# The Capacity of Count-Constrained ICI-Free Systems


Navin Kashyap
Dept. of Electrical Communication Eng.
Indian Institute of Science, Bangalore
Email: nkashyap@iisc.ac.in

Ron M. Roth
Dept. of Computer Science
Technion, Haifa, Israel
Email: ronny@cs.technion.ac.il

Paul H. Siegel
Dept. of Electrical & Computer Eng.
Univ. of California San Diego, USA
Email: psiegel@ucsd.edu



***Abstract*—A Markov chain approach is applied to determine the capacity of a general class of $q$-ary ICI-free constrained systems that satisfy an arbitrary count constraint.**


## I. Introduction

Let $\Sigma$ be an alphabet of a finite size $q \geq 2$. A *word* over $\Sigma$ is any finite string $\boldsymbol{w} = w_1 w_2 \ldots w_n$ where $w_i \in \Sigma$. Let $\mathcal{F}$ be a finite set of words over $\Sigma$. The *(finite-type) constrained system* $\mathsf{S}_{\mathcal{F}}$ consists of all words $\boldsymbol{w} = w_1 w_2 \ldots w_n$ over $\Sigma$ such that $\mathcal{F}$ contains none of their substrings $w_i w_{i+1} \ldots w_j$, for any $1 \leq i \leq j \leq n$. We refer to the set $\mathcal{F}$ as the set of *forbidden words* defining the constrained system $\mathsf{S}_{\mathcal{F}}$. The constrained system $\mathsf{S}_{\mathcal{F}}$ can be presented by a (finite) directed edge-labeled graph $G$, with edges labeled with symbols from $\Sigma$, such that $\mathsf{S}_{\mathcal{F}}$ is the set of all words obtained by reading off the labels along paths of $G$. For a proof of this fact, we refer the reader to [5], which provides a comprehensive introduction to the subject of constrained systems.

Our specific interest is in a general class of "inter-cell interference free" (in short, "ICI-free") constrained systems, which we now define. For prescribed positive integers $a$, $b$, and $q$ such that $a+b \leq q$, let $\Sigma$ be an alphabet of size $q$ which is assumed to be partitioned into three (disjoint) subsets $L$, $H$, and $I$, of sizes $a$, $b$, and $q-a-b$, respectively. The elements in $L$ (respectively, $H$) represent the "low" (respectively, "high") symbols of $\Sigma$, while those in $I$ are the "intermediate" symbols. The ICI-free constrained system that we consider is the constrained system[1] $\mathsf{S}_{q;a,b} := \mathsf{S}_{\mathcal{F}_{q;a,b}}$ defined by the set of forbidden words $\mathcal{F}_{q;a,b} := \{w_1 w_2 w_3 : w_1, w_3 \in H, w_2 \in L\}$. A graph $G_{q;a,b}$ presenting the constrained system $\mathsf{S}_{q;a,b}$ is shown in Fig. 1.

We additionally impose a count constraint defined by a given probability vector $\boldsymbol{p} = (p_s)_{s \in \Sigma}$ (with nonzero entries that sum to 1), which specifies the frequencies of occurrence of each $s \in \Sigma$ within words belonging to $\mathsf{S}_{q;a,b}$. To avoid trivialities, we will assume $\rho_L := \sum_{s \in L} p_s$ and $\rho_H := \sum_{s \in H} p_s$ to be strictly positive. The probability $\rho_I := \sum_{s \in I} p_s$ is allowed to be 0.

For $\varepsilon > 0$, let $\mathsf{S}_{q;a,b}(\boldsymbol{p}, \varepsilon)$ denote the subset of $\mathsf{S}_{q;a,b}$ consisting of all words $\boldsymbol{w} \in \mathsf{S}_{q;a,b}$ in which the number of occurrences of each symbol $s \in \Sigma$ lies in the interval $\big((p_s - \varepsilon)|\boldsymbol{w}|, (p_s + \varepsilon)|\boldsymbol{w}|\big)$, where $|\boldsymbol{w}|$ denotes the length of $\boldsymbol{w}$. The *capacity* (or the *asymptotic information rate*) of $\mathsf{S}_{q;a,b}$ under the count constraint specified by $\boldsymbol{p}$ is defined as[2]

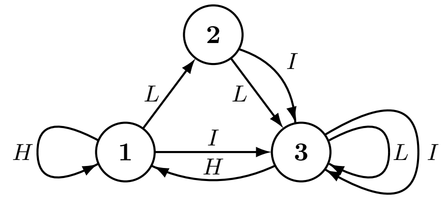


Fig. 1. The graph $G_{q;a,b}$ presenting the $q$-ary ICI-free constraint $\mathsf{S}_{q;a,b}$. Each arrowed line labeled by $X \in \{L, I, H\}$ represents $|X|$ parallel edges labeled by distinct symbols from $X$.

$$\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p}) := \lim_{\varepsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log |\mathsf{S}_{q;a,b}(\boldsymbol{p}, \varepsilon) \cap \Sigma^n| \ . \quad (1)$$

This quantifies, for large $n$, the exponential rate of growth of the number of length-$n$ words in $\mathsf{S}_{q;a,b}$ in which the relative frequency of occurrence of each symbol $s \in \Sigma$ is approximately $p_s$. Dropping the count constraint, we also define the (ordinary) capacity of the constrained system $\mathsf{S}_{q;a,b}$ to be[3]

$$\mathsf{cap}(\mathsf{S}_{q;a,b}) := \lim_{n \to \infty} \frac{1}{n} \log |\mathsf{S}_{q;a,b} \cap \Sigma^n| \ . \quad (2)$$

The quantities $\mathsf{cap}(\mathsf{S}_{q;a,b})$ and $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ were studied in [1], [6], [7], motivated by proposed coding schemes to mitigate inter-cell interference in flash memory devices.[4] Using standard techniques from the theory of constrained systems (see e.g., [5]) the (ordinary) capacity $\mathsf{cap}(\mathsf{S}_{q;a,b})$ was shown in [1] to be the largest real root of the cubic polynomial $x^3 - qx^2 + abx - ab(q-b)$. The analysis of $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ in [1] is based on combinatorial arguments, and a Stirling approximation of the resulting expressions then yields a bivariate function which needs to be maximized (numerically) in order to obtain the values of $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$.

In this work, we make use of a result from [4] to formulate the problem of determining the capacity $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ as an

[1]Since only the sizes of $\Sigma$, $L$, and $H$ will matter, we identify the constrained system by the sizes of these sets.

[2]All logarithms in this work are to the base 2.

[3]By a standard sub-additivity argument, the limit in this definition exists.

[4]These references used a different definition of $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$, which is shown in Appendix A in [2] to be equivalent to our definition in (1).

optimization problem over Markov chains defined on the graph $G_{q;a,b}$ shown in Fig. 1. By shifting to the dual optimization problem, we then derive an analytical solution to this optimization problem, which results in an exact expression for $\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p})$ given in Theorems 3 and 4 in Section III. While our analysis is tailored to count-constrained ICI-free systems, some of the tools that we use may be applicable to other constrained systems as well (see [3]).

## II. Markov Chains and Optimization

Let $G=(V,E)$ be a directed graph with vertex set $V$ and (directed) edge set $E$. For a vertex $v\in V$, we let $E_{\rm in}(v)$ and $E_{\rm out}(v)$ denote the set of incoming and outgoing edges, respectively, incident with $v$.

A stationary Markov chain on $G$ is a probability distribution $P=\big(P(e)\big)_{e\in E}$ on $E$, with the property that for each $v\in V$, the sum of the probabilities on the incoming edges of $v$ is equal to that on the outgoing edges of $v$:

$$\sum_{e\in E_{\rm in}(v)} P(e)=\sum_{e\in E_{\rm out}(v)} P(e)\ . \tag{3}$$

The induced stationary distribution on the vertex set $V$ is given by $\pi(v)=\sum_{e\in E_{\rm out}(v)}P(e)$, for all $v\in V$. The set of all stationary Markov chains on $G$ is denoted by $\Delta(G)$.

The *entropy rate* of a stationary Markov chain $P$ on $G$ is defined as

$$\mathsf{H}(P):=-\sum_{e\in E}P(e)\log P(e)-\left(-\sum_{v\in V}\pi(v)\log\pi(v)\right).$$

Since $\mathsf{H}(P)=-\sum_{v\in V}\sum_{e\in E_{\rm out}(v)}P(e)\log(P(e)/\pi(v))$, the convexity properties of relative entropy imply that $P\mapsto\mathsf{H}(P)$ is a concave function.

Given a Markov chain $P\in\Delta(G)$, along with a vector of real-valued functions $\boldsymbol{f}=(f_1\ f_2\ \ldots\ f_t):E\to\mathbb{R}^t$, we denote by $\mathbb{E}_P(\boldsymbol{f})$ the expected value of $\boldsymbol{f}$ with respect to $P$:

$$\mathbb{E}_P(\boldsymbol{f})=\sum_{e\in E}P(e)\boldsymbol{f}(e)\ .$$

We will only need the following special case of the function $\boldsymbol{f}$. Let $\mathsf{L}:E\to\Sigma$ be a labeling of the edges of the graph $G$ with symbols from $\Sigma$. For a subset $W$ of $\Sigma$ of size $t$, we define the vector indicator function $\mathcal{I}_W:E\to\mathbb{R}^t$ by $\mathcal{I}_W=(\mathcal{I}_s)_{s\in W}$, where $\mathcal{I}_s:E\to\mathbb{R}$ is the indicator function for a symbol $s\in\Sigma$:

$$\mathcal{I}_s(e)=\begin{cases}1 & \text{if } \mathsf{L}(e)=s\\ 0 & \text{otherwise}\end{cases}.$$

Then, $\mathbb{E}_P(\mathcal{I}_W)$ is a vector in $\mathbb{R}^t$ whose entry that is indexed by $s\in W$ is the probability that an edge chosen according to the distribution $P$ is labeled with the symbol $s$.

These definitions allow us to state the following result, which expresses $\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p})$ as the solution to a convex optimization problem.

**Proposition 1.** *We have*

$$\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p})=\sup_{\substack{P\in\Delta(G_{q;a,b}):\\ \mathbb{E}_P(\mathcal{I}_W)=\boldsymbol{p}'}}\mathsf{H}(P)\ ,$$

*for any $W\subset\Sigma$ of size $q-1$ and $\boldsymbol{p}'=(p_s)_{s\in W}$.*

*Proof.* As a consequence of [4, Lemma 2], for any $\varepsilon>0$, $\limsup_{n\to\infty}(1/n)\log|\mathsf{S}_{q;a,b}(\boldsymbol{p},\varepsilon)\cap\Sigma^n|$ is equal to $\sup\mathsf{H}(P)$, the supremum being over stationary Markov chains $P\in\Delta(G_{q;a,b})$ such that $\mathbb{E}_P(\mathcal{I}_\Sigma)\in(\boldsymbol{p}-\varepsilon\cdot\mathbf{1},\boldsymbol{p}+\varepsilon\cdot\mathbf{1})$ (where $\mathbf{1}$ denotes the all-one vector in $\mathbb{R}^q$). We claim that as $\varepsilon\to0$, these suprema converge to $\sup\mathsf{H}(P)$, the supremum now being over stationary Markov chains $P\in\Delta(G_{q;a,b})$ such that $\mathbb{E}_P(\mathcal{I}_\Sigma)=\boldsymbol{p}$. With this, we would have

$$\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p})=\sup_{\substack{P\in\Delta(G_{q;a,b}):\\ \mathbb{E}_P(\mathcal{I}_\Sigma)=\boldsymbol{p}}}\mathsf{H}(P)\ . \tag{4}$$

The constraint $\mathbb{E}_P(\mathcal{I}_\Sigma)=\boldsymbol{p}$ in the supremum on the right-hand side (RHS) above can be replaced by $\mathbb{E}_P(\mathcal{I}_W)=(p_s)_{s\in W}$, since the latter implies $\mathbb{E}_P(\mathcal{I}_{\{s\}})=p_s$ for the remaining symbol $s\in\Sigma\setminus W$. This would prove the proposition.

We now prove the claim above. To this end, for $\varepsilon>0$, define $\Delta_{\boldsymbol{p},\varepsilon}$ to be the set of all stationary Markov chains $P\in\Delta(G_{q;a,b})$ such that $\mathbb{E}_P(\mathcal{I}_\Sigma)\in(\boldsymbol{p}-\varepsilon\cdot\mathbf{1},\boldsymbol{p}+\varepsilon\cdot\mathbf{1})$. Its closure $\overline{\Delta_{\boldsymbol{p},\varepsilon}}$ is the set of all $P\in\Delta(G_{q;a,b})$ such that $\mathbb{E}_P(\mathcal{I}_\Sigma)\in[\boldsymbol{p}-\varepsilon\cdot\mathbf{1},\boldsymbol{p}+\varepsilon\cdot\mathbf{1}]$. By continuity of the mapping $P\mapsto\mathsf{H}(P)$, we have

$$\sup_{P\in\Delta_{\boldsymbol{p},\varepsilon}}\mathsf{H}(P)=\sup_{P\in\overline{\Delta_{\boldsymbol{p},\varepsilon}}}\mathsf{H}(P),$$

and the latter supremum is in fact a maximum. Finally, let $\Delta_{\boldsymbol{p},0}$ denote the set of all $P\in\Delta(G_{q;a,b})$ such that $\mathbb{E}_P(\mathcal{I}_\Sigma)=\boldsymbol{p}$. We wish to show that

$$\lim_{\varepsilon\to0^+}\sup_{P\in\overline{\Delta_{\boldsymbol{p},\varepsilon}}}\mathsf{H}(P)=\sup_{P\in\Delta_{\boldsymbol{p},0}}\mathsf{H}(P). \tag{5}$$

The limit on the left-hand side (LHS) of (5) exists since $\sup_{P\in\overline{\Delta_{\boldsymbol{p},\varepsilon}}}\mathsf{H}(P)$ is a monotone function of $\varepsilon$.

Since $\Delta_{\boldsymbol{p},0}\subseteq\overline{\Delta_{\boldsymbol{p},\varepsilon}}$ for all $\varepsilon>0$, the RHS above cannot exceed the LHS. To prove the reverse inequality, suppose that $P_\varepsilon$ achieves the supremum over $P\in\overline{\Delta_{\boldsymbol{p},\varepsilon}}$. Passing to a subsequence if necessary, $P_\varepsilon$ converges (as $\varepsilon\to0^+$) to some $P_0\in\Delta(G_{q;a,b})$. From the fact that $\mathbb{E}_P(\mathcal{I}_\Sigma)$ is continuous in $P$, it follows that $P_0\in\Delta_{\boldsymbol{p},0}$. Hence, again via the continuity of the mapping $P\mapsto\mathsf{H}(P)$, we obtain

$$\lim_{\varepsilon\to0^+}\sup_{P\in\overline{\Delta_{\boldsymbol{p},\varepsilon}}}\mathsf{H}(P)=\lim_{\varepsilon\to0^+}\mathsf{H}(P_\varepsilon)=\mathsf{H}(P_0)\le\sup_{P\in\Delta_{\boldsymbol{p},0}}\mathsf{H}(P),$$

which proves our claim. $\square$

Thus, computation of the quantity $\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p})$ requires the solution of a constrained optimization problem in which the objective function $P\mapsto\mathsf{H}(P)$ is concave, and the constraints are linear. The theory of convex duality based upon Lagrange multipliers provides a method to translate the problem into an unconstrained optimization with a convex objective function [4].

In order to reformulate the problem, we need to introduce a vector-valued matrix function that generalizes the adjacency matrix of a directed graph $G = (V, E)$. For a function $\boldsymbol{f} : E \to \mathbb{R}^t$ and $\boldsymbol{\xi} \in \mathbb{R}^t$, let $A_{G;\boldsymbol{f}}(\boldsymbol{\xi})$ be the matrix defined by

$$\Big( A_{G;\boldsymbol{f}}(\boldsymbol{\xi}) \Big)_{u,v} = \sum_{e \in E_{\mathrm{out}}(u) \cap E_{\mathrm{in}}(v)} 2^{-\boldsymbol{\xi} \cdot \boldsymbol{f}(e)} .$$

We remark that for any function $\boldsymbol{f}$, the matrix $A_{G;\boldsymbol{f}}(\boldsymbol{0})$ is precisely the adjacency matrix of $G$. Moreover, for any choice of $\boldsymbol{\xi} \in \mathbb{R}^t$, the matrix $A_{G;\boldsymbol{f}}(\boldsymbol{\xi})$ is (entry-wise) non-negative, so that it has a unique largest positive eigenvalue, called the Perron eigenvalue, which we denote by $\lambda(A_{G;\boldsymbol{f}}(\boldsymbol{\xi}))$.

The following lemma is the main tool in translating the constrained optimization problem to a more tractable form. It is a consequence of standard results in the theory of convex duality.

**Lemma 2.** *Let $G$ and $\boldsymbol{f}$ be as above. Then, for any $\boldsymbol{r} \in \mathbb{R}^t$,*

$$\sup_{\substack{P \in \Delta(G): \\ \mathbb{E}_P(\boldsymbol{f}) = \boldsymbol{r}}} \mathsf{H}(P) = \inf_{\boldsymbol{\xi} \in \mathbb{R}^t} \{ \, \boldsymbol{\xi} \cdot \boldsymbol{r} + \log \lambda(A_{G;\boldsymbol{f}}(\boldsymbol{\xi})) \, \} \ .$$

Note that since $P \mapsto \mathsf{H}(P)$ is a concave function, by convex duality, the objective function on the RHS of the lemma is a convex function of $\boldsymbol{\xi}$. Moreover, it is a differentiable function of $\xi$ whenever the graph $G$ is strongly-connected (as is the case when $G = G_{q;a,b}$): the matrix $A_{G;\boldsymbol{f}}(\boldsymbol{\xi})$ is then irreducible for all $\boldsymbol{\xi} \in \mathbb{R}^t$, so that its Perron eigenvalue is simple, and hence differentiable as a function of $\boldsymbol{\xi}$. Consequently, the objective function can be minimized by identifying the point at which its gradient with respect to $\boldsymbol{\xi}$ vanishes.

We illustrate the use of Proposition 1 and Lemma 2 to determine $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ in the case of $q = 3$ in Section III-A. We will later show in Section III-B that the general $q \geq 3$ case can be reduced to $q = 3$.

## III. Computation of $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$

The simplest case is that of $q = 2$, i.e., the $\mathsf{S}_{2;1,1}$ constrained system. This is the "no-101" constrained system, which forbids the occurrence of the string 101. The value of $\mathsf{cap}(\mathsf{S}_{2;1,1}, (1-p, p))$, for $p \in (0, 1)$, can be computed via Proposition 1 and Lemma 2, using an analysis similar to (but simpler than) that in Section III-A. However, we do not provide the details of this analysis, as it is not difficult to convince oneself that $\mathsf{cap}(\mathsf{S}_{2;1,1}, (1-p, p)) = \mathsf{cap}(\mathsf{S}_{3;1,1}, (1-p, 0, p))$. Thus, we start with the $q = 3$ case.

### A. The Case $q = 3$ and $a = b = 1$

The key to our analysis of the capacity $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ is the case $(q; a, b) = (3; 1, 1)$. As noted above, this case subsumes the case $(q; a, b) = (2; 1, 1)$. Moreover, as we will show in the next subsection, the computation of $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ for any $q \geq 3$, $a \geq 1$, and $b \geq 1$ can be reduced to the problem of computing $\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho})$, where the entries of $\boldsymbol{\rho}$ are $\rho_X = \sum_{s \in X} p_s$, for $X \in \{L, I, H\}$.

So, consider a ternary alphabet $\Sigma$ partitioned into singleton subsets $L$, $I$, and $H$. By abuse of notation, we will assume that $L$, $I$, and $H$ are the actual elements of the alphabet $\Sigma$. The graph presentation of $\mathsf{S}_{3;1,1}$ is given by Fig. 1, regarding each arrowed line in the figure as a single edge.

Let the count constraint vector be $\boldsymbol{\rho} = (\rho_L, \rho_I, \rho_H)$, with $\rho_L, \rho_H \in (0, 1)$ and $\rho_I \in [0, 1)$. From Proposition 1 and Lemma 2 (applied with $\boldsymbol{f} = (\mathcal{I}_I, \mathcal{I}_H)$), we obtain

$$\begin{aligned} &\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho}) \\ &\quad = \inf_{(\xi_I, \xi_H) \in \mathbb{R}^2} \left\{ \rho_I \xi_I + \rho_H \xi_H + \log \lambda(A_{G;(\mathcal{I}_I, \mathcal{I}_H)}(\xi_I, \xi_H)) \right\} , \end{aligned} \tag{6}$$

where

$$A_{G;(\mathcal{I}_I, \mathcal{I}_H)}(\xi_I, \xi_H) = \begin{pmatrix} 2^{-\xi_H} & 1 & 2^{-\xi_I} \\ 0 & 0 & 1 + 2^{-\xi_I} \\ 2^{-\xi_H} & 0 & 1 + 2^{-\xi_I} \end{pmatrix} . \tag{7}$$

As noted after Lemma 2, the objective function on the RHS of (6) can be minimized by identifying the point at which its gradient with respect to $(\xi_I, \xi_H)$ equals 0.

The case $\rho_I = 0$ needs a little extra care, as in this case the infimum in (6) is achieved by letting $\xi_I \to \infty$. This follows from the fact that for any fixed $\xi_H$, the Perron eigenvalue $\lambda(A_{G;(\mathcal{I}_I, \mathcal{I}_H)}(\xi_I, \xi_H))$ is strictly decreasing in $\xi_I$ (see Problem 3.12 in [5]). Thus, the RHS of (6) reduces to the single-variable optimization problem $\inf_{\xi_H} \{\rho_H \xi_H + \log \lambda(A_{G;\mathcal{I}_H}(\xi_H))\}$, where $A_{G;\mathcal{I}_H}(\xi_H)$ is the matrix obtained by setting $2^{-\xi_I} = 0$ in (7).

We first assume that $\rho_I > 0$ (describing later the minor modifications to be made to handle the case $\rho_I = 0$). We make the change of variables $y = 2^{-\xi_I}$ and $z = 2^{-\xi_H}$ to get

$$\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho}) = \log \left( \inf_{y, z \in (0, \infty)^2} \frac{\lambda(A(y, z))}{y^{\rho_I} z^{\rho_H}} \right) , \tag{8}$$

where $\lambda(A(y, z))$ is the Perron eigenvalue of the matrix

$$A(y, z) := \begin{pmatrix} z & 1 & y \\ 0 & 0 & 1 + y \\ z & 0 & 1 + y \end{pmatrix} .$$

It is easily checked that the determinant of the Jacobian of the transformation $(\xi_I, \xi_H) \mapsto (y, z)$ is nonzero for all $(\xi_I, \xi_H) \in \mathbb{R}^2$. It follows from this that for any $(\xi_I, \xi_H) \in \mathbb{R}^2$, the gradient of the objective function in (6) is 0 at $(\xi_I, \xi_H)$ if and only if the gradient of the objective function in (8) is 0 at $(y, z) = (2^{-\xi_I}, 2^{-\xi_H})$. Thus, the minimization in (8) can be carried out by identifying the positive values of $y, z$ at which the gradient of $\lambda(A(y, z))/(y^{\rho_I} z^{\rho_H})$ vanishes.

To do this, we make another convenient change of variables: $(y, z) \mapsto (y, \lambda)$ with $\lambda = \lambda(A(y, z))$. This mapping is invertible: since $A(y, z)$ is irreducible for $y, z > 0$, it follows from Problem 3.12 in [5] that $\lambda(A(y, z))$ is strictly increasing in $z$ for every fixed $y > 0$. Also, for each fixed $y > 0$, the mapping $z \mapsto \lambda(A(y, z))$ is a continuous function from $(0, \infty)$ onto $(y + 1, \infty)$ (as it is easy to see that $\lambda(A(y, 0)) = y + 1$). For every fixed $y > 0$, the inverse mapping $(y, \lambda) \mapsto (y, z)$ is determined by setting the characteristic polynomial of $A(y, z)$ equal to 0, and is given by

$$z = z(y, \lambda) = \frac{\lambda^2(\lambda - y - 1)}{\lambda^2 - \lambda + y + 1} . \tag{9}$$

It can be verified by direct computation[5] that $\partial z/\partial\lambda > 0$ whenever $\lambda > y+1 > 1$, and hence, the Jacobian determinant of the transformation $(y,z) \mapsto (y,\lambda)$ is nonzero for all $y,z > 0$. From this, arguing as above for the mapping $(\xi_I,\xi_H) \mapsto (y,z)$, we obtain via (8) and (9) that

$$\mathsf{cap}(\mathsf{S}_{3;1,1},\boldsymbol{\rho}) = \log\left(\inf_{(y,\lambda)\in\mathcal{U}} g(y,\lambda)\right), \tag{10}$$

where

$$g(y,\lambda) = \frac{\lambda}{y^{\rho_I}}\left(\frac{\lambda^2-\lambda+y+1}{\lambda^2(\lambda-y-1)}\right)^{\rho_H}$$

and

$$\mathcal{U} = \{(y,\lambda)\in\mathbb{R}^2 : \lambda > y+1 > 1\} .$$

Moreover, the infimum in (10) is obtained at any point $(y,\lambda) \in \mathcal{U}$ where the partial derivatives of $g(y,\lambda)$ vanish.

Turning now to the case $\rho_I = 0$, it can be handled by setting $y = 0$ in the discussion above, assuming the convention that $0^0 = 1$. Thus, the RHS of (8) reduces to $\log\left(\inf_{z>0}\lambda(A(0,z))/z^{\rho_H}\right)$, and the RHS of (10) becomes $\log\left(\inf_{\lambda>1} g_0(\lambda)\right)$, where $g_0(\lambda) := g(0,\lambda)$. This latter infimum is achieved at any point $\lambda > 1$ where the derivative $g_0'(\lambda)$ equals 0.

In Appendix C in the full version of this paper [2], we compute the partial derivatives $\partial g/\partial y$ and $\partial g/\partial\lambda$, each being a cubic multinomial in $y$ and $\lambda$. We then find explicitly their common root, thereby yielding the following result.

**Theorem 3.** *For $\boldsymbol{\rho} = (\rho_L,\rho_I,\rho_H) \in (0,1)\times[0,1)\times(0,1)$:*

$$\mathsf{cap}(\mathsf{S}_{3;1,1},\boldsymbol{\rho}) = \log\left[\frac{\lambda}{y^{\rho_I}}\left(\frac{\lambda^2-\lambda+y+1}{\lambda^2(\lambda-y-1)}\right)^{\rho_H}\right],$$

*where $(y,\lambda)$ is given as follows.*

- *If $\rho_L = \frac{1}{2}$ and $\rho_I = 0$, then $y = 0$ and $\lambda = 2$.*
- *If $\rho_L = \frac{1}{2}$ and $\rho_I > 0$, then*

$$y = -1-2\tau+2\sqrt{1+\tau+\tau^2} \quad \textit{and} \quad \lambda = 2 ,$$

*where $\tau := \rho_H/\rho_I$.*

- *If $\rho_L \neq \frac{1}{2}$, then*

$$y = \frac{\rho_I(\lambda-2)}{1-2\rho_L}$$

*and $\lambda$ is a root of the cubic polynomial*

$$\begin{aligned} Z(x) := (1-2\rho_L)\rho_L x^3 + \big((\rho_L-\rho_H)^2-(1-2\rho_L)\big)x^2 \\ -2(\rho_L-\rho_H)(1-2\rho_H)x+(1-2\rho_H)^2 \end{aligned}$$

*chosen as follows: if $\rho_L < \frac{1}{2}$, then $\lambda$ is the largest real (positive) root of $Z(x)$; and if $\rho_L > \frac{1}{2}$, then $\lambda$ is the smallest real (positive) root of $Z(x)$.*

It is worth noting that for $\boldsymbol{\rho} = (\frac{1}{2},0,\frac{1}{2})$ we obtain $\mathsf{cap}(\mathsf{S}_{3;1,1},(\frac{1}{2},0,\frac{1}{2})) = \frac{1}{2}\log 3$. Thus, $\mathsf{cap}(\mathsf{S}_{2;1,1},(\frac{1}{2},\frac{1}{2})) = \frac{1}{2}\log 3$, which agrees with the rate derived (using two different approaches) in [6].

[5] Also see Appendix B in [2] for an argument using Perron–Frobenius theory.

*B. The General Case of $q \geq 3$*

Consider now a constrained system $\mathsf{S}_{q;a,b}$ over a $q$-ary alphabet $\Sigma$ for some $q \geq 3$, where $\Sigma$ is partitioned into the subsets $L$, $H$, and $I$ of sizes $a \geq 1$, $b \geq 1$, and $q-a-b \geq 0$, respectively. Let $\boldsymbol{p} = (p_s)_{s\in\Sigma}$ be a given count constraint vector, and define $\rho_X = \sum_{s\in X} p_s$ for $X \in \{L,I,H\}$. If $I = \emptyset$, we set $\rho_I = 0$. The probabilities $\rho_L$ and $\rho_H$ are assumed to be strictly positive.

The aim of this subsection is to prove the result stated next. The statement requires the following standard definition: the *entropy* of a probability vector $\boldsymbol{u} = (u_i)_i$ is defined as $\mathsf{h}(\boldsymbol{u}) = -\sum_i u_i \log u_i$.

**Theorem 4.** *For $\mathsf{S}_{q;a,b}$ and $\boldsymbol{p}$ as above:*

$$\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p}) = \mathsf{cap}(\mathsf{S}_{3;1,1},\boldsymbol{\rho}) + \mathsf{h}(\boldsymbol{p}) - \mathsf{h}(\boldsymbol{\rho}) ,$$

*where the entries of $\boldsymbol{\rho}$ are $\rho_X = \sum_{s\in X} p_s$, for $X \in \{L,I,H\}$.*

Thus, the computation of $\mathsf{cap}(\mathsf{S}_{q;a,b},\boldsymbol{p})$ reduces to the problem of computing $\mathsf{cap}(\mathsf{S}_{3;1,1},\boldsymbol{\rho})$, which was solved explicitly in Theorem 3. The rest of this subsection is devoted to a proof of Theorem 4.

Let $\mathsf{P} = (\mathsf{P}(u,s))_{u,s}$ be a stationary Markov chain on the labeled graph $G_{q;a,b}$ on the vertex set $V = \{1,2,3\}$ in Fig. 1, where $\mathsf{P}(u,s)$ is the probability of the edge labeled $s$ leaving vertex $u$ (and $\mathsf{P}(u,s) \equiv 0$ if there is no such edge). Note that for any $s \in \Sigma$, we have $\mathbb{E}_{\mathsf{P}}(\mathcal{I}_s) = \sum_{u\in V}\mathsf{P}(u,s)$. Thus, the constraint $\mathbb{E}_{\mathsf{P}}(\mathcal{I}_\Sigma) = \boldsymbol{p}$ on the RHS of (4) is equivalently expressed as

$$\sum_{u\in V}\mathsf{P}(u,s) = p_s , \quad \text{for all } s \in \Sigma . \tag{11}$$

Now, for $u \in V$ and $X \in \{L,I,H\}$, define

$$\mathsf{Q}(u,X) = \sum_{s\in X}\mathsf{P}(u,s) .$$

Note that $\mathsf{Q} = (\mathsf{Q}(u,X))_{u,X}$ is a stationary Markov chain on the graph in Fig. 1, where each arrowed line in the figure is regarded as a single edge (this graph is $G_{3;1,1}$). Moreover, we have for $X \in \{L,I,H\}$,

$$\mathbb{E}_{\mathsf{Q}}(\mathcal{I}_X) = \sum_{u\in V}\mathsf{Q}(u,X) = \sum_{s\in X}\sum_{u\in V}\mathsf{P}(u,s) .$$

Thus, if we impose the constraint (11) on the Markov chain $\mathsf{P}$, we obtain

$$\mathbb{E}_{\mathsf{Q}}(\mathcal{I}_X) = \sum_{s\in X} p_s = \rho_X \text{ , for all } X \in \{L,I,H\} .$$

In other words, the constraint $\mathbb{E}_{\mathsf{P}}(\mathcal{I}_\Sigma) = \boldsymbol{p}$ on the Markov chain $\mathsf{P}$ induces the constraint $\mathbb{E}_{\mathsf{Q}}(\mathcal{I}_{\{L,I,H\}}) = \boldsymbol{\rho}$ on the Markov chain $\mathsf{Q}$. Finally, observe that $\mathsf{P}$ and $\mathsf{Q}$ induce the same stationary distribution on $V$:

$$\begin{aligned} \pi_{\mathsf{P}}(u) = \sum_{s\in\Sigma}\mathsf{P}(u,s) &= \sum_X\sum_{s\in X}\mathsf{P}(u,s) \\ &= \sum_X \mathsf{Q}(u,X) = \pi_{\mathsf{Q}}(u) . \end{aligned}$$

The following lemma is the key to proving Theorem 4.

**Lemma 5.** *For a Markov chain* $\mathsf{P} \in \Delta(G_{q;a,b})$ *with* $\mathbb{E}_{\mathsf{P}}(\mathcal{I}_\Sigma) = \boldsymbol{p}$*, and* $\mathsf{Q} \in \Delta(G_{3;1,1})$ *as above, we have*

$$\mathsf{H}(\mathsf{P}) \le \mathsf{H}(\mathsf{Q}) + \mathsf{h}(\boldsymbol{p}) - \mathsf{h}(\boldsymbol{\rho}) \ , \tag{12}$$

*with equality holding if and only if* $\mathsf{P}(u,s) = (p_s/\rho_X)\mathsf{Q}(u,X)$ *for every* $s \in X$ *(where* $\mathsf{P}(u,s) = 0$ *when* $p_s = \rho_X = 0$*).*

*Proof.* Let $(U,S)$ be a pair of random variables taking values $(u,s) \in V \times \Sigma$ with probability $\mathsf{P}(u,s)$. Let $\varphi : \Sigma \to \{L,I,H\}$ be the function that maps $s$ to $X$ if $s \in X$. Now, $U$—$S$—$\varphi(S)$ is a Markov chain, so that by the data processing inequality, $\mathrm{I}(U;S) \ge \mathrm{I}(U;\varphi(S))$. It is easily verified that $\mathrm{I}(U;S) = \mathsf{h}(\boldsymbol{p}) - \mathsf{H}(\mathsf{P})$ and $\mathrm{I}(U;\varphi(S)) = \mathsf{h}(\boldsymbol{\rho}) - \mathsf{H}(\mathsf{Q})$. Thus,

$$\mathsf{h}(\boldsymbol{p}) - \mathsf{H}(\mathsf{P}) \ge \mathsf{h}(\boldsymbol{\rho}) - \mathsf{H}(\mathsf{Q}) \ ,$$

re-arranging which we obtain (12).

Equality holds in the data processing inequality above if and only if $U$—$\varphi(S)$—$S$ is also a Markov chain, i.e., $U$ and $S$ are conditionally independent given $\varphi(S)$. Now, check that $\Pr\{U = u, S = s \mid \varphi(S) = X\}$ equals $\mathsf{P}(u,s)/\rho_X$ if $s \in X$, and equals 0 otherwise. Hence, $\Pr\{U = u, S = s \mid \varphi(S) = X\} = \sum_{s \in X} \mathsf{P}(u,s)/\rho_X = \mathsf{Q}(u,X)/\rho_X$. Finally, $\Pr\{S = s \mid \varphi(S) = X\}$ equals $p_s/\rho_X$ if $s \in X$, and equals 0 otherwise. Thus, the required conditional independence holds if and only if $\mathsf{P}(u,s) = \mathsf{Q}(u,X)\,(p_s/\rho_X)$ for all $s \in X$. □

We can now complete the proof of Theorem 4. Taking the supremum over $\mathsf{P}$ in (12), we obtain (by virtue of (4)) that

$$\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p}) \le \mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho}) + \mathsf{h}(\boldsymbol{p}) - \mathsf{h}(\boldsymbol{\rho}) \ . \tag{13}$$

We now argue that this is in fact an equality. Consider a $\mathsf{Q}^* = \big(\mathsf{Q}^*(u,X)\big)_{u,X}$ that achieves $\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho}) = \sup \mathsf{H}(\mathsf{Q})$, the supremum being over Markov chains $\mathsf{Q} \in \Delta(G_{3;1,1})$ such that $\mathbb{E}_{\mathsf{Q}}(\mathcal{I}_{\{L,I,H\}}) = \boldsymbol{\rho}$. Such a $\mathsf{Q}^*$ exists as $\mathsf{Q} \mapsto \mathsf{H}(\mathsf{Q})$ is a continuous function being maximized over a compact set. Recall that any outgoing edge from $u$ labeled by $X$ in $G_{3;1,1}$ is replaced in $G_{q;a,b}$ by $|X|$ parallel edges labeled by the distinct symbols $s \in X$. For each such edge $(u,s)$, set $\mathsf{P}(u,s) = (p_s/\rho_X)\mathsf{Q}^*(u,X)$. The resulting Markov chain $\mathsf{P} \in \Delta(G_{q;a,b})$ satisfies the conditions for equality in (12), from which it follows that equality holds in (13).

## IV. Discussion

Our computation of $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ consists of the following steps.

1) Applying Theorem 4 to reduce the problem to that of computing $\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho})$.
2) Expressing the computation of $\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho})$ as the bivariate minimization problem (8) in the variables $(y,z)$.
3) Eliminating the implicit expression $\lambda(A(y,z))$ in (8) through a change of variables, resulting in the bivariate minimization problem (10) in the variables $(y,\lambda)$.
4) Taking partial derivatives with respect to $y$ and $\lambda$, resulting in two cubic bivariate polynomials in $y$ and $\lambda$.
5) Finding the common root of these polynomials.

While Step 1 is specific to the constrained system $\mathsf{S}_{q;a,b}$, the other steps might be applicable to other count-constrained systems (albeit with varying degrees of difficulty). For any constrained system $\mathsf{S}$ over an alphabet $\Sigma$, the number of variables in Step 2 will be $|\Sigma| - 1$. As for Step 3, the explicit rational expression (9) for $z = z(y,\lambda)$ is attributed to the fact that the coefficients of the characteristic polynomial of $A(y,z)$ are linear terms in $z$. In general, this happens whenever there is a symbol $s \in \Sigma$ that has a "home state" in the graph presentation of $\mathsf{S}$, namely, all edges labeled by $s$ lead to the same vertex.

We mention that one could also compute $\mathsf{cap}(\mathsf{S}_{q;a,b}, \boldsymbol{p})$ based on Proposition 1 directly. Referring to the case $\mathsf{cap}(\mathsf{S}_{3;1,1}, \boldsymbol{\rho})$ and using the notation towards the end of Section III-B, such a computation would entail finding the nine edge probabilities of a Markov chain $\mathsf{Q} = \big(\mathsf{Q}(u,X)\big)_{u,X}$ (where $u \in V = \{1,2,3\}$ and $X \in \Sigma = \{L,I,H\}$) that maximizes $\mathsf{H}(\mathsf{Q})$, subject to the following six linear constraints:

- $\mathsf{Q}(2,H) = 0$,
- the constraints (3) for any two vertices in $V$ (the third is dependent on these two), and—
- the three constraints obtained from $\mathbb{E}(\mathcal{I}_\Sigma) = \boldsymbol{\rho}$ (these constraints imply that $\sum_{u,X} \mathsf{Q}(u,X) = 1$).

We would then end up with three linearly independent variables to optimize over.


## Acknowledgment

N. Kashyap and P. H. Siegel would like to acknowledge the support of the Indo–US Science & Technology Forum (IUSSTF), which funded in part the work reported here. This work was also supported in part by NSF Grant CCF-1619053, and by Grant 2015816 from the United-States–Israel Binational Science Foundation (BSF). Portions of this work were conducted while P. H. Siegel visited Technion in May 2013 and while R. M. Roth visited the Center for Memory and Recording Research (CMRR) at UC San Diego in summer 2018.


## References


[1] Y. M. Chee, J. Chrisnata, H. M. Kiah, S. Ling, T. T. Nguyen, and V. K. Vu, "Rates of constant-composition codes that mitigate intercell interference," to appear in *IEEE Trans. Inf. Theory*. DOI: 10.1109/TIT.2018.2884210

[2] N. Kashyap, R. M. Roth, and P. H. Siegel, "The capacity of count-constrained ICI-free channels," ArXiv e-print, 2019. http://arxiv.org/abs/1901.03261

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

[4] 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, July 1992.

[5] B. H. Marcus, R. M. Roth, and P. H. Siegel, *An Introduction to Coding for Constrained Systems, Fifth Ed.*, unpublished course textbook, Oct. 2001.

[6] 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, 2014.

[7] V. K. Vu, *Constrained Codes for Intercell Interference Mitigation and Dynamic Thresholding in Flash Memories*, Ph.D. thesis, School of Physical and Mathematical Sciences, Nanyang Technol. Univ., Singapore, 2017.