Semiconstrained Systems

Ohad Elishco
Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer Sheva 8410501, Israel
ohadeli@bgu.ac.il

Tom Meyerovitch
Department of Mathematics
Ben-Gurion University of the Negev
Beer Sheva 8410501, Israel
mtom@math.bgu.ac.il

Moshe Schwartz
Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer Sheva 8410501, Israel
schwartz@ee.bgu.ac.il

***Abstract*—When transmitting information over a noisy channel, two approaches 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 model, which is an extension of the channel with cost constraints, 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 study. 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.

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. Two common solutions to this problem were already described in Shannon's work [20]. The first solution 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 [15]). The second solution 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, and they have also been extensively studied and used (for example, see [7], [14]).

The two solutions 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 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 a middle road we call semiconstrained systems.

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 [7] 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.

While the capacity of the semiconstrained $(0,k)$-RLL system is known using the methods of [17], 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 [19]. 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 [19] as well 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 [13] 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 [11]) correspond to a semiconstrained system.

Although coding schemes for some of these systems exist, they are ad-hoc and tailored for each specific case, as in [13] and [11]. A more general coding scheme exists [12] for a channel with cost constraints model. However, it is not optimal, and it addresses only scalar cost functions, which is a different model than the semiconstrained systems we study.

The second main contribution of this paper is a general explicit 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 [9], [12] 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], [16], [17]. In [16] some embedding theorems and results concerning the entropy of a weight-per-symbol shift of finite type are presented, where the weights are given by functions which take values in $\mathbb{R}^d$. In [1], some large-deviation theorems are proved for empirical types of Markov chains that are constrained to thin sets. We also mention [17], in which an improved Gilbert-Varshamov bound for fully-constrained systems is found. Thus, [17] 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.

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 III we introduce a relaxation called weak semiconstrained systems, and study issues involving the existence of the limit in the capacity definition. In Section IV 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. Due to space constraints, some proofs are given in the appendix, and others are omitted. For a full version the reader is referred to the preprint [6].

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 $\boldsymbol{\omega}_{i,m} = \omega_i\omega_{i+1} \dots \omega_{i+m-1}$. We say $\boldsymbol{\omega}'$ is a *proper subword* of $\boldsymbol{\omega} \in \Sigma^*$ if $\boldsymbol{\omega}'$ is a subword of $\boldsymbol{\omega}$, and $|\boldsymbol{\omega}'| < |\boldsymbol{\omega}|$.

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.

For any two words $\boldsymbol{\tau}, \boldsymbol{\omega} \in \Sigma^*$, let $T(\boldsymbol{\tau}, \boldsymbol{\omega})$ denote the frequency of $\boldsymbol{\tau}$ as a subword of $\boldsymbol{\omega}$, i.e., $T(\boldsymbol{\tau}, \boldsymbol{\omega}) = \frac{1}{|\boldsymbol{\omega}|-|\boldsymbol{\tau}|+1} \sum_{i=0}^{|\boldsymbol{\omega}|-|\boldsymbol{\tau}|} [\boldsymbol{\omega}_{i,|\boldsymbol{\tau}|} = \boldsymbol{\tau}]$. 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 $T(\boldsymbol{\tau}, \boldsymbol{\omega}) = 0$.

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,*

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

We also define the set of words of length exactly n in $X(\mathcal{F}, P)$ as $\mathcal{B}_n(\mathcal{F}, P) = X(\mathcal{F}, P) \cap \Sigma^n$. An important object of interest is the capacity of an SCS.

Definition 2. *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 $(\mathcal{F}, P)$. Following the same logic as [19], we replace 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 $\mathsf{cap}(\mathcal{F}, P) = \log_2 |\Sigma| + \limsup_{n \to \infty} \frac{1}{n} \log_2 p_n$.

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 3. *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 4. *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 $\Gamma \subseteq \mathcal{X}$, $-\inf_{x \in \Gamma^\circ} I(x) \leqslant \liminf_{n \to \infty} \frac{1}{n} \log_2 \mu_n(\Gamma) \leqslant \limsup_{n \to \infty} \frac{1}{n} \log_2 \mu_n(\Gamma) \leqslant -\inf_{x \in \overline{\Gamma}} I(x)$.*

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

Definition 5. *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 $L_n^{\overline{\boldsymbol{Y}}}(\boldsymbol{y})$ the empirical occurrence frequency of the word $\boldsymbol{y}$ in the first n places of $\overline{\boldsymbol{Y}}$, $L_n^{\overline{\boldsymbol{Y}}}(\boldsymbol{y}) = T(\boldsymbol{y}, \overline{\boldsymbol{Y}}_{0,n+|\boldsymbol{y}|-1})$. We denote by $L_{n,k}^{\overline{\boldsymbol{Y}}} \in M_1(\Sigma^k)$ 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 $L_{n,k}^{\overline{\boldsymbol{Y}}}$ is $L_n^{\overline{\boldsymbol{Y}}}(\boldsymbol{y})$.*

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 $q(\sigma_0, \sigma_1, \ldots, \sigma_{k-1})$ the probability of the sequence $\sigma_0, \sigma_1, \ldots, \sigma_{k-1}$. The following theorem connects the empirical distribution with the large-deviation principle.

Theorem 6. *[5, §3.1] Let $\nu \in \mathcal{X} = M_1(\Sigma^k)$, and let $\overline{Y} = Y_0, Y_1, \ldots$ 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 $\mu_n(\Gamma) = \Pr[L^{\overline{Y}}_{n,k} \in \Gamma]$. 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, $\nu_1(\sigma_0, \ldots, \sigma_{k-2}) = \sum_{\sigma \in \Sigma} \nu(\sigma_0, \ldots, \sigma_{k-2}, \sigma)$. Then the rate function, $I : \mathcal{X} \to [0, \infty]$, governing the LDP of the empirical distribution $L^{\overline{Y}}_{n,k}$ with respect to Γ is, $I(\nu) = \sum_{\sigma \in \Sigma^k} \nu(\boldsymbol{\sigma}) \log_2 \frac{\nu(\sigma)}{\nu_1(\sigma_{0,k-1}) q(\sigma_{k-1})}$ if ν is shift invariant, and ∞ otherwise, where $\nu \in \mathcal{X} = M_1(\Sigma^k)$ is shift invariant if $\sum_{\sigma \in \Sigma} \nu(\sigma, \sigma_1, \ldots, \sigma_{k-1}) = \sum_{\sigma \in \Sigma} \nu(\sigma_1, \ldots, \sigma_{k-1}, \sigma)$.*

Finally, the following corollary shows cases in which the constraints in P are redundant. (proof omitted)

Corollary 7. *Let $\mathcal{F} \subseteq \Sigma^k$. If $P(\boldsymbol{\phi}) \geqslant |\Sigma|^{-k}$ for all $\boldsymbol{\phi} \in \mathcal{F}$, then $\mathsf{cap}(\mathcal{F}, P) = \log_2 |\Sigma|$.*

III. The Existence of the Limit in the Capacity Definition and Weak Semiconstrained Systems

Consider the following examples of binary semiconstraints.

Example 8. *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 9. *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 $\boldsymbol{p}$. We therefore suggest a more relaxed definition of semiconstrained systems.

Definition 10. *Let $\mathcal{F} \subseteq \Sigma^*$ be a finite set of words, and let $P \in [0, 1]^{\mathcal{F}}$. A* weak semiconstrained system (WSCS), *$\overline{X}(\mathcal{F}, P)$, is defined by*

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

where $\xi : \mathbb{N} \to \mathbb{R}^+$ is a function satisfying both $\xi(n) = o(1)$ and $\xi(n) = \Omega(1/n)$. Also $\overline{\mathcal{B}}_n(\mathcal{F}, P) = \overline{X}(\mathcal{F}, P) \cap \Sigma^n$.

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 $\xi(n) = \Omega(1/n)$.

Definition 11. *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. (for the proof, see [6])

Definition 12. *Let $\mathcal{F} \subseteq \Sigma^*$ be a finite set of words. Set $k = \max_{\boldsymbol{\phi} \in \mathcal{F}} |\boldsymbol{\phi}|$ 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 $\boldsymbol{\phi} \in \mathcal{F}$ and $\boldsymbol{\omega} \in \Sigma^k$, the $(\boldsymbol{\phi}, \boldsymbol{\omega})$ entry is given by $\mathcal{M}_{\boldsymbol{\phi}, \boldsymbol{\omega}} = [\boldsymbol{\omega}_{0, |\boldsymbol{\phi}|} = \boldsymbol{\phi}]$. 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.*

Theorem 13. *Let $\overline{X}(\mathcal{F}, P)$ be a WSCS, and*

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

Set $k = \max_{\boldsymbol{\phi} \in \mathcal{F}} |\boldsymbol{\phi}|$, and let $\mathcal{X} = M_\sigma(\Sigma^k)$ and $\mathcal{Y} = [0, 1]^{\mathcal{F}}$ (which is isomorphic to $[0, 1]^{|\mathcal{F}|}$). Define the linear function $f : \mathcal{X} \to \mathcal{Y}$ as in Definition 12. 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 from Theorem 6.

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, $\lim_{\boldsymbol{\epsilon} \to 0} \overline{\mathsf{cap}}(\mathcal{F}, P + \boldsymbol{\epsilon}) = \overline{\mathsf{cap}}(\mathcal{F}, P)$.

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

In this section we study 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. To that end, 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 some of them may be extended to general semiconstrained systems.

We consider the semiconstrained system $X(\mathcal{F}, P)$ defined by $\mathcal{F} = \{0^{k+1}\}$ and $P(0^{k+1}) = p$, 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. We note that due to Corollary 7 we must also take $p \leqslant \frac{1}{2^{k+1}}$ or else the capacity is exactly 1. In the case of fully-constrained $(0, k)$-RLL, the asymptotics of the capacity, as k tends to infinity, are well known [10], [19].

In order to obtain an upper bound on the capacity of $(0,k,p)$-RLL SCS we employ a result of Janson [8]. The first bound we present is the following: (proof in [6])

Theorem 14. *For $0 < p \leqslant \frac{1}{2^{k+1}}$, the capacity of the $(0,k,p)$-RLL SCS is bounded by*

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

The same method can be applied for the D-dimensional $(0,k,p)$-RLL SCS, extending the results of [19]. The definition of semiconstrained systems is extended in the natural way to higher dimensions. We obtain the following upper bound on $C_{k,p}^{(D)}$: (proof omitted)

Theorem 15. *The capacity of the D-dimensional $(0,k,p)$-RLL SCS is bounded by the following.*

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

We return to the one-dimensional case. The upper bound of Theorem 14 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 14 in the finite case, but does have a nice asymptotic form. (proof in [6])

Theorem 16. *For $p = p(k)$ let $c = \lim_{k\to\infty}\frac{p}{2^{-(k+1)}}$, $c \in [0,1]$, and let*

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

Then, $1 - C_{k,p} \geqslant \frac{b_{\mathrm{L}}}{2^{k+2}}(1+o(1))$, where $o(1)$ denotes a function of k tending to 0 as $k \to \infty$.

We note that taking $c = 0$ in Theorem 16 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 [19].

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 [19] 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 in Theorem 6, the capacity of $(0,k,p)$-RLL SCS is given by $\mathsf{cap}(\mathcal{F},p) = 1 - \inf_{\nu\in\Gamma} I(\nu)$, where $\Gamma = \{\nu \in M_\sigma(\Sigma^k) : \nu(1^{k+1}) \leqslant p\}$. 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. (proof in [6])

Theorem 17. *For all $k \geqslant 1$ and $0 \leqslant p \leqslant 2^{-(k+1)}$,*

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

The bound of Theorem 17 can now be used to prove an asymptotic form when $k \to \infty$. (proof in [6])

Theorem 18. *For $p = p(k)$ let $c = \lim_{k\to\infty}\frac{p}{2^{-(k+1)}}$, 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, $1 - C_{k,p} \leqslant \frac{b_{\mathrm{U}}}{2^{k+1}}(1+o(1))$.*

The ratio between the bounds of Theorem 18 and Theorem 16 is at most ≈ 1.5. Again, a higher-dimensional bound may be obtained (proof omitted).

Theorem 19. *The capacity of the D-dimensional $(0,k,p)$-RLL SCS is bounded by $C_{k,p}^{(D)} \geqslant 1 + D\left(C_{k,p/D}^{(1)} - 1\right)$.*

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].

The encoder we present is a block encoder. We analyze it for input that contains i.i.d. Bernoulli$(1/2)$ bits. We first 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.

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 $(\mathcal{F},\boldsymbol{P})$ 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 [6]).

Solving the appropriate LD problem (see Theorem 6) yields the capacity of the system, which is denoted by $C = \overline{\mathsf{cap}}(\mathcal{F},\boldsymbol{P})$, 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 $\boldsymbol{u} = (u_1,u_2,\ldots,u_{k-1}) \xrightarrow{u_k} (u_2,u_3,\ldots,u_k) = \boldsymbol{u'}$, 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 convenience, we define an operator $\mathcal{R} : \Sigma^+ \to \Sigma^*$, for $\boldsymbol{u} = (u_1,u_2,\ldots,u_n) \in \Sigma^n$, $\mathcal{R}(\boldsymbol{u}) = (u_2,u_3,\ldots,u_n) \in \Sigma^{n-1}$. 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 $\mathcal{L} : \Sigma^+ \to \Sigma$, $\mathcal{L}(\boldsymbol{u}) = u_1$.

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. We would like to find a Markov chain on G whose stationary distribution of the *edges* matches the vector $\boldsymbol{p}$. We note that since the vector $\boldsymbol{p}$ is shift invariant, the set of equations has a solution (see [4]).

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.

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 $\tilde{n}_i = H(q_i) v_i n / C$. 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, $\sum_{i=0}^{2^{k-1}-1} \tilde{n}_i = \sum_{i=0}^{2^{k-1}-1} n_i$ (proof omitted).

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 $\boldsymbol{u}_i$. 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). We stop the decoder 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."

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 [18].

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 as described in Algorithm 1.

The decoding process mirrors the encoding. Details may be found in [6]. Additionally, the following theorem summarizes the properties of the procedures. (proof in [6])

Theorem 20. *The encoding procedure described above produces the desired WSCS, asymptotically achieves capacity, and has a vanishing failure probability.*

Algorithm 1 Encoding – The Graph-Walking Stage

```
u ← 0^{k-1}                                   ▷ Set initial state
repeat
    if η̂_u is an empty sequence then
        Declare error and stop
    end if
    a ← L(η̂_u)                                ▷ Read first bit in queue
    Transmit a
    η̂_u ← R(η̂_u)                              ▷ Remove first bit from queue
    u ← R(ua)                                  ▷ Proceed to the next state
until ⌈n/C + n^{1/2+2ε}⌉ bits are transmitted
if ∃u ∈ Σ^{k-1} s.t. |η̂_u| > ⌈n^{1/2+2ε}⌉ then
    Declare error and stop
end if
```

References

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

[2] S. Aviran, P. H. Siegel, and J. K. Wolf, "An improvement to the bit stuffing algorithm," *IEEE Trans. Inform. Theory*, pp. 2885–2891, 2005.

[3] A. Berman and Y. Birk, "Constrained flash memory programming," in *Proceedings of the 2011 IEEE International Symposium on Information Theory (ISIT2011), 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: Springer, 1998.

[6] O. Elishco, T. Meyerovitch, and M. Schwartz, "Semiconstrained systems," Jan. 2015. [Online]. Available: http://arxiv.org/abs/1401.2612

[7] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Shannon Foundation Publishers, 2004.

[8] S. Janson, "Poisson approximation for large deviations," *Random Structures and Algorithms*, vol. 1, pp. 221–230, 1990.

[9] R. Karabed, D. Neuhoff, and A. Khayrallah, "The capacity of costly noiseless channels," IBM Research Report, Tech. Rep. RJ 6040 (59639), Jan. 1988.

[10] A. Kato and K. Zeger, "On the capacity of two-dimensional run-length constrained channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1527–1540, Jul. 1999.

[11] S. Kayser and P. H. Siegel, "Constructions for constant-weight ICI-free codes," in *Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT2014), Honolulu, HI, USA*, Jul. 2014, pp. 1431–1435.

[12] A. S. Khayrallah and D. L. Neuhoff, "Coding for channels with cost constraints," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 854–867, 1996.

[13] O. F. Kurmaev, "Constant-weight and constant-charge binary run-length limited codes," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4497–4515, Jul. 2011.

[14] D. Lind and B. H. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1985.

[15] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. North-Holland, 1978.

[16] B. Marcus and S. Tuncel, "Entropy at a weight-per-symbol and embeddings of Markov chains," *Inventiones Mathematicae*, vol. 102, no. 1, pp. 235–266, 1990.

[17] B. H. Marcus and R. M. Roth, "Improved Gilbert-Varshamov bound for constrained systems," *IEEE Transactions on Information Theory*, vol. 38, no. 4, pp. 1213–1221, 1992.

[18] A. Said, "Introduction to arithmetic coding – theory and practice," Hewlett Packard Laboratories, Tech. Rep. HPL-2004-76, Apr. 2004.

[19] M. Schwartz and A. Vardy, "New bounds on the capacity of multidimensional run-length constraints," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4373–4382, Jul. 2011.

[20] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, Jul. 1948.