > 0$ the maximum is attained by

$$t = \ln\left(\frac{1 + 4c + \sqrt{1+8c}}{8c}\right),$$

and then we get $b = b_{\mathrm{L}}$, i.e.,

$$b = \frac{3 - \sqrt{1+8c}}{4}\log_2 e - c\log_2\left(\frac{1+4c+\sqrt{1+8c}}{8c}\right).$$

When $c = 0$ we take the limit as $t \to \infty$ to obtain $b = b_{\mathrm{L}} = \frac{1}{2}\log_2 e$, which completes the proof. ■

We note that taking $c = 0$ in Theorem 17 gives $1 - C_{k,0} \geqslant \frac{\log_2 e}{4\cdot 2^k}(1 + o(1))$, which coincides with the capacity's rate of convergence for the fully-constrained system [28].

We also mention that in order to apply this method to more general constraints we can use [10]. Unlike [9], in which the events are positively correlated, in [10] a general dependence structure is allowed. The relevant bounds of [10] have the same flavor as those of [9].

B. A Lower Bound on the Capacity of (0, k, p)-RLL SCS

We turn to consider a lower bound on the capacity of the $(0, k, p)$-RLL SCS. We can extend the method of monotone families that was used in [28] to obtain such a bound. However, the result that we describe next, which is based on the theory of large deviations, outperforms the monotone-families approach.

As mentioned in Theorem 10, the capacity of $(0, k, p)$-RLL SCS is given by

$$\mathsf{cap}(\mathcal{F}, p) = 1 - \inf_{\nu\in\Gamma} I(\nu)$$

where $\Gamma = \left\{\nu \in M_\sigma(\Sigma^k) : \nu(1^{k+1}) \leqslant p\right\}$. The main idea of the bound is that by fixing some $\nu \in \Gamma$ we find a lower bound on the capacity. We do, however, have to keep in mind that the measure we choose must be shift invariant.

Theorem 18: For all $k \geqslant 1$ and $0 < p \leqslant 2^{-(k+1)}$,

$$C_{k,p} \geqslant 1 - \frac{1-p}{2^{k+1}-1}\log_2\left(\frac{2-2p}{1+2p(2^k-1)}\right) - p\log_2\left(\frac{2p(2^{k+1}-1)}{1+2p(2^k-1)}\right). \quad (6)$$

Proof: Construct the following measure,

$$\nu^*(i) = \begin{cases} p & i = 1^{k+1}, \\ \frac{1-p}{2^{k+1}-1} & \text{otherwise.}\end{cases}$$

It is easy to verify that ν^* is indeed a shift-invariant measure. Plugging ν^* into Theorem 10 gives,

$$C_{k,p} = 1 - \inf_{\nu\in\Gamma} I(\nu) \geqslant 1 - I(\nu^*) = 1 - \frac{1-p}{2^{k+1}-1}\log_2\left(\frac{2-2p}{1+2p(2^k-1)}\right) - p\log_2\left(\frac{2p(2^{k+1}-1)}{1+2p(2^k-1)}\right),$$

as claimed. ■

The bound of Theorem 18 can now be used to prove an asymptotic form when $k \to \infty$.

Corollary 19: For $p = p(k)$, assume $c = \lim_{k\to\infty} \frac{p}{2^{-(k+1)}}$ where $c \in [0, 1]$, and let $b_{\mathrm{U}} = (1+c)(1 - H(\frac{1}{c+1}))$, where $H(\cdot)$ is the binary entropy function. Then,

$$1 - C_{k,p} \leqslant \frac{b_{\mathrm{U}}}{2^{k+1}}(1 + o(1)).$$

Proof: We take the limit of the right-hand side of (6) divided by $b_{\mathrm{U}}/2^{k+1}$. We obtain that

$$\lim_{k\to\infty} \frac{2^{(k+1)}}{b_{\mathrm{U}}}\left(\frac{(1-p)\log_2\left(\frac{2-2p}{1+2p(2^k-1)}\right)}{2^{k+1}-1} + p\log_2\left(\frac{2p(2^{k+1}-1)}{1+2p(2^k-1)}\right)\right) = \frac{\log_2\left(\frac{2}{1+c}\right)}{b_{\mathrm{U}}} + \frac{c}{b_{\mathrm{U}}}\log_2\left(\frac{2c}{1+c}\right) = \frac{1}{b_{\mathrm{U}}}\left((1+c) - (1+c)H\left(\frac{1}{1+c}\right)\right) = 1,$$

which proves the claim. ■

In order to obtain a lower bound on the capacity of the D-dimensional $(0,k,p)$-RLL SCS we use the method of monotone families. The bound is recursive in the sense that it is given in terms of the one-dimensional capacity. Thus, the expression may be further simplified by plugging in lower bounds on the one-dimensional capacity from Theorem 18 or Theorem 19. We follow the steps presented in [28], and therefore, only sketch the proof.

Theorem 20: The capacity of the D-dimensional $(0,k,p)$-RLL SCS is bounded by the following,

$$C_{k,p}^{(D)} \geqslant 1 + D\left(C_{k,p/D}^{(1)} - 1\right).$$

(Sketch of Proof): Fix $j \in [D]$, and let A_j denote the set of all $\boldsymbol{\omega} \in \{0,1\}^{[n]^D}$ such that $\boldsymbol{\omega}_{\boldsymbol{i},n,\boldsymbol{e}_j}$ are each one-dimensional $(0,k,p/D)$-RLL semiconstrained strings. As in [28], we note that the D-dimensional $(0,k,p)$-RLL SCS is a superset of the intersection $\bigcap_{j\in[D]} A_j$. Additionally, each A_j is a monotone decreasing family in the sense that it is closed under the operation of turning 1's into 0's. Thus, as in [28, Corollary 8], we obtain the desired result. ■

As a final comment we note that the ratio between the bounds of Theorem 17 and Theorem 19 is at most ≈ 1.5.

IV. Weak Semiconstrained Systems

Let us consider the following two examples of binary semiconstrained systems.

Example 21: Let $X(\mathcal{F},P)$ be an SCS with $\mathcal{F}=\{0,1\}$ and $P(0)=P(1)=\frac{1}{2}$. Note that in this example the limit in the definition of the capacity does not exist. For an even number, n, one can calculate $|\mathcal{B}_n(\mathcal{F},P)|$ and obtain $\binom{n}{n/2}$ which gives $\mathsf{cap}(\mathcal{F},P)>0$. For an odd n we have that $|\mathcal{B}_n(\mathcal{F},P)|=0$. It is easy to construct more examples in the same spirit. □

Example 22: Let $X(\mathcal{F},P)$ be an SCS with $\mathcal{F}=\{0,1\}$ and $P(0)=r, P(1)=1-r$ where $r\in[0,1]$ is an irrational number. We have that the possible words are those with exactly an r-fraction of zeros and a $(1-r)$-fraction of ones. Since the capacity is defined on finite words, for every n we obtain $\mathcal{B}_n(\mathcal{F},P)=\emptyset$, which implies that $\mathsf{cap}(\mathcal{F},P)=-\infty$. □

These two examples are interesting because the first shows that the limit in the definition of the capacity does not always exist, and the second one shows that the capacity is not a continuous function of the restrictions P. That is, in the second example, from Theorem 10 we know that for every $\epsilon>0$,

$$\lim_{n\to\infty}\frac{1}{n}\log_2|\mathcal{B}_n(\mathcal{F},(r+\epsilon,1-r+\epsilon))|>0$$

exists. However, the second example shows that

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log_2|\mathcal{B}_n(\mathcal{F},(r+\epsilon,1-r+\epsilon))| \neq \mathsf{cap}(\mathcal{F},(r,1-r)).$$

We therefore suggest a more relaxed definition of semiconstrained systems.

Definition 23: Let $\mathcal{F}\subseteq\Sigma^*$ be a finite set of words, and let $P\in[0,1]^{\mathcal{F}}$ be a function from $\mathcal{F}$ to the real interval $[0,1]$. *A* weak semiconstrained system (WSCS)*, $\overline{X}(\mathcal{F},P)$, is defined by*

$$\overline{X}(\mathcal{F},P)=\left\{\boldsymbol{\omega}\in\Sigma^*:\forall\boldsymbol{\phi}\in\mathcal{F},\mathrm{fr}(\boldsymbol{\phi},\boldsymbol{\omega})\leqslant P(\boldsymbol{\phi})+\xi(|\boldsymbol{\omega}|)\right\},$$

where $\xi:\mathbb{N}\to\mathbb{R}^+$ is a function satisfying both $\xi(n)=o(1)$ and $\xi(n)=\Omega(1/n)$. In addition, we define

$$\overline{\mathcal{B}}_n(\mathcal{F},P)=\overline{X}(\mathcal{F},P)\cap\Sigma^n.$$

We can think of $\xi(|\boldsymbol{\omega}|)$ as an additive tolerance to the semiconstraints. The requirement that $\xi(n)=o(1)$ is in the spirit of having the WSCS $\overline{X}$ "close" to the SCS X. In the other direction, however, if we were to allow $\xi(n)=o(1/n)$, then for large enough n, we would have gotten $\mathcal{B}_n=\overline{\mathcal{B}}_n$, i.e., no relaxation at all. Thus, we require $\xi(n)=\Omega(1/n)$.

The capacity of WSCS is defined in a similar fashion.

Definition 24: Let $\overline{X}(\mathcal{F},P)$ be a WSCS. The capacity of $\overline{X}(\mathcal{F},P)$, which is denoted by $\overline{\mathsf{cap}}(\mathcal{F},P)$, is defined as

$$\overline{\mathsf{cap}}(\mathcal{F},P)=\limsup_{n\to\infty}\frac{1}{n}\log_2\left|\overline{\mathcal{B}}_n(\mathcal{F},P)\right|.$$

We show that under this definition of the capacity, the limit superior is actually a limit. Moreover, for cases such as the first example, $\mathsf{cap}(\mathcal{F},P)=\overline{\mathsf{cap}}(\mathcal{F},P)$, i.e. the capacity is continuous with respect to the restrictions P. We do however note that weak semiconstrained systems are not a generalization of fully-constrained systems since the set of sequences in the latter does not contain any word which belongs to $\mathcal{F}$, while the former does.

In order to show that the limit in the definition of $\overline{\mathsf{cap}}$ exists we need to show that it is possible to work only over the set of shift-invariant measures with the induced topology. In order to show this, we first show that the rate function governing the LDP as defined in Theorem 10 is a good rate function.

Lemma 25: The rate function governing the LDP of the empirical distribution $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$, as defined in Theorem 10, is a good rate function.

Proof: Recall that the rate function I governing the LDP of the empirical distribution $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$, as defined in Theorem 10, can be written as the relative-entropy function. Let us denote $\ell=|\Sigma|^k$. The set $M_1(\Sigma^k)$ is isomorphic to a closed and bounded subset of $[0,1]^\ell$ and hence, compact. The subset of shift-invariant measures in $M_1(\Sigma^k)$ is a closed subset of $M_1(\Sigma^k)$ as a finite intersection of closed sets. Every closed subset of a compact set is compact and therefore the set of shift-invariant measures on $M_1(\Sigma^k)$ is compact. Since I is a rate function, the level sets are a closed subset of the shift-invariant measures on $M_1(\Sigma^k)$ and hence compact. ■

Corollary 26: Let $\mathcal{X}=M_1(\Sigma^k)$ be the set of probability measures on k-tuples and let $\mathcal{Y}=M_\sigma(\Sigma^k)\subseteq\mathcal{X}$ be the set of shift-invariant probability measures on k-tuples. If $\overline{\boldsymbol{S}}$ is a sequence of i.i.d. symbols, denote $\mu_n(\Gamma)=\Pr[\mathrm{fr}_{n,k}^{\overline{\boldsymbol{S}}}\in\Gamma]$. Let $\mu_n|_{\mathcal{Y}}$ and $I|_{\mathcal{Y}}$ denote the restriction of μ_n and I to $\mathcal{Y}$, respectively. If the sequence $\{\mu_n\}$ satisfies the LDP with rate function I then $\mu_n|_{\mathcal{Y}}$ satisfies the LDP with the rate function $I|_{\mathcal{Y}}$.

Proof: The shift-invariant measures on $M_1(\Sigma^k)$ form a G_δ-subset, and the probability of a sequence to have a measure which is not shift invariant is 0, i.e., $\mu_n(\mathcal{X}\setminus\mathcal{Y})=0$ where $\mathcal{Y}$ is the set of shift-invariant measures. Moreover, by the definition of the rate function governing the LDP of the empirical distribution, $\mathrm{fr}_{n,k}^{\overline{\boldsymbol{Y}}}$, we have $I(x)=\infty$ for $x\in\mathcal{X}\setminus\mathcal{Y}$. Thus, we can use Lemma 8 which allows us to restrict ourselves to the set of shift-invariant measures with the induced topology, which is denoted by $M_\sigma(\Sigma^k)$. ■

The set $\mathcal{F}$ could contain words of various lengths, a fact that sometimes complicates proofs. In order to keep things as simple as possible we would like to work with a set of forbidden words of the same length, i.e., a set $\mathcal{F} \subseteq \Sigma^k$ for some $k \in \mathbb{N}$. The next definition and theorem help us achieve this goal.

Definition 27: Let $\mathcal{F} \subseteq \Sigma^*$ be a finite set of words. Set $k = \max_{\phi\in\mathcal{F}} |\phi|$ and define the operator $f : M_1(\Sigma^k) \to [0,1]^{|\mathcal{F}|}$ as follows. Let $\mathcal{M}$ be a $|\mathcal{F}| \times |\Sigma|^k$ matrix, where for $\phi \in \mathcal{F}$ and $\omega \in \Sigma^k$, the (ϕ, ω) entry is given by

$$\mathcal{M}_{\phi,\omega} = \left[\omega_{0,|\phi|} = \phi\right].$$

Then, for any $\nu \in M_1(\Sigma^k)$, we define $f(\nu) = \mathcal{M}\nu$, where ν is viewed as a vector indexed by Σ^k.

Example 28: Let $\mathcal{F} = \{1, 100\}$ thus $k = 3$. The matrix $\mathcal{M}$ has two rows and eight columns. Each rows corresponds to a word from $\mathcal{F}$: the first row to 1 and the second row to 100. Each column corresponds to a triple 000, 001, 010, ..., 111.

$$\mathcal{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

□

The key point we make is that by using f^{-1} we are able to convert constraints on a general set $\mathcal{F}$ to constraints on k-tuples. The following theorem shows that the limit in the definition of the capacity of the WSCS exists.

Theorem 29: Let $\overline{X}(\mathcal{F}, P)$ be a WSCS, and

$$\Delta = \left\{\nu \in [0,1]^{\mathcal{F}} : \forall \phi \in \mathcal{F}, \nu(\phi) \leqslant P(\phi)\right\}.$$

Set $k = \max_{\phi\in\mathcal{F}} |\phi|$, and let $\mathcal{X} = M_\sigma(\Sigma^k)$ and $\mathcal{Y} = [0,1]^{\mathcal{F}}$. Define the linear function $f : \mathcal{X} \to \mathcal{Y}$ as in Definition 27. Then, if $f^{-1}(\Delta) \cap \mathcal{X} \neq \emptyset$ the following equality holds (and the limit exists):

$$\overline{\mathsf{cap}}(\mathcal{F}, P) = \lim_{n\to\infty} \frac{1}{n} \log_2 \left|\overline{\mathcal{B}}_n(\mathcal{F}, P)\right| = 1 - \inf_{\mu\in f^{-1}(\Delta)} I(\mu),$$

provided the tolerance function for the WSCS satisfies $\xi(n) \geqslant 2|\Sigma|^{k-1} n^{-1}$, and where I is the rate function defined in Theorem 10.

Before proving Theorem 29 we need a technical lemma.

Lemma 30: Let $\overline{S} = S_0, S_1, \cdots$ be a sequence of Σ-valued i.i.d. random variables. Denote $\mathcal{Y} = [0,1]^{|\mathcal{F}|}$, and let $\mathrm{fr}^{\overline{S}}_{n,\mathcal{F}}$ denote the vector of empirical distribution of words in $\mathcal{F}$ in the first n places of $\overline{S}$, i.e., the coordinate $\mathrm{fr}^{\overline{S}}_{n,\mathcal{F}}(\phi)$ that corresponds to $\phi \in \mathcal{F}$ is $\mathrm{fr}(\phi, \overline{S}_{0,n+|\phi|-1})$. Set $k = \max_{\phi\in\mathcal{F}} |\phi|$. Then, for all $n \geqslant k$, and for every Borel set $\Delta \subseteq \mathcal{Y}$,

$$\Pr[\mathrm{fr}^{\overline{S}}_{n,\mathcal{F}} \in \Delta] = \Pr[\mathrm{fr}^{\overline{S}}_{n,k} \in f^{-1}(\Delta)].$$

Proof: Let $\mathrm{fr}^{\overline{S}}_{n,k} = v \in f^{-1}(\Delta)$, and $\mathrm{fr}^{\overline{S}}_{n,\mathcal{F}} = u$. We show that $u \in \Delta$. Note that $f(v) \in \Delta$ hence if $u = f(v)$ we are done. For all $\phi \in \mathcal{F}$, we get,

$$\begin{aligned} f(v)(\phi) &= \sum_{\omega\in\Sigma^k} \mathcal{M}_{\phi,\omega} v(\omega) \\ &= \sum_{\omega\in\Sigma^k} \left[\omega_{0,|\phi|} = \phi\right] \mathrm{fr}(\omega, \overline{S}_{0,n+k-1}) \\ &= \frac{1}{n} \sum_{\omega\in\Sigma^k} \sum_{i=0}^{n-1} \left[\overline{S}_{i,k} = \omega\right] \left[\omega_{0,|\phi|} = \phi\right] \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \left[\overline{S}_{i,|\phi|} = \phi\right] \\ &= \mathrm{fr}(\phi, \overline{S}_{0,n+|\phi|-1}) = u(\phi). \end{aligned}$$

The proof for the other direction is symmetric. ■

TABLE I

EMPIRICAL DISTRIBUTION OF TRIPLES $\mathrm{fr}^{\overline{S}}_{10,3}$ IN THE SEQUENCE $\omega = 101001101000\ldots$

Triple	Distribution	Triple	Distribution
000	$^1/_{10}$	100	$^2/_{10}$
001	$^1/_{10}$	101	$^2/_{10}$
010	$^2/_{10}$	110	$^1/_{10}$
011	$^1/_{10}$	111	0

Example 31: Let $\mathcal{F} = \{1, 100\}$ and $k = 3$ as in Example 28. Consider the sequence $\overline{S}_{0,12} = 101001101000$. The empirical distribution of triples in $\overline{S}_{0,12}$ is shown in Table I.

Thus,

$$v = \left(\frac{1}{10}, \frac{1}{10}, \frac{2}{10}, \frac{1}{10}, \frac{2}{10}, \frac{2}{10}, \frac{1}{10}, 0\right)^T$$

and we have that

$$\mathcal{M}v = \left(\frac{5}{10}, \frac{2}{10}\right)^T.$$

Indeed, $\mathrm{fr}^{\overline{S}}_{10,\mathcal{F}}(1) = \mathrm{fr}(1, \overline{S}_{0,10}) = {}^5/_{10}$ *and* $\mathrm{fr}^{\overline{S}}_{10,\mathcal{F}}(100) = \mathrm{fr}(100, \overline{S}_{0,12}) = {}^4/_{12}$ *and we obtain that* $\mathrm{fr}^{\overline{S}}_{10,\mathcal{F}} = \mathcal{M}\mathrm{fr}^{\overline{S}}_{10,3}$. □

We are now ready to prove Theorem 29.

Proof of Theorem 29: We first note that Δ is closed and hence compact. Let $\{\Delta_n\}$ be the sequence

$$\Delta_n = \left\{\nu \in [0,1]^{\mathcal{F}} : \forall \phi \in \mathcal{F}, \nu(\phi) \leqslant P(\phi) + \frac{2|\Sigma|^{k-1}}{n}\right\}.$$

We let $\Gamma = f^{-1}(\Delta)$, and define the sequence $\{\Gamma_n\}$ where for every n, $\Gamma_n = f^{-1}(\Delta_n)$, i.e.,

$$\Gamma_n = \left\{\nu \in M_\sigma\left(\Sigma^k\right) : \forall \phi \in \mathcal{F}, f(\nu(\phi)) \leqslant P(\phi) + \frac{2|\Sigma|^{k-1}}{n}\right\}.$$

From Corollary 26 we can restrict ourselves to the shift-invariant measures $M_\sigma(\Sigma^k)$. Clearly, for every n, Γ_n is a closed and compact set, and $\Gamma_n \to \Gamma$ when $n \to \infty$ (in the sense that for every open neighborhood U of Γ there exists $N \in \mathbb{N}$ such that for all $n > N$, $\Gamma_n \subseteq U$).

Moreover, $\bigcap_n \Gamma_n = \Gamma$. Since Γ is not empty and is closed, from LD theory we obtain that for every $l \in \mathbb{N}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma_n) \leqslant \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma_l) \leqslant - \inf_{\nu\in\Gamma_l} I(\nu),$$

where $\mu_n(\cdot) = \Pr[\mathrm{fr}^{\overline{S}}_{n,k} \in \cdot]$. Since the rate function is continuous we obtain,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma_n) \leqslant \lim_{l\to\infty} \left(- \inf_{\nu\in\Gamma_l} I(\nu) \right) \overset{(a)}{=} - \inf_{\nu\in\Gamma} I(\nu)$$

where (a) follows from the convergence of Γ_n to Γ.

Now we argue that for every n, the set Γ_n contains a probability measure which belongs to the support of μ_n, i.e., $\exists \omega_1 \in \Gamma_n$ such that $\omega_1 \in \mathrm{fr}^{\overline{S}}_{n,k}$. Moreover, if $q_1 \in \Gamma$ then we have $\omega_1 \in \Gamma_n$ such that $\omega_1 \in \mathrm{fr}^{\overline{S}}_{n,k}$ and $\|q_1 - \omega_1\|_\infty \leqslant \frac{2}{n}$ (see appendix for a proof). Note that a tolerance of $2/n$ becomes, after applying f, a tolerance of at most $2\,|\Sigma|^{k-1}\, n^{-1}$ since the maximum sum of entries in a row of $\mathcal{M}$ is $|\Sigma|^{k-1}$. It is known [1] that

$$\frac{1}{n} \log \mu_n(\omega_1) = -I(\omega_1) + O(n^{-1} \log n).$$

For every $n \in \mathbb{N}$

$$\frac{1}{n} \log \mu_n(\Gamma_n) \geqslant - \inf_{\omega\in\Gamma_n \cap L^{\overline{S}}_{n,k}} I(\omega) + O(n^{-1} \log n).$$

Since the rate function is continuous and since $\|q_1 - \omega_1\|_\infty \leqslant \frac{2}{n}$ we obtain

$$\frac{1}{n} \log \mu_n(\Gamma_n) \geqslant - \inf_{\omega\in\Gamma_n} I(\omega) + O(n^{-1} \log n) + o(1),$$

which implies

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma_n) \geqslant - \inf_{\omega\in\Gamma} I(\omega).$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma_n) = - \inf_{\omega\in\Gamma} I(\omega).$$

From Lemma 30 and since

$$\mu_n(\Gamma_n) = \Pr[\mathrm{fr}^{\overline{S}}_{n,k} \in f^{-1}(\Delta_n)] = \Pr[\mathrm{fr}^{\overline{S}}_{n,\mathcal{F}} \in \Delta_n]$$

we have that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma_n) &= \lim_{n\to\infty} \frac{1}{n} \log \left(\Pr[L^{\overline{S}}_{n,\mathcal{F}} \in \Delta_n] \right) \\ &= - \inf_{\omega\in f^{-1}(\Delta)} I(\omega) \end{aligned}$$

as claimed. ■

The proof shows another important property, namely, the continuity of the capacity as a function of P. Note that the function f is continuous and the rate function is continuous when reduced to its support. Thus, if P is not empty and $P + \boldsymbol{\epsilon}$ is not empty, $\lim_{\boldsymbol{\epsilon}\to 0} \overline{\mathsf{cap}}(\mathcal{F}, P + \boldsymbol{\epsilon}) = \overline{\mathsf{cap}}(\mathcal{F}, P)$.

V. Encoder and Decoder Construction for WSCS

In this section we describe an encoding and decoding scheme for *general* weak semiconstrained systems, that asymptotically achieves capacity. The scheme relies on LD theory, and its implementation is inspired by the coding scheme briefly sketched in [2].

We outline the strategy used to construct the encoder. Given a general semiconstrained system, by LD theory we can solve an optimization problem to find the empirical distribution of k-tuples that both satisfies the semiconstraints, as well as maximizes the entropy. We then use this empirical distribution to construct a Markov chain over a De-Bruijn graph of order $k-1$, with a stationary distribution of edges matching the empirical distribution given by LD theory. We then use this Markov chain to translate a stream of input symbols into symbols that are sent over a channel. The decoder simply reverses the process to obtain the input symbols.

The encoder we present is a block encoder which is also a constant bit rate to constant bit rate encoder. We analyze it for input blocks that contain i.i.d. Bernoulli$(1/2)$ bits. In what follows we present some notation, then describe the encoder and decoder, and finally, analyze the scheme and show its rate is asymptotically optimal, and its probability of failure tends to 0.

A. Preliminaries

Several assumptions will be made in this section, all of them solely for the purpose of simplicity of presentation. We will make these assumptions clear. We further note that the results easily apply to the general case as well.

Let $(\mathcal{F}, P)$ be a WSCS. The first assumption we make is that the system is over the binary alphabet $\Sigma = \{0, 1\}$. Another assumption we make is that $\mathcal{F} \subseteq \Sigma^k$, i.e., every word $\boldsymbol{\phi} \in \mathcal{F}$ is of the same length k (see Theorem 27).

Solving the appropriate LD problem (see Theorem 10) yields the capacity of the system, which is denoted by $C = \overline{\mathsf{cap}}(\mathcal{F}, P)$, together with an optimal probability vector, $\boldsymbol{p}$, of length 2^k. Each entry of the vector $\boldsymbol{p}$ corresponds to a k-tuple and contains the probability that a k-tuple should appear in order to achieve the capacity of the system, as well as satisfy the constraints. We denote the entries $\boldsymbol{p} = (p_0, p_1, \ldots, p_{2^k-1})$.

Let G be the binary De-Bruijn graph of order $k-1$, i.e., the vertices are all the binary $(k-1)$-tuples, and the directed labeled edges are

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_{k-1}) \xrightarrow{u_k} (u_2, u_3, \ldots, u_k) = \boldsymbol{u}', \qquad (7)$$

where $u_i \in \Sigma$. Thus, each vertex has 2 outgoing edges labeled 0 and 1. Additionally, each edge corresponds to a binary k-tuple. For example, the edge from (7) corresponds to $\boldsymbol{u}u_k = u_1\boldsymbol{u}'$.

For convenience, we define an operator $\mathcal{R} : \Sigma^+ \to \Sigma^*$, (where Σ^+ denotes the set of positive-length finite strings over Σ) which removes the first bit of a sequence. Namely, for a sequence $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \Sigma^n$, we define $\mathcal{R}(\boldsymbol{u}) = (u_2, u_3, \ldots, u_n) \in \Sigma^{n-1}$. Thus, the edges of the De-Bruijn graph are of the form $\boldsymbol{u} \to \mathcal{R}(\boldsymbol{u}a)$, for all $\boldsymbol{u} \in \Sigma^{k-1}$ and $a \in \Sigma$. Another operator we require is $\mathcal{L} : \Sigma^+ \to \Sigma$, which maps to the first bit of the sequence. That is, $\mathcal{L}(\boldsymbol{u}) = u_1$.

We can construct a Markov chain over G, whose transition matrix, A, is a $2^{k-1} \times 2^{k-1}$ matrix whose i, j entry, A_{ij}, is the probability of choosing the edge going from vertex $\boldsymbol{u}_i \in \Sigma^{k-1}$ to vertex $\boldsymbol{u}_j \in \Sigma^{k-1}$ given that we are in state $\boldsymbol{u}_i$. At this point, for simplicity of presentation, we assume that from each vertex emanate exactly two outgoing edges with positive probability.

Denote by $\boldsymbol{v} = (v_0, v_1, \ldots, v_{2^{k-1}-1})$ the stationary distribution of the vertices of the Markov chain, i.e., $\boldsymbol{v}$ is the unique left eigenvector of A associated with the eigenvalue 1, whose entry sum is also 1. We would like to find a Markov chain on G whose stationary distribution of the *edges* matches the vector $\boldsymbol{p}$. More precisely, the variables appear in the non-zero entries of A (we have 2^{k-1} variables), and we would like to find a vector $\boldsymbol{v}$ as above (another set of 2^{k-1} variables) satisfying

$$\boldsymbol{v}A = \boldsymbol{v}, \tag{8}$$

as well as, for each edge $\boldsymbol{u}_i \xrightarrow{a} \boldsymbol{u}_j$, $a \in \Sigma$,

$$v_i A_{i,j} = p_{\boldsymbol{u}_i a},$$

where $p_{\boldsymbol{u}_i a}$ is the entry in $\boldsymbol{p}$ that corresponds to the k-tuple $\boldsymbol{u}_i a$. We note that since the vector $\boldsymbol{p}$ is shift invariant, the set of equations has a solution (see [4]).

B. Encoder

Assume $\boldsymbol{\omega} \in \Sigma^n$ is a sequence of n input bits at the encoder, which are i.i.d. Bernoulli $(1/2)$. The encoding process is comprised of three steps: partitioning, biasing, and graph walking.

1) Partitioning: The first step in the encoding process is partitioning the sequence $\boldsymbol{\omega}$ of n input bits into 2^{k-1} subsequences of, perhaps, varying lengths, denoted n_i, $0 \leqslant i \leqslant 2^{k-1} - 1$. Obviously, $n_i \geqslant 0$ for all i, as well as $\sum_{i=0}^{2^{k-1}-1} n_i = n$. Each subsequence is to be associated with a vertex of the Markov chain, or equivalently, with a $(k-1)$-tuple. The first n_0 bits of the input are associated with state $\boldsymbol{u}_0$, the following n_1 bits are associated with state $\boldsymbol{u}_1$, and so on.

For every vertex $\boldsymbol{u}_i$, let $\boldsymbol{u}_j$ be the vertex for which $\boldsymbol{u}_i \xrightarrow{0} \boldsymbol{u}_j$, i.e., $\boldsymbol{u}_j = \mathcal{R}(\boldsymbol{u}_i 0)$, and denote by q_i the entry A_{ij}. For every k-tuple i, let

$$\tilde{n_i} = H(q_i) v_i \cdot \frac{n}{C}.$$

For all $0 \leqslant i \leqslant 2^{k-1} - 1$ take $n_i = \lfloor \tilde{n_i} \rceil$, where $\lfloor \cdot \rceil$ denotes either a rounding down or a rounding up. The rounding is done in such a manner as to preserve the sum,

$$\sum_{i=0}^{2^{k-1}-1} \tilde{n_i} = \sum_{i=0}^{2^{k-1}-1} n_i.$$

This is always possible, for example, by taking 2^{k-1} steps, where at the ith step, $\tilde{n_i}$ is rounded in a direction that keeps the intermediate sum strictly less than 1 away from the original sum. We additionally note that indeed

$$\begin{aligned}
\sum_{i=0}^{2^{k-1}-1} n_i &= \sum_{i=0}^{2^{k-1}-1} \tilde{n_i} \\
&\overset{(a)}{=} \frac{n}{C} \sum_{i \in \Sigma^{k-1}} H(q_i) v_i \\
&= \frac{n}{C} \sum_{i \in \Sigma^{k-1}} \left(-q_i v_i \log_2 q_i - (1-q_i) v_i \log_2 (1-q_i)\right) \\
&= \frac{n}{C} \sum_{i \in \Sigma^{k-1}} \left(-p_{i0} \log_2 q_i - p_{i1} \log_2 (1-q_i)\right) \\
&= \frac{n}{C} \sum_{i \in \Sigma^{k-1}} \bigg(-p_{i0} \log_2 \frac{p_{i0}}{p_{i0}+p_{i1}} \\
&\qquad - p_{i1} \log_2 \frac{p_{i0}}{p_{i0}+p_{i1}}\bigg) \\
&\overset{(b)}{=} \frac{n}{C} \cdot C = n,
\end{aligned}$$

where (a) follows from the one-to-one correspondence between states and $(k-1)$-tuples, and (b) follows from Theorem 10.

2) Biasing: After obtaining 2^{k-1} subsequences, we take each subsequence and bias it to create subsequences that are typical for a Bernoulli (q) source, for some q. To that end, we use an arithmetic decoding process on each subsequence.

Let $\boldsymbol{\eta}_i$ be the subsequence that corresponds to vertex $\boldsymbol{u}_i$, namely,

$$\boldsymbol{\eta}_i = \boldsymbol{\omega}_{\sum_{j=0}^{i-1} n_j, n_i}.$$

For every i, we decode $\boldsymbol{\eta}_i$ using an arithmetic decoder with probability q_i to obtain a new sequence $\hat{\boldsymbol{\eta}}_i$ distributed Bernoulli(q_i). Since the decoding process can continue indefinitely, we stop the process when the obtained sequence $\hat{\boldsymbol{\eta}}_i$ is of length $\left\lceil n_i / H(q_i) + n^{\frac{1}{2}+\epsilon} \right\rceil$ bits for some known arbitrarily small $\epsilon \in (0, \frac{1}{4})$. For every state $\boldsymbol{u}_i$, we call the obtained sequence "the information bits of state $\boldsymbol{u}_i$."

The resulting arithmetically-decoded sequence, $\hat{\boldsymbol{\eta}}_i$, corresponds to a closed segment in $[0, 1]$. If there exists a state $\boldsymbol{u}_i$ for which $\hat{\boldsymbol{\eta}}_i$ corresponds to a segment of length greater than 2^{-n_i}, an error is declared. For a detailed description of arithmetic coding see [27].

3) Graph Walking: The encoder now has the sequences $\hat{\boldsymbol{\eta}}_i$, which are of various lengths. The encoder appends to each sequence $\hat{\boldsymbol{\eta}}_i$ an extra $\left\lceil n^{\frac{1}{2}+2\epsilon} \right\rceil$ bits distributed Bernoulli(q_i). These extra bits carry no information and are used for padding only.

Then, the encoder starts the transmission in the following manner:

Intuitively, when arriving at a state, the encoder takes a bit from the sequence associated with the state. This bit is transmitted, removed from the sequence, and determines the next state. The encoder fails if a bit is required and the sequence associated with the state is already empty, or if at the end of the main loop, not all information bits have been transmitted. This is described formally in Algorithm 1

C. Decoder

The decoding process mirrors the encoding. A simple graph walking is the first stage of the decoding (see Algorithm 2).

After the graph walking is completed, the decoder takes from each received subsequence $\tilde{\boldsymbol{\eta}}_i$ only the first $\left\lceil n_i / H(q_i) + n^{\frac{1}{2}+\epsilon} \right\rceil$ bits and passes them through an arithmetic encoder, thus reversing the second stage of the encoder. The resulting sequences are now $\boldsymbol{\eta}_i$ of length n_i. Finally, the decoder takes each $\boldsymbol{\eta}_i$ and concatenates them in order to obtain the desired input sequence $\boldsymbol{\omega} = \boldsymbol{\eta}_0 \ldots \boldsymbol{\eta}_{2^{k-1}-1}$.

Algorithm 1 Encoding – The Graph-Walking Stage

Input: The sequences $\hat{\boldsymbol{\eta}}_i$
Output: Transmitted bits
$\boldsymbol{u} \leftarrow 0^{k-1}$ ▷ *Set initial state*
repeat
  if $\hat{\boldsymbol{\eta}}_{\boldsymbol{u}}$ is an empty sequence **then**
    Declare error and stop
  end if
  $a \leftarrow \mathcal{L}(\hat{\boldsymbol{\eta}}_{\boldsymbol{u}})$ ▷ *Read first bit in queue*
  Transmit a
  $\hat{\boldsymbol{\eta}}_{\boldsymbol{u}} \leftarrow \mathcal{R}(\hat{\boldsymbol{\eta}}_{\boldsymbol{u}})$ ▷ *Remove first bit from queue*
  $\boldsymbol{u} \leftarrow \mathcal{R}(\boldsymbol{u}a)$ ▷ *Proceed to the next state*
until $\left\lceil \frac{n}{C} + n^{\frac{1}{2}+2\epsilon} \right\rceil$ bits are transmitted
if $\exists \boldsymbol{u} \in \Sigma^{k-1}$ s.t. $\left|\hat{\boldsymbol{\eta}}_{\boldsymbol{u}}\right| > \left\lceil n^{\frac{1}{2}+2\epsilon} \right\rceil$ **then**
  Declare error and stop
end if

Algorithm 2 Decoding – The Graph-Walking Stage

Input: Received bits
Output: The sequences $\hat{\boldsymbol{\eta}}_i$
$\boldsymbol{u} \leftarrow 0^{k-1}$ ▷ *Set initial state*
Set $\hat{\boldsymbol{\eta}}_i$ to be empty sequences, for all i
repeat
  Receive a bit a
  $\hat{\boldsymbol{\eta}}_{\boldsymbol{u}} \leftarrow \hat{\boldsymbol{\eta}}_{\boldsymbol{u}} a$ ▷ Append received bit to queue
  $\boldsymbol{u} \leftarrow \mathcal{R}(\boldsymbol{u}a)$ ▷ *Proceed to the next state*
until $\left\lceil \frac{n}{C} + n^{\frac{1}{2}+2\epsilon} \right\rceil$ bits are received

D. Analysis

We first show that the transmitted sequence indeed admits the constraints given by $\boldsymbol{P}$. Let G be the De-Bruijn graph and A be the associated transition matrix with the stationary distribution vector $\boldsymbol{v}$. It is easy to see that G is irreducible and aperiodic. It is well known that for such graphs, starting with any vertex-probability vector $\boldsymbol{u}$, $\lim_{n\to\infty} \boldsymbol{u}A^n = \boldsymbol{v}$. For $\epsilon > 0$, a divergence of ϵ in some coordinate of $\boldsymbol{v}$ induces a divergence of ϵ in some coordinate in $\boldsymbol{p}$. Although the WSCS allows some tolerance, we need to make sure that the tolerance is indeed $o(1)$. To show that for large enough n the transmitted words satisfy the semiconstraints we need the following theorem.

Theorem 32 ([18, Ch. 4]): Suppose A is the transition matrix of an irreducible and aperiodic Markov chain, with stationary distribution $\boldsymbol{v}$. Then there exist constants $\alpha \in (0, 1)$ and $c > 0$ such that

$$\max_i \|(A^n)_{i,\cdot} - \boldsymbol{v}\|_{TV} \leqslant c\alpha^n,$$

where $(A^n)_{i,\cdot}$ denotes the ith row of A^n, and $\|\cdot\|_{TV}$ denotes the total variation norm.

This implies that the rate of convergence to the stationary distribution is exponential and as such, the divergence from the semiconstraints decays as $o(1)$.

We now examine the rate of the presented coding scheme. The encoder takes n input bits and transmits $\left\lceil \frac{n}{C} + n^{\frac{1}{2}+2\epsilon} \right\rceil$ bits over the channel. Since $\epsilon \in (0, \frac{1}{4})$, the asymptotic rate of the scheme is

$$\lim_{n\to\infty} \frac{n}{\left\lceil \frac{n}{C} + n^{\frac{1}{2}+2\epsilon} \right\rceil} = C,$$

and the coding scheme is asymptotically capacity achieving.

We now show that the error probability vanishes as n grows. We define the following events:

1) E_1: There exists a state $\boldsymbol{u}_i$ for which the arithmetic-decoded word $\hat{\boldsymbol{\eta}}_i$ corresponds to a segment of length greater than 2^{-n_i}.
2) E_2: Some of the information bits have not been transmitted, i.e., there exists a state j which, during the graph-walking stage, is visited strictly less than $\left\lceil n_j/H(q_j) + n^{\frac{1}{2}+\epsilon} \right\rceil$ times.
3) E_3: There exists a state j which, during the graph – walking stage, is visited strictly more than $\left\lceil n_j/H(q_j) + n^{\frac{1}{2}+\epsilon} + n^{\frac{1}{2}+2\epsilon} \right\rceil$ times.

Thus, the total error probability is

$$P_{\text{err}} = \Pr[E_1 \cup E_2 \cup E_3] \leqslant \Pr[E_1] + \Pr[E_2 \cup E_3].$$

We bound the two probabilities appearing on the right-hand side separately, showing each of them vanishes.

We start by considering $\Pr[E_1]$. The arithmetic-coding scheme used here receives a sequence of n_i bits distributed Bernoulli(1/2), employs the *decoding* process first, and then uses the encoding process. The main obstacle in arithmetic decoding is that the arithmetic decoder does not know when to stop the decoding process. In our construction we stop the arithmetic decoder after $\left\lceil n_i/H(q_i) + n^{\frac{1}{2}+\epsilon} \right\rceil$ bits are obtained. It is well-known (for example, see [27]) that the error probability in the arithmetic-coding scheme vanishes as the block length grows, and therefore, using a simple union bound $\Pr[E_1]$ tends to 0.

We continue to the case of bounding $\Pr[E_2 \cup E_3]$. The encoder transmits exactly $\left\lceil n/C + n^{\frac{1}{2}+2\epsilon} \right\rceil$ bits. Let V be the $2^{k-1} \times 2^{k-1}$ matrix all of whose rows are the stationary vector $\boldsymbol{v}$ from (8). We denote by Z the matrix

$$Z = (I - A + V)^{-1},$$

where A is from (8) and I is the identity matrix. The matrix A is invertible by [6, Ch. 11]. We also define, for each i,

$$\sigma_i^2 = 2v_i Z_{ii} - v_i - v_i^2.$$

Let $S_i^{(n)}$ denote the number of times a walk of length n on G visits the vertex i. Let us denote $f(n) = n^{\frac{1}{2}+\epsilon}$ and $g(n) = n^{\frac{1}{2}+2\epsilon}$. For any state $\boldsymbol{u}_i$, it is easy to verify that since $v_i \neq 0, 1$ for every i,

$$\lim_{n\to\infty} \frac{f(n)(1-v_i) - g(n)v_i}{\sqrt{\sigma_i^2\left(\frac{n}{C} + f(n) + g(n)\right)}} \approx \lim_{n\to\infty} n^{\epsilon}(1 - v_i(1+n^{\epsilon})) = -\infty$$

and that

$$\lim_{n\to\infty} \frac{(f(n)+g(n))(1-v_i)}{\sqrt{\sigma_i^2\left(\frac{n}{C} + f(n) + g(n)\right)}} \approx \lim_{n\to\infty} n^{2\epsilon} = \infty.$$

$$
\begin{aligned}
&\Pr\left[\frac{n_i}{H(q_i)} + n^{\frac{1}{2}+\epsilon} < S_i^{\left(\frac{n}{C}+n^{\frac{1}{2}+\epsilon}+n^{\frac{1}{2}+2\epsilon}\right)} < \frac{n_i}{H(q_i)} + n^{\frac{1}{2}+\epsilon} + n^{\frac{1}{2}+2\epsilon}\right] \\
&\quad = \Pr\left[\frac{nv_i}{C} + f(n) < S_i^{\left(\frac{n}{C}+f(n)+g(n)\right)} < \frac{nv_i}{C} + f(n) + g(n)\right] \\
&\quad = \Pr\left[\frac{\frac{nv_i}{C} + f(n) - \left(\frac{n}{C} + f(n) + g(n)\right) v_i}{\sqrt{\left(\frac{n}{C} + f(n) + g(n)\right)\sigma_i^2}} < \frac{S_i^{\left(\frac{n}{C}+f(n)+g(n)\right)} - \left(\frac{n}{C} + f(n) + g(n)\right) v_i}{\sqrt{\left(\frac{n}{C} + f(n) + g(n)\right)\sigma_i^2}}\right. \\
&\qquad\qquad \left. < \frac{\frac{nv_i}{C} + f(n) + g(n) - \left(\frac{n}{C} + f(n) + g(n)\right) v_i}{\sqrt{\left(\frac{n}{C} + f(n) + g(n)\right)\sigma_i^2}}\right] \\
&\quad = \Pr\left[\frac{f(n)(1-v_i) - g(n)v_i}{\sqrt{\left(\frac{n}{C} + f(n) + g(n)\right)\sigma_i^2}} < \frac{S_i^{\left(\frac{n}{C}+f(n)+g(n)\right)} - \left(\frac{n}{C} + f(n) + g(n)\right) v_i}{\sqrt{\left(\frac{n}{C} + f(n) + g(n)\right)\sigma_i^2}} < \frac{(f(n)+g(n))(1-v_i)}{\sqrt{\left(\frac{n}{C} + f(n) + g(n)\right)\sigma_i^2}}\right] \\
&\quad \xrightarrow[n\to\infty]{} \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \mathrm{d}x = 1
\end{aligned} \tag{9}
$$

Using the central limit theorem (CLT) for Markov chains [6, Ch. 11] we bound $1 - \Pr[E_2 \cup E_3]$. For any starting vertex and for every state i, the probability that a walk of length $\left\lceil n/C + n^{\frac{1}{2}+\epsilon} + n^{\frac{1}{2}+2\epsilon} \right\rceil$ on G visits state $\boldsymbol{u}_i$ at least $n_i/H(q_i) + n^{\frac{1}{2}+\epsilon}$ times but no more than $\left\lceil n_i/H(q_i) + n^{\frac{1}{2}+\epsilon} + n^{\frac{1}{2}+2\epsilon} \right\rceil$ times is given in (9), as shown at the top of this page. Thus, as n increases, the probability $\Pr[E_2 \cup E_3]$ tends to 0.

VI. Conclusion

In this paper we studied semiconstrained systems, as well as a relaxation in the form of weak semiconstrained systems. We used tools from probability theory, and in particular, large deviations theory, to formulate closed-form bounds on the capacity of the $(0, k, p)$-RLL SCS. These enabled us to bound the capacity's rate of convergence as k grows. We also examined the limit in the definition of the capacity for these systems does exist, unlike SCS. We also showed the capacity is continuous, again, unlike the case of SCS. Finally, we devised encoding and decoding schemes for WSCS with rate that asymptotically achieves capacity, and with a vanishing failure probability.

Many questions remain open. An important one is the study of multi-dimensional semiconstrained systems, and in particular, the two-dimensional case. While problems concerning two-dimensional fully-constrained systems are notoriously difficult (e.g., finding their exact capacity), perhaps the generalization suggested by semiconstrained systems may bring new insight to the problem. Other research goals suggested by this work are the study of the complexity of the associated encoding and decoding algorithms, as well as the goal of finding encoders that have a zero-probably of failure. We leave these problems for future works.

Appendix

We provide a proof that for every n, the set Γ_n contains a probability measure which belongs to the support of μ_n.

Definition 33: Let $G = (V, E)$ be a directed graph. An n-circulation *is an assignment of weights $w(\cdot)$ to the edges such that:*

1) $w(e) \geqslant 0$ *for all* $e \in E$.
2) $\sum_{e\in \mathrm{In}(v)} w(e) = \sum_{e\in \mathrm{Out}(v)} w(e)$ *for all* $v \in V$.
3) $\sum_{e\in E} w(e) = n$.

An integer n-circulation *is an n-circulation for which $w(e) \in \mathbb{Z}$ for all $e \in E$.*

We assume throughout that a directed graph has no parallel edges.

Definition 34: Let $G = (V, E)$ be a directed graph. A cycle is a sequence $v_0, v_1, \ldots, v_{k-1}$, such that $v_i \in V$, and $(v_i, v_{i+1}) \in E$ for all i, where the indices are taken modulo k. The cycle is vertex simple *if the vertices are all distinct. It is* edge simple *if the edges are all distinct.*

We note that using this notation, a cycle of length 1 is described by a sequence with one vertex only. We say two cycles are distinct if they do not contain the exact same set of edges.

The underlying graph of a direct graph, is the undirected graph obtained by removing the orientation of the edges. An underlying graph may contain parallel edges.

Definition 35: Let $G = (V, E)$ be a directed graph. Let $C = v_0, \ldots, v_{k-1}$ be a cycle in the underlying graph. For all i, we say (v_i, v_{i+1}) is cooriented if $(v_i, v_{i+1}) \in E$, and disoriented if $(v_{i+1}, v_i) \in E$. The set of cooriented and disoriented edges are defined as:

$$
\mathrm{CO}(C) = \{(v_i, v_{i+1}) : (v_i, v_{i+1}) \in E\},
$$
$$
\mathrm{DO}(C) = \{(v_i, v_{i+1}) : (v_i, v_{i+1}) \notin E\}.
$$

We say the effective length *of the cycle is* $\mathrm{CO}(C) - \mathrm{DO}(C)$.

A cycle with effective length of 0 is called *balanced*.

Definition 36: Let $G = (V, E)$ be a directed graph, w a weight assignment to the edges, and $C = v_0, \ldots, v_{k-1}$ an edge-simple cycle in the underlying graph. Let $\epsilon \in \mathbb{R}$. An ϵ-adjustment of the cycle C is a weight assignment w'

such that,

$$w'(e) = \begin{cases} w(e) + \epsilon & e \in \mathrm{CO}(C), \\ w(e) - \epsilon & e \in \mathrm{DO}(C), \\ w(e) & \textit{otherwise.} \end{cases}$$

Lemma 37: Let $G = (V, E)$ be a directed graph, and let C be a balanced edge-simple cycle in the underlying graph. Assume w is an n-circulation, and $\epsilon \in \mathbb{R}$ is some real number. Denote by w' the edge-weighing function obtained from w by ϵ-adjusting C. If $w'(e) \geqslant 0$ for all $e \in E$, then w' is also an n-circulation.

Proof: Property 1 is satisfied by requirement. It is easily verifiable that an adjustment preserves property 2. Finally, the overall weight of the edges is not changed. ■

Lemma 38: Let $G = (V, E)$ be a directed graph, and let C_1 and C_2 be two distinct edge-simple cycles in the underlying graph of effective lengths k_1 and k_2 respectively. Assume w is an n-circulation, and $\epsilon \in \mathbb{R}$ is some real number. Denote by w' the edge-weighing function obtained from w by $k_2\epsilon$-adjusting C_1, and then $-k_1\epsilon$-adjusting C_2. If $w'(e) \geqslant 0$ for all $e \in E$, then w' is also an n-circulation.

Proof: Property 1 is satisfied by requirement. It is easily verifiable that an adjustment preserves property 2. Finally, the overall weight of the edges if increased by $k_1k_2\epsilon$ after the first adjustment, and decreased by the same amount after the second adjustment. ■

Theorem 39: Let $G = (V, E)$ be the De-Bruijn graph of order m over the finite alphabet Σ. Assume w is an n-circulation for some $n \in \mathbb{N}$. Then there exists an integer n-circulation w' such that

$$\lfloor w(e) \rfloor \leqslant w'(e) \leqslant \lceil w(e) \rceil + 1,$$

for all $e \in E$.

Proof: We first look at the underlying unoriented graph. This is a regular graph of degree $2|\Sigma|$. Because of property 2, there is no vertex with exactly one incident edge of non-integer weight. It follows, that every edge of non-integer weight is on an unoriented cycle in the underlying graph, all of whose edges have non-integer weights. We call such cycles, *non-integer cycles*.

Assume there is a balanced non-integer cycle C. By Lemma 37, and since all of the weights on the cycle's edges are non-integers, there exists a minimal $\epsilon > 0$ such that ϵ-adjusting C creates a new n-circulation with at least one edge of the cycle having an integer weight. Furthermore, for this edge e, since we took the minimal ϵ possible, the new weight of the edge is either $\lfloor w(e) \rfloor$ or $\lceil w(e) \rceil$.

We can repeat the process as long as we have balanced non-integer cycles. If we do not, assume we have two distinct non-integer cycles, C_1 and C_2. We can assume they are edge simple. Again, there exists a minimal $\epsilon > 0$ such that adjusting by Lemma 38 turns at least one of the cycle-edge weights to an integer weight. Like before, choosing the minimal such ϵ ensures the new weight is either a rounding down or a rounding up of the original weight.

After this, we go back to looking for balanced non-integer cycles, an continue this way. Repeating the above, we must end up with either an integer n-circulation w' as desired, or with a circulation all of whose non-integer weights form a single vertex-simple non-balanced non-integer cycle. Denote this cycle as C, and assume it has an effective length of k. It is easy to verify that the fractional part of the weight of all cooriented edges is equal to some constant $0 < \alpha < 1$, whereas the fractional part of the disoriented edges is $1 - \alpha$. Since the sum of the edges of C is an integer, we have

$$|\mathrm{CO}(C)|\,\alpha + |\mathrm{DO}(C)|\,(1 - \alpha) = |\mathrm{DO}(C)| + k\alpha,$$

is an integer. Thus, $k\alpha$ is an integer and $0 < k\alpha < k$.

It is well known [17] that the De Bruijn graph of order m has an edge-simple directed cycle for each length between 1 and $|\Sigma|^m$. We find such a cycle of length $k\alpha$. We then round down all the weights of the cycle C, and add 1 to all the edges of the $k\alpha$-cycle. We call the resulting n-circulation w'. Since the weights of the edges of the $k\alpha$ cycle may have already been increased in a previous rounding operation, we have

$$\lfloor w(e) \rfloor \leqslant w'(e) \leqslant \lceil w(e) \rceil + 1,$$

for all $e \in E$, as claimed. ■

Finally, in the proof of Theorem 29 we are given q_1, a shift-invariant distribution over Σ^k. By identifying the elements of Σ^k with the edges of the De Bruijn graph of order $k - 1$ over Σ, and assigning each edge $\boldsymbol{\phi} \in \Sigma^k$ the weight $n \cdot q_1(\boldsymbol{\phi})$, we obtain a circulation. Using Theorem 39, we can obtain an integer circulation, which we denote w'. If we define $\omega_1 = w'/n$, then ω_1 is a shift-invariant distribution in $L^{\overline{S}}_{n,k}$ satisfying

$$\|q_1 - \omega_1\|_\infty \leqslant \frac{2}{n},$$

as claimed.

ACKNOWLEDGMENTS

The authors would like to thank the associate editor and the anonymous reviewers, whose comments improved the presentation of the paper.

REFERENCES

[1] P. H. Algoet and B. H. Marcus, "Large deviation theorems for empirical types of Markov chains constrained to thin sets," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1276–1291, Jul. 1992.

[2] S. Aviran, P. H. Siegel, and J. K. Wolf, "An improvement to the bit stuffing algorithm," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2885–2891, Aug. 2005.

[3] A. Berman and Y. Birk, "Constrained flash memory programming," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Aug. 2011, pp. 2128–2132.

[4] J.-R. Chazottes, J.-M. Gambaudo, M. Hochman, and E. Ugalde, "On the finite-dimensional marginals of shift-invariant measures," *Ergodic Theory Dyn. Syst.*, vol. 32, no. 5, pp. 1485–1500, 2012.

[5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY, USA: Springer, 1998.

[6] C. M. Grinstead and L. J. Snell, *Introduction to Probability*. Providence, RI, USA: AMS, 2006.

[7] K. A. S. Immink, "Weakly constrained codes," *Electron. Lett.*, vol. 33, no. 23, pp. 1943–1944, Nov. 1997.

[8] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Eindhoven, The Netherlands: Shannon Foundation Publishers, 2004.

[9] S. Janson, "Poisson approximation for large deviations," *Random Struct. Algorithms*, vol. 1, no. 2, pp. 221–230, 1990.

[10] S. Janson, “Large deviations for sums of partly dependent random variables,” *Random Struct. Algorithms*, vol. 24, no. 3, pp. 234–248, 2004.
[11] R. Karabed, D. L. Neuhoff, and A. Khayrallah, “The capacity of costly noiseless channels,” IBM Res. Rep., New York, NY, USA, Tech. Rep. RJ 6040 (59639), Jan. 1988.
[12] A. Kato and K. Zeger, “On the capacity of two-dimensional run-length constrained channels,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1527–1540, Jul. 1999.
[13] S. Kayser and P. H. Siegel, “Constructions for constant-weight ICI-free codes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 1431–1435.
[14] A. S. Khayrallah and D. L. Neuhoff, “Coding for channels with cost constraints,” *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 854–867, May 1996.
[15] H. M. Kiah, G. J. Puleo, and O. Milenkovic. (2015). “Codes for DNA sequence profiles.” [Online]. Available: http://arxiv.org/abs/1502.00517.
[16] O. F. Kurmaev, “Constant-weight and constant-charge binary run-length limited codes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4497–4515, Jul. 2011.
[17] A. Lempel, “m-ary closed sequences,” *J. Combinat. Theory, A*, vol. 10, no. 3, pp. 253–258, 1971.
[18] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*. Providence, RI, USA: AMS, 2006.
[19] S. Lin and D. J. Costello, *Error Control Coding*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.
[20] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[21] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1978.
[22] B. Marcus and S. Tuncel, “Entropy at a weight-per-symbol and embeddings of Markov chains,” *Inventiones Math.*, vol. 102, no. 1, pp. 235–266, 1990.
[23] 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.
[24] M. Qin, E. Yaakobi, and P. H. Siegel, “Constrained codes that mitigate inter-cell interference in read/write cycles for flash memories,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 836–846, May 2014.
[25] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[26] R. M. Roth, *Introduction to Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[27] A. Said, “Introduction to arithmetic coding—Theory and practice,” Hewlett Packard Lab., Palo Alto, CA, USA, Tech. Rep. HPL-2004-76, Apr. 2004.
[28] M. Schwartz and A. Vardy, “New bounds on the capacity of multidimensional run-length constraints,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4373–4382, Jul. 2011.
[29] A. Shafarenko, A. Skidin, and S. K. Turitsyn, “Weakly-constrained codes for suppression of patterning effects in digital communications,” *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2845–2854, Oct. 2010.
[30] A. Shafarenko, K. S. Turitsyn, and S. K. Turitsyn, “Information-theory analysis of skewed coding for suppression of pattern-dependent errors in digital communications,” *IEEE Trans. Commun.*, vol. 55, no. 2, pp. 237–241, Feb. 2007.
[31] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

Ohad Elishco (S'12) received the B.Sc. in mathematics and his B.Sc. in electrical engineering in 2012 from Ben-Gurion University of the Negev, Israel; the M.Sc. degree in electrical engineering in 2013 from Ben-Gurion University of the Negev. In 2013 he started his Ph.D in electrical engineering also in Ben-Gurion University. His research interests are constrained coding and dynamical systems.

Tom Meyerovitch received the B.Sc. in mathematics and computer science from from Tel-Aviv University, Israel, in 1999; the M.Sc. and Ph.D. degrees in mathematics in 2004 and 2010, respectively, also from Tel-Aviv University. In 2010-2012 he held a post doctoral position at the University of British Columbia, and the Pacific Institute for Mathematical studies in Vancouver Canada. Since 2012, he has been on the faculty of the Department of Mathematics at Ben-Gurion University, Beer-Sheva, Israel. He studies various mathematical aspects of dynamical systems, in particular symbolic dynamics.

Moshe Schwartz (M'03–SM'10) is an associate professor at the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include algebraic coding, combinatorial structures, and digital sequences.

Prof. Schwartz received the B.A. (*summa cum laude*), M.Sc., and Ph.D. degrees from the Technion - Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004 respectively, all from the Computer Science Department. He was a Fulbright post-doctoral researcher in the Department of Electrical and Computer Engineering, University of California San Diego, and a post-doctoral researcher in the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012-2014, he was a visiting scientist at the Massachusetts Institute of Technology (MIT).

Prof. Schwartz received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage, and the 2010 IEEE Communications Society Best Student Paper Award in Signal Processing and Coding for Data Storage.