Weakly Constrained Codes via Row-by-Row Coding

Sarit Buzaglo and **Paul H. Siegel**
University of California San Diego, La Jolla, CA 92093, USA
{*sbuzaglo, psiegel*}*@ucsd.edu*

***Abstract*—A *constrained code* is a set of finite-length codewords that entirely avoid the occurrences of certain patterns. In some applications, it may be preferable to merely limit the number of occurrences of certain patterns in codewords rather than to completely forbid them. Constrained codes that involve such weaker constraints are called *weakly constrained codes*.**

In this paper we construct capacity-achieving weakly constrained codes. The construction is based on a *row-by-row coding scheme* in which messages are encoded into the rows of a 2-dimensional array in which the frequency of occurrence of patterns along columns is controlled.

I. Introduction

In many applications in coding and information theory there is a need to completely avoid or limit the number of occurrences of certain patterns as substrings of codewords. One example of such an application is a *multi-level cell* flash memory. In this example, programming the memory cells into certain levels wear out the cells. Hence, it is preferable to use the levels that are more likely to damage the cells less often [9], [12]. More examples arise in every application that involves a transmission over a channel in which some patterns may cause errors when they appear as substrings of codewords (e.g., [2]).

Constrained sequences are sequences that completely avoid certain unwanted patterns as their substrings. A *constrained code* is simply a set of constrained sequences. These codes have been studied extensively (a survey can be found in [14]). However, in some cases the rate penalty that is incurred by imposing such strong constraints is too severe, and there is a need to weaken the constraints so as to allow some of the bad patterns to appear, yet not very often. Constrained codes that admit such weak constraints are called *weakly constrained codes*.

The study of weakly constrained codes is much less extensive, however. Moreover, in the previous literature one finds somewhat different definitions of weak constraints that arise from the specific restrictions imposed on the statistics of some of the allowed patterns in codewords. For example, in [4] the authors impose only upper bounds on these statistics, and to avoid having zero-capacity codes, they further weaken the upper bounds by adding a *tolerance factor* that approaches zero as the code-length grows to infinity. On the other hand, in [13], [15] the authors impose both lower and upper bounds on the statistics of some patterns and use a constant tolerance factor. In this paper we propose a third definition in which both lower and upper bounds are imposed but the tolerance factor is a vanishing function of the code-length.

In all of these cases, there exist analytic expressions for the capacity of the corresponding constraints. More specifically, for classical constraints as well as weak constraints, code sequences can be obtained by reading the labels of paths in some labeled directed graph, and the structure of the graph and the labeling function guarantee that every path produces a legal constrained sequence. The capacity of the weakly constrained sequences can then be expressed in terms of the entropy of a suitably defined *Markov chain* on the graph. An expression for a capacity-achieving Markov chain was given in [10] for weak (d, k)-RLL (runlength-limited) constraints. In [11], the corresponding expression was obtained for the case when there is only one linear constraint on the statistics of patterns, although the results can be extended to other classes of weak constraints. In general, having a capacity-achieving Markov chain is useful for code constructions, as demonstrated in [4] as well as in this paper. For completeness, we will present the expressions for the capacity and for a capacity-achieving Markov chain in Section II.

The main contribution of this paper is an explicit construction of capacity-achieving weakly constrained codes. The construction is based on the *row-by-row coding scheme* that was first presented in [6] and [16], and that was later adapted in [2] to design codes to mitigate inter-cell interference effects in flash memory. In the row-by-row coding scheme, messages are encoded and then written sequentially into the rows of a 2-dimensional (2D) array, such that the columns of the array are constrained sequences. This coding scheme admits an implementation that also allows one to control the total number of occurrences of some patterns along columns. In our construction, we show how to concatenate the columns of such an array into a longer codeword that can satisfy classical constraints that completely avoid certain forbidden patterns, as well as weak constraints that dictate the occurrence probabilities of specific allowed patterns.

The rest of the paper is organized as follows. In Section II we present notations and definitions for constrained systems, Markov chains, and weakly constrained systems. In Section III we review the row-by-row coding scheme. This coding scheme is then used to construct weakly constrained codes in Section IV.

II. Preliminaries

In this section we present the basic notations and definitions for constrained systems, Markov chains, and weakly constrained systems.

978-1-5090-3097-2/17/$31.00 ©2017 IEEE

For a positive integer n, denote by $[n]$ the set of n integers $\{1, 2, \ldots, n\}$. We use the notations $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_+$ to denote the set of non-negative real numbers and the set of positive real numbers, respectively.

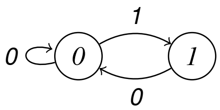
A. Labeled Graphs

Let Σ be an alphabet of size μ and denote by Σ^* the set of all sequences of finite length over Σ. For a word $\mathbf{w} = w_1 w_2 \ldots w_\ell \in \Sigma^*$ we denote by $|\mathbf{w}| \stackrel{\text{def}}{=} \ell$ the length of $\mathbf{w}$. A *labeled directed graph* $G(V, E, L)$ over Σ is a directed graph with a set of states V, an edge set $E \subseteq V \times V$, and an edge labeling function $L : E \to \Sigma$.

Let $G = G(V, E, L)$ be a labeled directed graph. For $e \in E$ we denote by $\sigma(e)$ and $\tau(e)$ the *initial state* and *terminal state* of e, respectively, i.e., $e = \sigma(e)\,\tau(e)$. A *path* γ in G of length $|\gamma| \stackrel{\text{def}}{=} \ell$ is a sequence of ℓ edges $e_1 e_2 \ldots e_\ell \in E$, such that for all $i \in [\ell - 1]$, $\sigma(e_{i+1}) = \tau(e_i)$. Let Γ be the set of all paths in G of finite length. The edge labeling function L can be extended to $L : \Gamma \to \Sigma^*$, where for all $\gamma \in \Gamma$, $L(\gamma)$ is the word obtained by reading the labels of the edges in γ. We call $L(\gamma)$ the *labeling* of the path γ.

The graph G is called *irreducible* if for every two distinct states $u, v \in V$ there exists a path $\gamma \in \Gamma$ connecting u to v. The graph G is primitive if there exists some integer $N_G > 0$ such that for every two states $u, v \in V$ there exists a path of length N_G that connects u to v. The graph G is called *lossless* if every two distinct paths in Γ with the same initial state and terminal state have different labelings.

Example 1. *Let $G = G(V, E, L)$ be the graph shown below.*



Then, G is lossless and primitive with $N_G = 2$.

B. Markov Chains

A *Markov chain* $\mathcal{P}$ on the labeled graph $G = G(V, E, L)$ is a probability mass function over the edge set of G. That is, $\mathcal{P} : E \to \mathbb{R}_{\geq 0}$ such that $\sum_{e \in E} \mathcal{P}(e) = 1$. For all $u \in V$, the *state probability mass function* for the Markov chain $\mathcal{P}$, $\pi : V \to \mathbb{R}_{\geq 0}$, is defined by

$$\pi(u) \stackrel{\text{def}}{=} \sum_{\substack{e \in E: \\ \sigma(e) = u}} \mathcal{P}(e).$$

A *Markov chain* $\mathcal{P}$ is called *stationary* if for all $u \in V$

$$\pi(u) = \sum_{\substack{e \in E: \\ \tau(e) = u}} \mathcal{P}(e).$$

Let Δ be the set of all stationary Markov chains on G. For $\gamma = e_1 e_2 \ldots e_\ell \in \Gamma$, define the *empirical Markov chain* of the path γ to be the Markov chain on G, $\mathcal{P}_\gamma : E \to \mathbb{R}_{\geq 0}$, such that for all $e \in E$,

$$\mathcal{P}_\gamma(e) \stackrel{\text{def}}{=} \frac{1}{\ell} |\{i \in [\ell] \ : \ e_i = e\}|.$$

The *entropy* of a Markov chain $\mathcal{P}$ is defined by

$$H(\mathcal{P}) \stackrel{\text{def}}{=} \sum_{e \in E} \mathcal{P} \log_2 \frac{\mathcal{P}(e)}{\pi(\sigma(e))}.$$

For a positive integer n, we say that $\mathcal{P}$ is *n-integral* if for every $e \in E$, $\mathcal{P}(e)n$ is an integer.

C. Constrained Systems

A *constrained system* $S(G)$ over the labeled graph $G = G(V, E, L)$ is the set of all words in Σ^* that are obtained by reading the labels of paths in Γ, i.e.,

$$S(G) \stackrel{\text{def}}{=} \{\mathbf{w} \in \Sigma^* \ : \ \exists \gamma \in \Gamma, \ \mathbf{w} = L(\gamma)\}.$$

We say that $S(G)$ is *represented* by G. Every constrained system can be represented by a lossless graph and therefore we will assume throughout this paper that G is lossless. A *constrained code* is simply a subset of $S(G)$. For every $B \subset \Sigma^*$, the base-2 capacity of B is defined by

$$cap(B) \stackrel{\text{def}}{=} \limsup_{\ell \to \infty} \frac{\log |B \cap \Sigma^\ell|}{\ell}.$$

There exists a lossless and primitive labeled graph F such that $S(F) \subseteq S(G)$ and $cap(S(F)) = cap(S(G))$. Since we are interested in this paper only in capacity achieving-constrained codes, we can assume w.l.o.g. that G is also primitive.

Let $A = (A_{u,v})$ be the adjacency matrix of G, i.e., A is a $|V| \times |V|$ matrix and $A_{u,v}$ is the number of edges that start at u and terminates at v. Let $\lambda_1, \lambda_2, \ldots, \lambda_{|V|} \in \mathbb{C}$ be all the $|V|$ eigenvalues of A (perhaps with repetitions). The *spectral radius* of A is defined by $\lambda \stackrel{\text{def}}{=} \max\{|\lambda_i| \ : \ i \in [|V|]\}$. By the Perron-Frobenius Theorem [5, Ch. 8], the spectral radius λ is an eigenvalue of A and it admits a positive right eigenvector as well as a positive left eigenvector (by a positive vector we mean a vector whose entries take positive real values). Let $\mathbf{y} = (y_1, y_2, \ldots, y_{|V|})$ and $\mathbf{x} = (x_1, x_2, \ldots, x_{|V|})^T$ be positive left and right eigenvectors of A, respectively, for the eigenvalue λ, normalized such that $\mathbf{yx} = 1$.

It is well known (see, for example, [14]) that the capacity of $S(G)$ is equal to $\log_2 \lambda$. Moreover,

$$cap(S(G)) = \max_{\mathcal{P} \in \Delta} H(\mathcal{P}),$$

and a capacity-achieving stationary Markov chain is given by

$$\widehat{\mathcal{P}}(e) \stackrel{\text{def}}{=} \frac{A_{\sigma(e),\tau(e)} y_{\sigma(e)} x_{\tau(e)}}{\lambda}.$$

Example 2. *The set $S \subset \Sigma^*$ of all words that do not contain two consecutive 1's is a constrained system and it is represented by the primitive and lossless graph G from Example 1. We call S the "no 11" constrained system. The adjacency matrix of G is the matrix*

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

with spectral radius $\lambda = \frac{1+\sqrt{5}}{2}$. Hence, $cap(S) = \log_2 \lambda \approx 0.694$.

D. Weakly-constrained Systems

For a function $\varphi : E \to \mathbb{R}^t$ and for a Markov chain $\mathcal{P}$ on G, denote the expected value of φ with respect to $\mathcal{P}$ by $\mathbb{E}_{\mathcal{P}}[\varphi]$, i.e.,

$$\mathbb{E}_{\mathcal{P}}[\varphi] \stackrel{\text{def}}{=} \sum_{e \in E} \mathcal{P}(e) \varphi(e).$$

For $\varphi : E \to \mathbb{R}^t$, $\mathbf{r} \in \mathbb{R}^t$, and $\epsilon : \mathbb{N} \to \mathbb{R}_{\geq 0}$, a $(\varphi, \mathbf{r}, \epsilon)$-*weakly constrained system* is the set

$$S_{\varphi,\mathbf{r},\epsilon}(G) \stackrel{\text{def}}{=} \left\{ \mathbf{w} \in S(G) \ : \ \begin{array}{c} \exists \gamma \in \Gamma,\ L(\gamma) = \mathbf{w}, \\ \text{and } \forall s \in [t], \\ |(\mathbb{E}_{\mathcal{P}_\gamma}[\varphi])_s - r_s| \leq \epsilon(|\gamma|) \end{array} \right\}.$$

That is, $S_{\varphi,\mathbf{r},\epsilon}(G)$ consists of the words in $S(G)$ that can be formed by paths in G with empirical Markov chains under which the expected value of φ is "close" to $\mathbf{r}$. The notion of closeness is formalized using the *tolerance function* ϵ that limits the difference between the expected value of φ and $\mathbf{r}$. Therefore, we are interested in functions ϵ, such that $\epsilon(n) = o(1)$. A subset of $S_{\varphi,\mathbf{r},\epsilon}(G)$ is called a $(\varphi, \mathbf{r}, \epsilon)$-*weakly constrained code*.

For $\mathbf{z} \in \mathbb{R}^t_+$, let $A(\mathbf{z})$ be the $|V| \times |V|$ matrix defined by

$$A(\mathbf{z})_{u,v} \stackrel{\text{def}}{=} \sum_{\substack{e: \\ \sigma(e)=u, \tau(e)=v}} \prod_{s=1}^{t} z_s^{\varphi(e)_s}.$$

As for the adjacency matrix A, the Perron-Frobenius Theorem states that the spectral radius of $A(\mathbf{z})$, $\lambda(\mathbf{z})$, is an eigenvalue of $A(\mathbf{z})$ with positive left and right eigenvectors $\mathbf{y}(\mathbf{z})$ and $\mathbf{x}(\mathbf{z})$, normalized such that $\mathbf{y}(\mathbf{z})\mathbf{x}(\mathbf{z}) = 1$.

The following theorem is an immediate consequence of Lemmas 2 and 5 from [13].

Theorem 1. *If $\epsilon(n) = o(1)$, then*

$$\begin{aligned} cap(S_{\varphi,\mathbf{r},\epsilon}(G)) &\leq \sup_{\mathcal{P} \in \Delta_{\varphi,\mathbf{r}}} H(\mathcal{P}) \\ &= \inf_{\mathbf{z} \in \mathbb{R}^t_+} \{-\sum_{s=1}^{t} r_s \log z_s + \log \lambda(\mathbf{z})\}, \end{aligned} \tag{1}$$

where $\Delta_{\varphi,\mathbf{r}}$ is the set of all stationary Markov chains on G, $\mathcal{P}$, for which $\mathbb{E}_{\mathcal{P}}[\varphi] = \mathbf{r}$.

As for constrained systems, one can define a Markov chain $\widehat{\mathcal{P}} \in \Delta_{\varphi,\mathbf{r}}$ for which

$$\sup_{\mathcal{P} \in \Delta_{\varphi,\mathbf{r}}} H(\mathcal{P}) = H(\widehat{\mathcal{P}}).$$

This Markov chain is defined by

$$\widehat{\mathcal{P}}_e \stackrel{\text{def}}{=} \prod_{s=1}^{t} \frac{z_s^{\varphi(e)_s} y(\mathbf{z})_{\sigma(e)} x(\mathbf{z})_{\tau(e)}}{\lambda(\mathbf{z})}, \tag{2}$$

where $\mathbf{z} \in \mathbb{R}^t_+$ is a solution to

$$\mathbb{E}_{\widehat{\mathcal{P}}}(\varphi) = \mathbf{r}. \tag{3}$$

Example 3. *Let $S = S(G)$ be the "no 11" constrained system, where G is the graph from Example 1 and let $\widehat{S}$ be the set of words $\mathbf{w} \in S$ such that $\mathbf{w}$ has exactly $0.25|\mathbf{w}|$ ones. Then*

$$\widehat{S} = S_{\varphi,0.25}(G),$$

where $\varphi : E \to \mathbb{R}$ is defined by $\varphi(0\,1) = 1$ and $\varphi(0\,0) = \varphi(1\,0) = 0$, $\mathbf{r} = r = 0.25$, and ϵ is just the zero function and thus omitted from the notation.

The matrix $A(z)$ is given by

$$A(z) \stackrel{\text{def}}{=} \begin{pmatrix} 1 & z \\ 1 & 0 \end{pmatrix},$$

the spectral radius is

$$\lambda(z) = \frac{1 + \sqrt{1 + 4z}}{2},$$

and

$$\mathbf{y}(z) = (1, \lambda(z) - 1) \ \ \textit{and} \ \ \mathbf{x}(z) = \frac{1}{2\lambda(z) - 1} \begin{pmatrix} \lambda(z) \\ 1 \end{pmatrix}$$

are positive left and right eigenvectors of $A(z)$ corresponding to $\lambda(z)$ and normalized such that $\mathbf{y}(z)\mathbf{x}(z) = 1$. A capacity-achieving Markov chain is defined by

$$\widehat{\mathcal{P}}(e) = \frac{z^{\varphi(e)} y(z)_{\sigma(e)} x(z)_{\tau(e)}}{\lambda(z)},$$

where $z = 0.75$ is a positive solution to $\mathbb{E}_{\widehat{\mathcal{P}}}[\varphi] = \widehat{\mathcal{P}}(0\,1) = 0.25$. Hence $\lambda(z) = 1.5$, $\widehat{\mathcal{P}}(0\,1) = \widehat{\mathcal{P}}(1\,0) = 0.25$, $\widehat{\mathcal{P}}(0\,0) = 0.5$, and $cap(\widehat{S}) \leq \log 1.5 - 0.25 \log 0.75 \approx 0.688$.

Remark 1. *Clearly, if $\epsilon(n) = o(1)$ and the rate of convergence of $\epsilon(n)$ to zero is high enough then $cap(S_{\varphi,\mathbf{r},\epsilon})$ might be zero. In particular, if $\epsilon(n) = o(1/n)$ then for large enough n most choices of φ and $\mathbf{r}$ will result in zero capacity. Using a slightly different definition of weakly constrained codes the authors of [4] showed that when $\epsilon(n) = \Omega(1/n)$ the bound on capacity in (1) is tight. The result should hold for our definition of weakly constrained codes as well. The main result of our paper is a construction of a $(\varphi, \mathbf{r}, \epsilon)$-weakly constrained code of length n, where $\epsilon(n) = O(1/n^{1-\beta})$ (β can be arbitrarily close to 0) for which the capacity satisfies (1) with equality.*

III. Row-by-row constrained coding

In this section we review the row-by-row constrained coding technique presented in [6], [16] that will be useful for our code construction presented in Section IV.

For $\mu \geq 2$ and $k \geq 1$, the k-dimensional De Bruijn graph of μ symbols, $D_{k,\mu}$, is the labeled directed graph over an alphabet Σ of size μ, whose vertex set $V_{k,\mu}$ is the set Σ^k and whose edge set, $E_{k,\mu}$, is the set $\{\mathbf{u}\,\mathbf{v} \in V_{k,\mu} \times V_{k,\mu} \ : \ u_{i+1} = v_i, \text{ for all } i \in [k-1]\}$. We represent an edge $e = \mathbf{u}\,\mathbf{v} \in E_{k,\mu}$ by the vector $\mathbf{e} = e_1\, e_2\, \ldots\, e_{k+1} \in \Sigma^{k+1}$, where $e_i = u_i$, for all $i \in [k]$, and $e_{k+1} = v_k$. The edge labeling function $L_{k,\mu} : E_{k,\mu} \to \Sigma$ of $D_{k,\mu}$ is defined by $L_{k,\mu}(\mathbf{e}) \stackrel{\text{def}}{=} e_{k+1}$.

Throughout this section we assume that $G = G(V, E, L)$ is an irreducible subgraph of $D_{k,\mu}$, for some k and μ. Notice that this assumption holds without the loss of generality since G is a subgraph of $D_{1,|V|}$. Let n be a positive integer and let $\mathcal{P} : E \to \mathbb{R}_{\geq 0}$ be an n-integral stationary Markov chain on G with state probability mass function π. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{|V|}$ be the lexicographic order of the states. Since $\mathcal{P}$ is n-integral, it follows that for all $\ell \in [|V|]$, $n_\ell \stackrel{\text{def}}{=} \pi(\mathbf{v}_\ell)n$ is an integer. Let U_π be the $k \times n$ matrix over Σ such that for all $\ell \in [|V|]$, all the columns of U_π with column index $\sum_{s=0}^{\ell-1} n_s + 1 \leq j \leq \sum_{s=1}^{\ell} n_s$, where $n_0 \stackrel{\text{def}}{=} 0$, are equal to the vector $\mathbf{v}_\ell$.

A *row-by-row weakly constrained coding scheme* over G, with a Markov chain $\mathcal{P}$, encodes a sequence of m messages $M_1, M_2, \ldots, M_m \in [M]$ into the rows of some $(m + k) \times n$ array $W = (W_{i,j})$, with row index $-k + 1 \leq i \leq m$, column index $j \in [n]$, and with rows

$W_{-k+1}, \ldots, W_{-1}, W_0, W_1, W_2, \ldots, W_m$, such that the following conditions are satisfied.

1) For every $i \in [m]$ the message M_i is encoded to a (unique) length-n vector over Σ which is stored in W_i.
2) For every $i_1, i_2 \in [m]$, if $i_1 < i_2$ then W_{i_1} is programmed before W_{i_2}.
3) For every $\mathbf{e} \in E$ and for every $i \in [m]$, $\mathbf{e}$ appears as a column of the rows $W_{i-k}, W_{i-k+1}, \ldots, W_i$ exactly $\mathcal{P}(\mathbf{e})n$ times.

To implement a row-by-row weakly constrained coding scheme one can set the first k rows of W, $W_{-k+1}, W_{-k+2}, \ldots, W_0$ to be the rows of U_π and encode the input messages using a constant-weight code of the form

$$\mathcal{C}_{rbr} \stackrel{\text{def}}{=} \mathbb{C}_1 \times \mathbb{C}_2 \times \cdots \times \mathbb{C}_{|V|},$$

where

$$\mathbb{C}_\ell \stackrel{\text{def}}{=} \left\{ \mathbf{w} \in \Sigma^{n_\ell} \ : \ \begin{array}{c} \text{for all } \alpha \in \Sigma, \ \alpha \text{ appears} \\ \text{in } \mathbf{w} \ \mathcal{P}(\mathbf{v}_\ell \, \alpha)n \text{ times} \end{array} \right\}.$$

If M_i is encoded to $\mathbf{c}_1 \, \mathbf{c}_2 \, \ldots \, \mathbf{c}_{|V|}$, where $\mathbf{c}_\ell \in \mathbb{C}_\ell$, for all $\ell \in [|V|]$, then the codeword $\mathbf{c}_\ell$ is stored in the row W_i in the n_ℓ positions j for which $W_{i-k,j} W_{i-k+1,j}, \ldots, W_{i-1,j} = \mathbf{v}_\ell$.

Remark 2. *From the stationarity of $\mathcal{P}$, the definition of the codes $\mathbb{C}_\ell$, and the way the codewords are stored it is guaranteed that the number of positions j for which $W_{i-k,j} W_{i-k+1,j}, \ldots, W_{i-1,j} = \mathbf{v}_\ell$ is indeed $n_\ell (= \pi(\mathbf{v}_\ell)n)$. For more details on the implementation of a row-by-row coding scheme, see [16].*

It can be readily verified that this implementation indeed satisfies the properties of a row-by-row weakly constrained coding scheme. By definition, the asymptotic coding rate of any row-by-row weakly constrained coding scheme over G, with the Markov chain $\mathcal{P}$, cannot exceed $H(\mathcal{P})$. Since

$$\lim_{n \to \infty} \frac{1}{n} \log_2 |\mathcal{C}| = H(\mathcal{P}),$$

it follows that the above row-by-row weakly constrained coding scheme is also capacity-achieving.

Example 4. *Let G be the graph from Example 1 that represents the "no 11" constrained system. Then G is a primitive subgraph of $D_{1,2}$. Let $\mathcal{P}$ be the Markov chain on G from Example 3, i.e., $\mathcal{P}(0\,1) = \mathcal{P}(1\,0) = 0.25$ and $\mathcal{P}(0\,0) = 0.5$, and let $n = 8$. The state probability mass function, π, associated with $\mathcal{P}$ is given by $\pi(0) = 0.75$, $\pi(1) = 0.25$, and therefore $n_1 = 6$ and $n_2 = 2$. The matrix U_π is a 1×8 matrix and its single row is equal to $0\,0\,0\,0\,0\,0\,1\,1$. The code $\mathcal{C}_{rbr}$ that is used to encode the m messages is defined by $\mathcal{C}_{rbr} = \mathbb{C}_1 \times \mathbb{C}_2$, where*

$$\mathbb{C}_1 = \left\{ \mathbf{x} \in \{0,1\}^6 \ : \ \begin{array}{c} 1 \text{ appears in } \mathbf{x} \\ 2 \text{ times} \end{array} \right\} \text{ and } \mathbb{C}_2 = \{0\,0\}.$$

The 2D array W that is obtained by programming the codewords $\mathbf{c}, \widehat{\mathbf{c}} \in \mathcal{C}_{rbr}$, where $\mathbf{c} = 1\,0\,0\,0\,0\,1\,0\,0$ and $\widehat{\mathbf{c}} = 0\,0\,0\,0\,1\,1\,0\,0$ is the following.

0	0	0	0	0	0	1	1
1	0	0	0	0	1	0	0
0	0	0	0	0	0	1	1

The first row of W is just the single row of U_π. The first 6 entries of $\mathbf{c}$ are stored in W_1 below the zeros of W_0 and the last 2 (zero) entries of $\mathbf{c}$ are stored in W_1 below the ones of W_0. Similarly, the first 6 entries of $\widehat{\mathbf{c}}$ are stored in W_2 below the zeros of W_1 and the last 2 (zero) entries of $\widehat{\mathbf{c}}$ are stored in W_2 below the ones of W_1. Notice that every column satisfies the "no 11" constraint and that for every $\mathbf{e} \in E \subset \{0,1\}^2$, $\mathbf{e}$ appears as a column of any two consecutive rows exactly $\mathcal{P}(\mathbf{e})n$ times.

IV. Capacity-Achieving Codes

The goal of this section is to present an explicit construction of capacity achieving weakly constrained codes. To this end we assume that G is some primitive subgraph of $D_{k,\mu}$ and that $\mathcal{P} : E \to \mathbb{R}_{\geq 0}$ is an n-integral stationary Markov chain over G, for some positive integer n, and use the row-by-row weakly constrained coding to produce length-$N \approx n^2$ codewords in $S(G)$, in which every pattern $\mathbf{e} \in E$ appears exactly $\mathcal{P}(\mathbf{e})(N-k)$ times. We show that as n approaches infinity, the coding rate approaches $H(\mathcal{P})$. We then show how to apply this method to obtain a capacity-achieving $(\varphi, \mathbf{r}, \epsilon)$-weakly constrained code, where $\epsilon(N) = O(1/N^{1-\beta})$, where β can be arbitrarily small.

Let $W = (W_{i,j})$, $-k+1 \leq i \leq m$ and $j \in [n]$, be the $(m+k) \times n$ 2D array that is obtained by encoding some m messages from $[|\mathcal{C}_{rbr}|]$ using the row-by-row weakly constrained coding scheme over G, with the Markov chain $\mathcal{P}$. Recall that we denote the rows of W by $W_{-k+1}, W_{-k+2}, \ldots, W_0, W_1, \ldots, W_m$. We denote the columns of W by $B_1, B_2, \ldots, B_n$. Since all the columns of W belongs to $S(G)$ and since every $\mathbf{e} \in E$ appears vertically in W exactly $\mathcal{P}(\mathbf{e})mn$ times, it is tempting to connect the columns into one long codeword. This is the key idea behind our code construction, but it requires some additional steps in order to guarantee that the codeword belongs to $S(G)$ and that its empirical distribution over E matches the Markov chain $\mathcal{P}$.

For two vectors $\mathbf{w}_1, \mathbf{w}_2 \in \Sigma^*$ of lengths $m_1, m_2 \geq k$, we say that $\mathbf{w}_1$ is *extendable* by $\mathbf{w}_2$ if the last k entries of $\mathbf{w}_1$ are equal to the first k entries of $\mathbf{w}_2$. In this case, the result of extending $\mathbf{w}_1$ with $\mathbf{w}_2$ is the length $m_1 + m_2 - k$ vector over Σ,

$$\mathbf{w}_1 \| \mathbf{w}_2 \stackrel{\text{def}}{=} \mathbf{w}_1 \, w_{2,k+1} \, w_{2,k+2} \, \ldots \, w_{m_2}.$$

The following lemma can be readily verified.

Lemma 1. *If $\mathbf{w}_1, \mathbf{w}_2 \in S(G)$ and $\mathbf{w}_1$ is extendable by $\mathbf{w}_2$ then $\mathbf{w}_1 \| \mathbf{w}_2 \in S(G)$. Moreover, if for some $\mathbf{e} \in E$, the number of times the pattern $\mathbf{e}$ appears in $\mathbf{w}_1$ and $\mathbf{w}_2$ is t_1 and t_2, respectively, then the number of times $\mathbf{e}$ appears in $\mathbf{w}_1 \| \mathbf{w}_2$ is $t_1 + t_2$.*

From Lemma 1 it follows that if there exists some permutation $\rho : [n] \to [n]$ such that for all $j \in [n-1]$, $B_{\rho(j)}$ is extendable by $B_{\rho(j+1)}$, then

$$\mathbf{c} = B_{\rho(1)} \| B_{\rho(2)} \| \cdots \| B_{\rho(n)}$$

is a codeword of length $N = mn + k$ in $S(G)$ such that for all $\mathbf{e} \in E$, $\mathbf{e}$ appears in $\mathbf{c}$ exactly $\mathcal{P}(\mathbf{e})mn = \mathcal{P}(\mathbf{e})(N-k)$ times. In that case, W is called *good*. Hence, the first step towards our code construction is to show that by appending some extra rows to W using the row-by-row coding technique, we can obtain a good $(m'+k) \times n$ 2D array W'.

Lemma 2. *Suppose that for all $\ell \in [|V|]$ and for all $\mathbf{e} \in E$ with $\sigma(e) = \mathbf{v}_\ell$ we have $\mathcal{P}(\mathbf{e})n \geq |V|$. Then, using the row-by-row weakly constrained coding scheme, one can append some N_G rows to W to obtain a good $(m + N_G + k) \times n$ 2D array W'.*

Due to space limitations, we omit the proof. The key idea is that finding the permutation ρ is equivalent to finding an Eulerian path in a graph with $|V|$ vertices and n edges, where the edges are determined by the first and last k entries of each column in the array W'. This graph already has the property that the in-degree of each vertex matches its out-degree. Therefore, one only needs to ensure the graph is irreducible,which is easily achieved for n large enough.

Although Lemmas 1 and 2 imply that we can encode a sequence of m messages in $\mathbb{Z}_{|\mathcal{C}_{rbr}|}$ to a codeword $\mathbf{c} \in S(G)$ of length $N = m'n + k$ for which every pattern $\mathbf{e} \in E$ appears exactly $\mathcal{P}(\mathbf{e})(N - k)$ times, we still need to make sure that $\mathbf{c}$ is decodable, i.e., that we can retrieve the m messages from $\mathbf{c}$. If we knew how to recreate the $(m'+k)\times n$ 2D array W' from which $\mathbf{c}$ was obtained, we could use the row-by-row decoding algorithm to retrieve the original messages. However, without knowing the permutation ρ according to which the columns of W' were linked to one another to create $\mathbf{c}$, it is not clear how we can obtain W'. To overcome this problem we again suggest to append rows to W'. Recall that the first k rows of W' are equal to the rows of the matrix U_π, defined in Section III. We would like to append rows to the bottom of the matrix U_π, using the row-by-row weakly constrained coding technique, to create an $\hat{m} \times n$ matrix $\widehat{U}_\pi$ in which all columns are distinct. The 2D array that we will use will then consist of $\hat{m}$ rows that are equal to the rows of $\widehat{U}_\pi$, followed by m rows that are the result of encoding m messages using the row-by-row weakly constrained coding, and finally N_G rows that guarantee that the final 2D array is good.

Lemma 3. *An $\hat{m} \times n$ matrix $\widehat{U}_\pi$ with distinct columns can be constructed with $\hat{m} = O(\log n)$ rows.*

Due to space limitations, the proof is omitted. We summarize with the following corollary.

Corollary 1. *If for all $\mathbf{e} \in E$, $\mathcal{P}(e)n \geq |V|$, then there exists a constant c that depends only on G and the Markov chain $\mathcal{P}$ such that one can encode m messages from $\mathcal{C}_{rbr}$ to a length-$N = (c \log n + m + N_G)n + k$ codeword $\mathbf{c} \in S(G)$ in which every pattern $\mathbf{e} \in E$ appears exactly $\mathcal{P}(\mathbf{e})(N - k)$ times.*

Example 5. *Let $n = 8$, let G be the graph from Example 1, and let $\mathcal{P}$ be the Markov chain with $\mathcal{P}(01) = \mathcal{P}(10) = 0.25$ and $\mathcal{P}(00) = 0.5$. We will show how to create a 9×8 2D array, constructed using the row-by-row weakly constrained coding and Lemmas 2 and 3, which we will then assemble into a length-65 codeword. The first five rows of the 2D array are the rows of the matrix $\widehat{U}_\pi$. The next $m = 2$ rows are information rows. Assume we wish to write the codewords $\mathbf{c}, \widehat{\mathbf{c}} \in \mathcal{C}_{rbr}$, where $\mathbf{c} = 10010000$ and $\widehat{\mathbf{c}} = 00001100$. Then after writing these two codewords and adding $N_G = 2$ rows we get the following good 2D array W:*

0	0	0	0	0	0	1	1
0	0	0	0	1	1	0	0
0	0	1	1	0	0	0	0
0	0	0	0	0	1	0	1
0	1	0	1	0	0	0	0
1	0	0	0	0	1	0	0
0	0	0	0	0	0	1	1
0	0	0	0	1	1	0	0
0	0	0	1	0	0	0	1

Then the codeword

$$\mathbf{c} = B_1 \| B_2 \| B_3 \| B_4 \| B_8 \| B_7 \| B_5 \| B_6$$

is of length 65 and it has the property that each of the patterns 0 1 and 1 0 appears in it exactly $16 = \mathcal{P}(0\,1)64$ times, and the pattern 0 0 appears in it exactly $32 = \mathcal{P}(0\,0)64$ times.

Finally, using a technique suggested in [16], given the capacity-achieving Markov chain $\widehat{\mathcal{P}}$ for a $(\varphi, \mathbf{r}, \epsilon)$-weakly constrained system, one can obtain an n-integral Markov chain $\mathcal{P}$ such that $|\widehat{\mathcal{P}}(\mathbf{e}) - \mathcal{P}(\mathbf{e})| = O(1/n)$. As a result the entropy of $\mathcal{P}$ goes to the entropy of $\widehat{\mathcal{P}}$, as n goes to infinity. Applying our construction with the n-integral Markov chain $\mathcal{P}$ setting $m = \Theta(n^\delta)$ ($\delta > 0$ should be much larger than $\log_2 \log_2 n / \log_2 n$), one can construct a capacity-achieving $(\varphi, \mathbf{r}, \epsilon)$-weakly constrained code of length $N \approx n^{1+\delta}$ with $\epsilon(N) = O(1/n) = O(1/N^{(1+\delta)^{-1}})$.

V. Acknowledgments

This work was supported by the ISEF Foundation and by NSF Grant CCF-1619053.

References

[1] R. L. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes," *IEEE Trans. on Inform. Theory*, vol. 29, no. 1, pp. 5–22, Jan. 1983.

[2] S. Buzaglo and P. H. Siegel, "Row-by-row coding schemes for inter-cell interference in Flash memory," to appear in *IEEE Trans. Commun.*.

[3] N. G. de Bruijn, "A combinatorial problem," *Koninklijke Nederlandse Akademie v. Wetenschappen*, vol. 49, pp. 758-764, 1946.

[4] O. Elishco, T. Meyerovitch, and M. Schwartz, "Semiconstrained systems," *IEEE Trans. on Inform. Theory,* vol. 62, no. 4, pp. 1688–1702, Apr. 2016.

[5] F. R. Gantmacher *Matrix Theory, Volume II*, Chelsea Publishing Company, New York, 1960.

[6] S. Halevy and R. M. Roth, "Parallel constrained coding with application to two-dimensional constraints," *IEEE Trans. on Inform. Theory*, vol. 48, no. 5, pp. 1009–1020, Aug. 2002.

[7] K. A. S. Immink, "Weakly constrained codes," *Electron. Lett.*, vol. 33, pp. 1943-1944, Nov. 1997.

[8] K. A. S. Immink, "A practical method for approaching the channel capacity of constrained channels," *IEEE Trans. Inform. Theory,* vol. 43, no. 5, pp. 1389-1399, Sep. 1997.

[9] A. Jagmohan, M. Franceschini, L. A. Lastras-Montao and J. Karidis, "Adaptive endurance coding for NAND Flash," *IEEE Globecom Workshops,* pp. 1841–1845, Miami, FL, 2010.

[10] A. J. E. M. Janssen and K. A. S. Immink, "An entropy theorem for computing the capacity of weakly (d, k)-constrained sequences," *IEEE Trans. Inform. Theory,* vol. 46, no. 3, pp. 1034-1038, May 2000.

[11] J. Justesen and T. Høholdt, "Maxentropic Markov chains," *IEEE Trans. Inform. Theory,* vol. 30, no. 4, pp. 665-667, July 1984.

[12] Y. Liu and PH. Siegel, "Shaping codes for structured data," *IEEE GLOBECOM,* Washington, DC, December 2016.

[13] B. H. Marcus and R. M. Roth, "Improved Gilbert-Varshamov bound for constrained systems," *IEEE Trans. Inform. Theory,* vol. 38, pp. 1213-1221, July 1992.

[14] B. H. Marcus, R. M. Roth, and P. H. Siegel, *Constrained systems and coding for recording channels,* in Handbook of Coding Theory, V. Pless and W. Huffman, Eds. Amsterdam: Elsevier, pp. 1635–1764, 1998.

[15] B. Moison, unpublished notes, 2000.

[16] I. Tal, T. Etzion, and R. M. Roth, "On row-by-row coding for 2-D constraints," *IEEE Trans. on Inform. Theory*, vol. 55, pp. 3565–3576, 2009.