Convex Programming Upper Bounds on the Capacity of 2-D Constraints*

Ido Tal, *Member, IEEE* and Ron M. Roth, *Fellow, IEEE*

***Abstract*—The capacity of 1-D constraints is given by the entropy of a corresponding stationary maxentropic Markov chain. Namely, the entropy is maximized over a set of probability distributions, which is defined by some linear equalities and inequalities. In this paper, certain aspects of this characterization are extended to 2-D constraints. The result is a method for calculating an upper bound on the capacity of 2-D constraints.**

The key steps are as follows. The maxentropic stationary probability distribution on square configurations is considered; a set of linear equalities and inequalities is derived from this stationarity; the result is then a convex program, which can be easily solved numerically. Our method improves upon previous upper bounds for the capacity of the 2-D "no isolated bits" constraint, as well as certain 2-D RLL constraints.

***Index Terms*—Concave function maximization, Convex programming, Runlength-limited constraints, No isolated bits constraint, Two-dimensional constraints.**

I. Introduction

Let Σ be a finite alphabet. A one-dimensional (1-D) constraint is a set S of words over Σ. For the set S to be called a 1-D constraint, there must exist an edge-labeled graph G with the following property: a word $\mathbf{w} = w_0 w_2 \ldots w_{n-1}$ is in S iff there exists a path in G for which the successive edge labels are $w_0, w_2, \ldots, w_{n-1}$ (see [1]).

A two dimensional (2-D) constraint over Σ is a generalization of a 1-D constraint. However, let us first define a *configuration*, the 2-D generalization of a word. Denote the set of integers by $\mathbb{Z}$. A 2-D index set $\mathsf{U} \subseteq \mathbb{Z}^2$ is a set of integer pairs. A 2-D configuration over Σ with index set U is a function $w : \mathsf{U} \to \Sigma$. We denote such a configuration as $w = (w_{i,j})_{(i,j)\in\mathsf{U}}$, where for all $(i,j) \in \mathsf{U}$, we have that $w_{i,j} \in \Sigma$. A configuration $w : \mathsf{B} \to \Sigma$ is *rectangular* if B is rectangular. Namely, there exist integers $\alpha_1 < \alpha_2$ and $\beta_1 < \beta_2$ such that

$$\mathsf{B} = \{(i,j) : \alpha_1 \le i < \alpha_2 \ , \quad \beta_1 \le j < \beta_2\} \ .$$

A 2-D constraint is a set $\mathbb{S}$ of rectangular configurations over Σ and is defined through a pair of *vertex*-labeled graphs $(G_{\mathrm{row}}, G_{\mathrm{col}})$, where $G_{\mathrm{row}} = (V, E_{\mathrm{row}}, L)$ and $G_{\mathrm{col}} = (V, E_{\mathrm{col}}, L)$. Namely, both graphs share the same vertex set and the same vertex labeling function $L : V \to \Sigma$. The constraint $\mathbb{S} = \mathbb{S}(G_{\mathrm{row}}, G_{\mathrm{col}})$ consists of all rectangular configurations $(w_{i,j})_{(i,j)\in\mathsf{B}}$ over Σ with the following property: There exists a configuration $(u_{i,j})_{(i,j)\in\mathsf{B}}$ over the vertex set V such that (a) for each $(i,j) \in \mathsf{B}$ we have $w_{i,j} = L(u_{i,j})$; (b) each row in $(u_{i,j})$ is a path in G_{row}; (c) each column in $(u_{i,j})$ is a path in G_{col}. Examples of 2-D constraints include the square constraint [2], 2-D runlength-limited (RLL) constraints [3], 2-D symmetric runlength-limited (SRLL) constraints [4], and the "no isolated bits" (n.i.b.) constraint [5].

Let $\mathbb{S}$ be a given 2-D constraint over a finite alphabet Σ. For positive integers $M, N > 0$, denote by $\mathbb{S}_{M,N}$ all the configurations in $\mathbb{S}$ with index set

$$\mathsf{B}_{M,N} = \{(i,j) : 0 \le i < M \ , \quad 0 \le j < N\} \ .$$

Also, let

$$\mathsf{B}_M = \mathsf{B}_{M,M} \quad \text{and} \quad \mathbb{S}_M = \mathbb{S}_{M,M} \ .$$

The capacity of $\mathbb{S}$ is defined as

$$\mathsf{cap}(\mathbb{S}) = \lim_{M\to\infty} \frac{1}{M^2} \cdot \log_2 |\mathbb{S}_M| \ . \tag{1}$$

In this paper, we show a method for calculating an upper bound on $\mathsf{cap}(\mathbb{S})$. Two other methods for calculating an upper bound on the capacity of a 2-D constraint are as follows. The first method is the so called "stripe method," in which we fix a positive integer N and bound $\mathsf{cap}(\mathbb{S})$ by

$$\mathsf{cap}(\mathbb{S}) \le \lim_{M\to\infty} \frac{1}{M \cdot N} \cdot \log_2 |\mathbb{S}_{M,N}| \ . \tag{2}$$

Namely, we consider only stripes of width N, and essentially get a 1-D constraint (since we may regard each of the possible row values as a symbol in an auxiliary alphabet). The RHS of (2) is easily calculated for modest values of N: Let G be the edge-labeled graph corresponding to the 1-D constraint, and let A_G be the adjacency matrix of G. Denote by $\lambda(A_G)$ the Perron eigenvalue of A_G. By [1, §3.2], the RHS of (2) is equal to $\lambda(A_G)$. The second method for bounding $\mathsf{cap}(\mathbb{S})$ from above is the generalization presented by Forchhammer and Justesen [6] to the method of Calkin and Wilf [7].

The capacity of a given 1-D constraint is known to be equal to the value of an optimization program, where the optimization is on the entropy of a certain stationary Markov

* This work was presented in part at the *IEEE International Symposium on Information Theory (ISIT2009)*, Seoul, Korea, (June 2009), pp. 1060–1064 in the proceedings.
I. Tal is with the Information Theory and Applications (ITA) Center and is affiliated with the Center for Magnetic Recording Research (CMRR), both at the University of California at San Diego, La Jolla, CA 92093-0401, USA. This work was done while he was with the Computer Science Department, Technion, Haifa 32000, Israel (email: idotal@ieee.org).
R. M. Roth is with the Computer Science Department, Technion, Haifa 32000, Israel (email: ronny@cs.technion.ac.il).
This work was supported in part by grant No. 2008258 from the United-States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. The work of Ido Tal was also supported by University of California Lab Fees Research Program, Award No. 09-LR-06-118620-SIEP.

chain, and is carried out over the conditional probabilities of that chain (see [1, Section 3.2.3]). We try to extend certain aspects of this characterization of capacity to 2-D constraints. What results is a (generally non-tight) upper bound on $\mathsf{cap}(\mathbb{S})$.

The structure of this paper is as follows. In Section II, we set up some notation. Then, in Section III, we show the existence of a certain stationary random variable taking values on $\mathbb{S}_M$ and having entropy approaching the capacity of $\mathbb{S}$, as $M \to \infty$. We then consider a relatively small sub-configuration of that random variable, and denote it by $X^{(M)}$. The section concludes with an upper bound on the capacity of $\mathbb{S}$, which is a function of the probability distribution of $X^{(M)}$. In Section IV, we derive a set of linear equations which hold on the probability distribution of $X^{(M)}$. In Section V, we argue as follows: The bound derived in Section III is a function of the probability distribution of $X^{(M)}$, which we do not know how to calculate; however, by Section IV we know that this probability distribution is subject to a set of linear requirements (equalities and inequalities). Thus, we formalize an optimization problem, where the unknown probability distribution is replaced by a set of variables, subject to the above-mentioned linear requirements. The maximum of this optimization problem is an upper bound on the capacity of $\mathbb{S}$. We then show that this optimization problem is easily solved, since it is an instance of convex programming. In Section VI, we show our computational results. Finally, in Section VII we present an asymptotic analysis of our method.

We note at this point that although this paper deals with 2-D constraints, our method can be easily generalized to higher dimensions as well.

II. Notation

This section is devoted to setting up some notation.

A. Shifts and restrictions

In this paper, index sets will always be denoted by upper-case Greek letters or upper-case Roman letters in the sans-serif font. For integers α, β we denote the shifting of an index set U by (α, β) as

$$\sigma_{\alpha,\beta}(\mathsf{U}) = \{(i+\alpha, j+\beta) : (i,j) \in \mathsf{U}\} \ .$$

Let $(w_{i,j})_{(i,j)\in\mathsf{U}}$ be a configuration with index set U. By abuse of notation, let $\sigma_{\alpha,\beta}(w)$ be the shifted configuration (with index set $\sigma(\mathsf{U})$):

$$\sigma_{\alpha,\beta}(w)_{i+\alpha,j+\beta} = w_{i,j} \ .$$

For a configuration w with index set U, and an index set $\mathsf{V} \subseteq \mathsf{U}$, denote the restriction of w to V by $w[\mathsf{V}] = (w[\mathsf{V}]_{i,j})_{(i,j)\in\mathsf{V}}$; namely,

$$w[\mathsf{V}]_{i,j} = w_{i,j} \ , \quad \text{where} \quad (i,j) \in \mathsf{V} \ .$$

We denote the restriction of $\mathbb{S}$ to U by $\mathbb{S}[\mathsf{U}]$:

$$\mathbb{S}[\mathsf{U}] = \{w : \text{there exists } w' \in \mathbb{S} \text{ such that } w'[\mathsf{U}] = w\} \ . \quad (3)$$

B. Strict total order

A *strict total order* $\prec$ is a relation on $\mathbb{Z}^2 \times \mathbb{Z}^2$, satisfying the following conditions for all $(i_1,j_1),(i_2,j_2),(i_3,j_3) \in \mathbb{Z}^2$.

- If $(i_1,j_1) \neq (i_2,j_2)$, then either $(i_1,j_1) \prec (i_2,j_2)$ or $(i_2,j_2) \prec (i_1,j_1)$, but not both.
- If $(i_1,j_1) = (i_2,j_2)$, then neither $(i_1,j_1) \prec (i_2,j_2)$ nor $(i_2,j_2) \prec (i_1,j_1)$.
- If $(i_1,j_1) \prec (i_2,j_2)$ and $(i_2,j_2) \prec (i_3,j_3)$, then $(i_1,j_1) \prec (i_3,j_3)$.

For $(i,j) \in \mathbb{Z}^2$, define $\mathsf{T}^{(\prec)}_{i,j}$ as all the indexes preceding (i,j). Namely,

$$\mathsf{T}^{(\prec)}_{i,j} = \{(i',j') \in \mathbb{Z}^2 : (i',j') \prec (i,j)\} \ .$$

C. Entropy

Let X and Y be two random variables. Denote

$$p_x = \mathrm{Prob}(X = x) \ .$$

and

$$p_{y|x} = \mathrm{Prob}(X = x, Y = y)/\mathrm{Prob}(X = x) \ .$$

The entropy of X is denoted by $H(X)$ and is equal to

$$H(X) = \sum_x p_x \log p_x \ ,$$

where the sum is on all x for which $\mathrm{Prob}(X = x)$ is positive. Similarly, we define the conditional entropy $H(Y|X)$ as

$$H(Y|X) = \sum_x p_x \sum_y p_{y|x} \log p_{y|x} \ ,$$

where we sum on all x for which p_x is positive and all y for which $p_{y|x}$ is positive.

III. A preliminary upper bound on $\mathsf{cap}(\mathbb{S})$

Let M be a positive integer and let W be a random variable taking values on $\mathbb{S}_M$. We say that W is *stationary* if for all $\mathsf{U} \subseteq \mathsf{B}_M$, all $\alpha, \beta \in \mathbb{Z}$ such that $\sigma_{\alpha,\beta}(\mathsf{U}) \subseteq \mathsf{B}_M$, and all $w' \in \mathbb{S}[\mathsf{U}]$, we have that

$$\mathrm{Prob}(W[\mathsf{U}] = w') = \mathrm{Prob}(W[\sigma_{\alpha,\beta}(\mathsf{U})] = \sigma_{\alpha,\beta}(w')) \ .$$

The following is a corollary of [8, Theorem 1.4]. The proof is given in Appendix A.

Theorem 1: There exists a series of random variables $(W^{(M)})_{M=1}^{\infty}$ with the following properties: (i) Each $W^{(M)}$ takes values on $\mathbb{S}_M$. (ii) The probability distribution of $W^{(M)}$ is stationary. (iii) The normalized entropy of $W^{(M)}$ approaches $\mathsf{cap}(\mathbb{S})$,

$$\mathsf{cap}(\mathbb{S}) = \lim_{M\to\infty} \frac{1}{M^2} \cdot H(W^{(M)}) \ . \quad (4)$$

We now proceed towards deriving Lemma 2 below, which gives an upper bound on $\mathsf{cap}(\mathbb{S})$, and makes use of the stationarity property. We note in advance that this bound is not actually meant to be calculated. Thus, its utility will be made clear in the following sections. In order to enhance the exposition, we accompany the derivation with two running examples.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

$\prec_{\text{lex}}$

1	2	3	4	5
16	17	18	19	20
6	7	8	9	10
21	22	23	24	25
11	12	13	14	15

$\prec_{\text{irs}}$

Fig. 1. An entry labeled i in the left (right) configuration precedes an entry labeled j according to $\prec_{\text{lex}}$ ($\prec_{\text{irs}}$) iff $i < j$.

Running Example I: Define the lexicographic order $\prec_{\text{lex}}$ as follows: $(i_1, j_1) \prec_{\text{lex}} (i_2, j_2)$ iff

- $i_1 < i_2$, or
- ($i_1 = i_2$ and $j_1 < j_2$).

Running Example II: Define the "interleaved raster scan" order $\prec_{\text{irs}}$ as follows: $(i_1, j_1) \prec_{\text{irs}} (i_2, j_2)$ iff

- $i_1 \equiv 0 \pmod 2$ and $i_2 \equiv 1 \pmod 2$, or
- $i_1 \equiv i_2 \pmod 2$ and $i_1 < i_2$, or
- $i_1 = i_2$ and $j_1 < j_2$.

(See Figure 1 for both examples.)

For the rest of this section, fix positive integers r and s, and define the index set

$$\Lambda = \mathsf{B}_{r,s} \ .$$

We will refer to Λ as "the patch." The bound we derive in Lemma 2 will be a function of the following:

- the strict total order $\prec$,
- the integers r and s, which determine the order $r \times s$ of the patch Λ,
- an integer c, which will denote the number of "colors" we encounter,
- a coloring function $f : \mathbb{Z}^2 \to \{1, 2, \ldots, c\}$, mapping each point in $\mathbb{Z}^2$ to one of c colors,
- c indexes, $(a_\gamma, b_\gamma)_{\gamma=1}^c$, such that for all $1 \le \gamma \le c$,

$$(a_\gamma, b_\gamma) \in \Lambda$$

(namely, each color γ has a designated point in the patch, *which may or may not be of color* γ).

The function f must satisfy two requirements, which we now elaborate on. Our first requirement is: for all $1 \le \gamma \le c$,

$$\lim_{M \to \infty} \frac{|\{(i,j) \in \mathsf{B}_M : f(i,j) = \gamma\}|}{M^2} = \frac{1}{c} \ . \qquad (5)$$

Namely, as the orders of $W^{(M)}$ tend to infinity, each color appears with the same frequency[1]. Our second requirement is as follows: there exist index sets $\Psi_1, \Psi_2, \ldots, \Psi_c \subseteq \Lambda$ such that for all indexes $(i,j) \in \mathbb{Z}^2$,

$$\sigma_{i',j'}(\Psi_\gamma) = \mathsf{T}_{i,j}^{(\prec)} \cap \sigma_{i',j'}(\Lambda) \ , \qquad (6)$$

where $\gamma = f(i,j)$, $i' = i - a_\gamma$, and $j' = j - b_\gamma$. Namely, let (i,j) be such that $f(i,j) = \gamma$, and shift Λ such that (a_γ, b_γ) is shifted to (i,j). Now, consider the set of all indexes in the shifted Λ which precede (i,j): this set must be equal to the correspondingly shifted Ψ_γ.

[1] In fact, it is possible to generalize (5) and require only that the limit exists for all γ. We have not found this generalization useful.

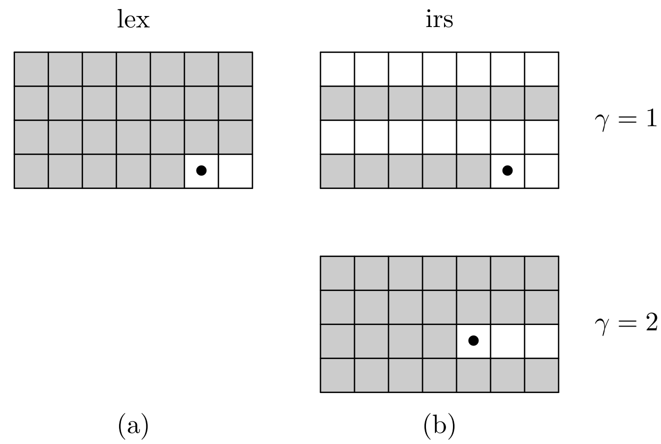

Fig. 2. The left (right) column corresponds to Running Example I (II). The configurations are of order $r \times s$ and represent the index set Λ. The $\bullet$ symbol is in position (a_γ, b_γ). The shaded part is Ψ_γ.

Running Example I: Take $r = 4$ and $s = 7$ as the patch orders. Let the number of colors be $c = 1$. Thus, we must define $f = f_{\text{lex}}$ as follows: for all $(i,j) \in \mathbb{Z}^2$, $f_{\text{lex}}(i,j) = 1$. Take the point corresponding to the single color as $(a_1 = 3, b_1 = 5)$. See also Figure 2(a).

Running Example II: As in the previous example, take $r = 4$ and $s = 7$ as the patch orders. Let the number of colors be $c = 2$. Define $f = f_{\text{irs}}$ as follows:

$$f_{\text{irs}}(i,j) = \begin{cases} 1 & i \equiv 0 \pmod 2 \\ 2 & i \equiv 1 \pmod 2 \end{cases} \ .$$

Take $(a_1 = 3, b_1 = 5)$ and $(a_2 = 2, b_2 = 4)$. See also Figure 2(b).

Lemma 2: Let $(W^{(M)})_{M=1}^\infty$ be as in Theorem 1 and define

$$X^{(M)} = W^{(M)}[\Lambda] \ .$$

Let $\prec$, r, s, c, f, $(\Psi_\gamma)_{\gamma=1}^c$, and $(a_\gamma, b_\gamma)_{\gamma=1}^c$ be given. For $1 \le \gamma \le c$, define

$$\Upsilon_\gamma = \{(a_\gamma, b_\gamma)\} \cup \Psi_\gamma \ .$$

Let

$$Y_\gamma = X^{(M)}[\Upsilon_\gamma] \ \text{ and } \ Z_\gamma = X^{(M)}[\Psi_\gamma]$$

(note that Y_γ and Z_γ are functions of M). Then,

$$\mathsf{cap}(\mathbb{S}) \le \liminf_{M \to \infty} \frac{1}{c} \sum_{\gamma=1}^{c} H(Y_\gamma | Z_\gamma) \ .$$

Proof: Let X, W and $\mathsf{T}_{i,j}$ be shorthand for $X^{(M)}$, $W^{(M)}$ and $\mathsf{T}_{i,j}^{(\prec)}$, respectively. First note that

$$Y_\gamma = W[\Upsilon_\gamma] \ \text{ and } \ Z_\gamma = W[\Psi_\gamma] \ .$$

We show that

$$\lim_{M\to\infty} \frac{1}{M^2} H(W) \le \liminf_{M\to\infty} \frac{1}{c} \sum_{\gamma=1}^{c} H(Y_\gamma | Z_\gamma) \ .$$

Once this is proved, the claim follows from (4).

By the chain rule [9, Theorem 2.5.1], we have

$$H(W) = \sum_{(i,j)\in \mathsf{B}_M} H(W_{i,j} | W[\mathsf{T}_{i,j} \cap \mathsf{B}_M]) \ .$$

We now recall (6) and define the index set $\bar{\partial}$ to be the largest subset of B_M for which the following condition holds: for all $(i,j) \in \bar{\partial}$, we have that

$$\sigma_{i',j'}(\Psi_\gamma) \subseteq \mathsf{B}_M \ , \tag{7}$$

where hereafter in the proof, $\gamma = f(i,j)$, $i' = i - a_\gamma$, and $j' = j - b_\gamma$. Define $\partial = \mathsf{B}_M \setminus \bar{\partial}$. Note that since r and s are constant, and $\Psi_1, \Psi_2, \ldots, \Psi_c \subseteq \Lambda$, then

$$\frac{|\partial|}{M^2} = O(1/M) \ .$$

Thus, on the one hand, we have

$$\frac{1}{M^2} \sum_{(i,j)\in\partial} H(W_{i,j} | W[\mathsf{T}_{i,j} \cap \mathsf{B}_M]) \le \log_2 |\Sigma| \cdot O(1/M) \ .$$

On the other hand, from (6) and (7) we have for all $(i,j) \in \bar{\partial}$,

$$\sigma_{i',j'}(\Psi_\gamma) \subseteq \mathsf{T}_{i,j} \cap \mathsf{B}_M \ .$$

Hence, since conditioning reduces entropy [9, Theorem 2.6.5],

$$\begin{aligned}
&\frac{1}{M^2} \sum_{(i,j)\in\bar{\partial}} H(W_{i,j} | W[\mathsf{T}_{i,j} \cap \mathsf{B}_M]) \\
&\le \frac{1}{M^2} \sum_{(i,j)\in\bar{\partial}} H(W_{i,j} | W[\sigma_{i',j'}(\Psi_\gamma)]) \\
&= \frac{1}{M^2} \sum_{(i,j)\in\bar{\partial}} H(W[\{(i,j)\} \cup \sigma_{i',j'}(\Psi_\gamma)] | W[\sigma_{i',j'}(\Psi_\gamma)]) \\
&= \frac{1}{M^2} \sum_{(i,j)\in\bar{\partial}} H(Y_\gamma | Z_\gamma) \ ,
\end{aligned}$$

where the last step follows from the stationarity of $W^{(M)}$. Recalling (5), the proof follows. ■

The following is a simple corollary of Lemma 2.

Corollary 3: Let $(W^{(M)})_{M=1}^{\infty}$ be as in Theorem 1 and define

$$X^{(M)} = W^{(M)}[\Lambda] \ .$$

Fix positive integers r and s. Let ℓ be a positive integer, and let $(\rho^{\langle k\rangle})_{k=1}^{\ell}$ be non-negative reals such that $\sum_{k=1}^{\ell} \rho^{\langle k\rangle} = 1$. For every $1 \le k \le \ell$, let $\prec^{\langle k\rangle}$, $c = c^{\langle k\rangle}$, $f^{\langle k\rangle}$, $(\Psi_\gamma^{\langle k\rangle})_{\gamma=1}^{c}$, and $(a_\gamma^{\langle k\rangle}, b_\gamma^{\langle k\rangle})_{\gamma=1}^{c}$ be given. Also, for $1 \le \gamma \le c^{\langle k\rangle}$, let

$$\Upsilon_\gamma^{\langle k\rangle} = \{(a_\gamma^{\langle k\rangle}, b_\gamma^{\langle k\rangle})\} \cup \Psi_\gamma^{\langle k\rangle} \ .$$

Define

$$Y_\gamma^{\langle k\rangle} = X^{(M)}[\Upsilon_\gamma^{\langle k\rangle}] \ \text{ and } \ Z_\gamma^{\langle k\rangle} = X^{(M)}[\Psi_\gamma^{\langle k\rangle}]$$

(note that $Y_\gamma^{\langle k\rangle}$ and $Z_\gamma^{\langle k\rangle}$ are functions of M). Then,

$$\mathsf{cap}(\mathbb{S}) \le \liminf_{M\to\infty} \sum_{k=1}^{\ell} \frac{\rho^{\langle k\rangle}}{c^{\langle k\rangle}} \sum_{\gamma=1}^{c^{\langle k\rangle}} H(Y_\gamma^{\langle k\rangle} | Z_\gamma^{\langle k\rangle}) \ .$$

Corollary 3 is the most general way we have found to state our results. This generality will indeed help us later on. However, almost none of the intuition is lost if the reader has in mind the much simpler case of

$$\begin{gathered}
\ell = 1 \ , \quad \rho^{\langle 1\rangle} = 1 \ , \quad c^{\langle 1\rangle} = 1 \ , \quad \prec^{\langle 1\rangle} = \prec_{\text{lex}} \ , \\
(a_1^{\langle 1\rangle}, b_1^{\langle 1\rangle}) = (r-1, t) \ , \quad \text{and} \quad \Psi_1^{\langle 1\rangle} = \Lambda \cap \mathsf{T}_{(a_1^{\langle 1\rangle}, b_1^{\langle 1\rangle})} \ ,
\end{gathered} \tag{8}$$

where $0 \le t < s$. This simpler case was demonstrated in Running Example I.

IV. Linear requirements

Recall that $X^{(M)} = W^{(M)}[\Lambda]$ is an $r \times s$ sub-configuration of $W^{(M)}$, and thus stationary as well. In this section, we formulate a set of linear requirements (equalities and inequalities) on the probability distribution of $X^{(M)}$. For the rest of this section, let M be fixed and let X be shorthand for $X^{(M)}$.

A. Linear requirements from stationarity

In this subsection, we formulate a set of linear requirements that follow from the stationarity of $X^{(M)}$. Let $x \in \mathbb{S}[\Lambda]$ be a realization of X. Denote

$$p_x = \text{Prob}(X = x) \ .$$

We start with the trivial requirements. Obviously, we must have for all $x \in \mathbb{S}[\Lambda]$ that

$$p_x \ge 0 \ .$$

Also,

$$\sum_{x\in\mathbb{S}[\Lambda]} p_x = 1 \ .$$

Next, we show how we can use stationarity to get more linear equations on $(p_x)_{x\in\mathbb{S}[\Lambda]}$. Let

$$\Lambda' = \{(i,j) : 0 \le i < r-1 \ , \ \ 0 \le j < s\} \ .$$

For $x' \in \mathbb{S}[\Lambda']$ we must have by stationarity that

$$\text{Prob}(X[\Lambda'] = x') = \text{Prob}(X[\sigma_{1,0}(\Lambda')] = \sigma_{1,0}(x')) \ . \tag{9}$$

As a concrete example, suppose that $r = s = 3$. We claim that

$$\text{Prob}\left(X = \begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ * & * & * \end{smallmatrix}\right) = \text{Prob}\left(X = \begin{smallmatrix} * & * & * \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right) \ ,$$

where $*$ denotes "don't care."

Both the left-hand and right-hand sides of (9) are marginalizations of $(p_x)_x$. Thus, we get a set of linear equations on $(p_x)_x$, namely, for all $x' \in \mathbb{S}[\Lambda']$,

$$\sum_{x : x[\Lambda'] = x'} p_x = \sum_{x : x[\sigma_{1,0}(\Lambda')] = \sigma_{1,0}(x')} p_x \ .$$

To get more equations, we now apply the same rationale horizontally, instead of vertically. Let

$$\Lambda'' = \{(i,j) : 0 \le i < r \ , \ \ 0 \le j < s-1\} \ .$$

for all $x'' \in \mathbb{S}[\Lambda'']$,

$$\sum_{x \,:\, x[\Lambda'']=x''} p_x = \sum_{x \,:\, x[\sigma_{0,1}(\Lambda'')]=\sigma_{0,1}(x'')} p_x \ .$$

B. Linear equations from reflection, transposition, and complementation

We now show that if $\mathbb{S}$ is reflection, transposition, or complementation invariant (defined below), then we can derive yet more linear equations.

Define $v_M(\cdot)$ ($h_M(\cdot)$) as the vertical (horizontal) reflection of a rectangular configuration with M rows (columns). Namely,

$$(v_M(w))_{i,j} = w_{M-1-i,j} \ , \quad \text{and} \quad (h_M(w))_{i,j} = w_{i,M-1-j} \ .$$

Define τ as the transposition of a configuration. Namely,

$$\tau(w)_{i,j} = w_{j,i} \ .$$

For $\Sigma = \{0,1\}$, denote by $\mathrm{comp}(w)$ the bitwise complement of a configuration w. Namely,

$$\mathrm{comp}(w)_{i,j} = \begin{cases} 1 & \text{if } w_{i,j} = 0 \\ 0 & \text{otherwise} \ . \end{cases}$$

We state three similar lemmas, and prove the first. The proofs of the other two is similar.

Lemma 4: Suppose that $\mathbb{S}$ is such that for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in \mathbb{S} \iff h_M(w) \in \mathbb{S} \iff v_M(w) \in \mathbb{S} \ .$$

Then, w.l.o.g., the probability distribution of W is such that for all $w \in \mathbb{S}_M$,

$$\begin{aligned} &\mathrm{Prob}(W = w) = \\ &\qquad \mathrm{Prob}(W = h_M(w)) = \mathrm{Prob}(W = v_M(w)) \ . \quad (10) \end{aligned}$$

Lemma 5: Suppose that $\mathbb{S}$ is such that for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in \mathbb{S} \iff \tau(w) \in \mathbb{S} \ .$$

Then, w.l.o.g., W is such that for all $w \in \mathbb{S}_M$,

$$\mathrm{Prob}(W = w) = \mathrm{Prob}(W = \tau(w)) \ . \qquad (11)$$

Lemma 6: Suppose that $\Sigma = \{0,1\}$ and $\mathbb{S}$ is such that for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in \mathbb{S} \iff \mathrm{comp}(w) \in \mathbb{S} \ .$$

Then, w.l.o.g., W is such that for all $w \in \mathbb{S}_M$,

$$\mathrm{Prob}(W = w) = \mathrm{Prob}(W = \mathrm{comp}(w)) \ . \qquad (12)$$

Proof of Lemma 4: Let h and v be shorthand for h_M and v_M, respectively. For M fixed, we define a new random variable W^{new} taking values on $\mathbb{S}_M$, with the following distribution: for all $w \in \mathbb{S}_M$,

$$\mathrm{Prob}(W^{\mathrm{new}}{=}w) = \frac{1}{4} \sum_{\substack{w' \in \\ \{w, h(w), v(w), h(v(w))\}}} \mathrm{Prob}(W{=}w') \ .$$

Since $h(h(w)) = v(v(w)) = w$ and $h(v(w)) = v(h(w))$ we get that (10) holds for W^{new}. Moreover, by the concavity of the entropy function,

$$H(W) \le H(W^{\mathrm{new}}) \ .$$

Also, it is easily proved that

$$H(W^{\mathrm{new}}) \le H(W) + 2$$

(to see this, define an auxiliary random variable indicating which one of the 4 equally likely reflections was "chosen"). Thus, the properties defined in Theorem 1 hold for W^{new}. ∎

If the condition of Lemma 4 holds, then we get the following equations by stationarity. For all $x \in \mathbb{S}[\Lambda]$,

$$p_x = p_{v_r(x)} = p_{h_s(x)} \ .$$

If the condition of Lemma 5 holds, then we have the following. Assume w.l.o.g. that $r \le s$, and let

$$\tilde{\Lambda} = \{(i,j) : 0 \le i, j < r\} \ .$$

For all $\chi \in \mathbb{S}[\tilde{\Lambda}]$,

$$\sum_{x \,:\, x[\tilde{\Lambda}]=\chi} p_x = \sum_{x \,:\, x[\tilde{\Lambda}]=\tau(\chi)} p_x \ .$$

If the condition of Lemma 6 holds, then we have the following. For all $x \in \mathbb{S}[\Lambda]$,

$$p_x = p_{\mathrm{comp}(x)} \ .$$

V. An upper bound on $\mathsf{cap}(\mathbb{S})$

For the rest of this section, let r, s, ℓ, $\rho^{\langle k \rangle}$, $\prec^{\langle k \rangle}$, $c^{\langle k \rangle}$, $f^{\langle k \rangle}$, $\Psi^{\langle k \rangle}_\gamma$, and $(a^{\langle k \rangle}_\gamma, b^{\langle k \rangle}_\gamma)$ be given as in Corollary 3. Recall from Corollary 3 that we are interested in $H(Y^{\langle k \rangle}_\gamma | Z^{\langle k \rangle}_\gamma)$, in order to bound $\mathsf{cap}(\mathbb{S})$ from above.

As a first step, we fix M and express $H(Y^{\langle k \rangle}_\gamma | Z^{\langle k \rangle}_\gamma)$ in terms of the probabilities $(p_x)_x$ of the random variable $X^{(M)}$. For given $1 \le k \le \ell$ and $1 \le \gamma \le c^{\langle k \rangle}$, let

$$y \in \mathbb{S}[\Upsilon^{\langle k \rangle}_\gamma] \ \text{ and } \ z \in \mathbb{S}[\Psi^{\langle k \rangle}_\gamma]$$

be realizations of $Y^{\langle k \rangle}_\gamma$ and $Z^{\langle k \rangle}_\gamma$, respectively. Let

$$p^{\langle k \rangle}_{\gamma,y} = \mathrm{Prob}(Y^{\langle k \rangle}_\gamma = y) \quad \text{and} \quad p^{\langle k \rangle}_{\gamma,z} = \mathrm{Prob}(Z^{\langle k \rangle}_\gamma = z)$$

($p^{\langle k \rangle}_{\gamma,y}$ and $p^{\langle k \rangle}_{\gamma,z}$ are functions of M). From here onward, let p_y and p_z be shorthand for $p^{\langle k \rangle}_{\gamma,y}$ and $p^{\langle k \rangle}_{\gamma,z}$, respectively. Both p_y and p_z are marginalizations of $(p_x)_x$, namely,

$$p_y = \sum_{x \in \mathbb{S}[\Lambda] \,:\, x[\Upsilon^{\langle k \rangle}_\gamma]=y} p_x \ , \quad p_z = \sum_{x \in \mathbb{S}[\Lambda] \,:\, x[\Psi^{\langle k \rangle}_\gamma]=z} p_x \ .$$

Thus, for given γ and k,

$$H(Y_\gamma^{\langle k\rangle}|Z_\gamma^{\langle k\rangle}) = \sum_{y\in\mathbb{S}[\Upsilon_\gamma^{\langle k\rangle}]} -p_y \log_2 p_y + \sum_{z\in\mathbb{S}[\Psi_\gamma^{\langle k\rangle}]} p_z \log_2 p_z$$

is a function of the probabilities $(p_x)_x$ of $X^{(M)}$.

Our next step will be to reason as follows: We have found linear requirements that the p_x's satisfy and expressed $H(Y_\gamma^{\langle k\rangle}|Z_\gamma^{\langle k\rangle})$ as a function of $(p_x)_x$. However, *we do not know of a way to actually calculate* $(p_x)_x$. So, instead of the probabilities $(p_x)_x$, consider the *variables* $(\bar{p}_x)_x$. From this line of thought we get our main theorem.

Theorem 7: The value of the optimization program given in Figure 3 is an upper bound on $\mathsf{cap}(\mathbb{S})$.

Proof: First, notice that if we take $\bar{p}_x = p_x$, then (by Section IV) all the requirements which the $\bar{p}_x$'s are subject to indeed hold, and the objective function is equal to

$$\sum_{k=1}^{\ell} \frac{\rho^{\langle k\rangle}}{c^{\langle k\rangle}} \sum_{\gamma=1}^{c^{\langle k\rangle}} H(Y_\gamma^{\langle k\rangle}|Z_\gamma^{\langle k\rangle}) \; .$$

So, the maximum is an upper bound on the above equation. Next, by compactness, a maximum indeed exists. Since the maximum is not a function of M, the claim now follows from Corollary 3. ∎

We now proceed to show that the optimization problem in Figure 3 is an instance of convex programming[2] [10, p. 137, Equation (4.16)], and thus easily calculated. Since the requirements that the variables $(\bar{p}_x)_x$ are subject to are linear, this reduces to showing that the objective function is concave in $(\bar{p}_x)_x$.

Lemma 8: The objective function in Figure 3 is concave in the variables $(\bar{p}_x)_{x\in\mathbb{S}[\Lambda]}$, subject to them being non-negative.

Proof: Recall that for all $1 \le k \le \ell$ we have that $\frac{\rho^{\langle k\rangle}}{c^{\langle k\rangle}}$ is non-negative. Thus, it suffices to prove that for all $1 \le k \le \ell$ and $1 \le \gamma \le c^{\langle k\rangle}$, the function $\Xi(k,\gamma)$ is concave in the variables $(\bar{p}_x)_x$. So, let k and γ be fixed, and let $\bar{p}_y$ and $\bar{p}_z$ be shorthand for $\bar{p}_{\gamma,y}^{\langle k\rangle}$ and $\bar{p}_{\gamma,z}^{\langle k\rangle}$, respectively.

Recalling the definition of $\bar{p}_{\gamma,y}^{\langle k\rangle}$ and $\bar{p}_{\gamma,z}^{\langle k\rangle}$ in Figure 3 and the fact that $\Psi_\gamma^{\langle k\rangle} \subseteq \Upsilon_\gamma^{\langle k\rangle}$, we get that

$$\Xi(k,\gamma) = \sum_{\substack{y\in\mathbb{S}[\Upsilon_\gamma^{\langle k\rangle}] \\ z=y[\Psi_\gamma^{\langle k\rangle}]}} -\bar{p}_y \log_2 \frac{\bar{p}_y}{\bar{p}_z} \; .$$

Thus, it suffices to show that each summand is concave in $(\bar{p}_x)_x$. This is indeed the case: let $(\bar{p}_x^{(1)})_{x\in\mathbb{S}[\Lambda]}$ and $(\bar{p}_x^{(2)})_{x\in\mathbb{S}[\Lambda]}$ be non-negative. Let $0 \le \xi \le 1$ be given, and define $(\bar{p}_x^{(3)})_{x\in\mathbb{S}[\Lambda]}$ as

$$\bar{p}_x^{(3)} = \xi\bar{p}_x^{(1)} + (1-\xi)\bar{p}_x^{(2)} \; , \quad x\in\mathbb{S}[\Lambda] \; .$$

For $t = 1,2,3$, denote by $\bar{p}_y^{(t)}$ and $\bar{p}_z^{(t)}$ the marginalizations corresponding to $(\bar{p}_x^{(t)})_x$. Obviously,

$$\bar{p}_y^{(3)} = \xi\bar{p}_y^{(1)} + (1-\xi)\bar{p}_y^{(2)} \; , \quad y\in\mathbb{S}[\Upsilon_\gamma^{\langle k\rangle}] \; .$$

[2]Although the objective function is *concave*, the standard terminology seems to be *convex* programming

---

maximize

$$\sum_{k=1}^{\ell} \frac{\rho^{\langle k\rangle}}{c^{\langle k\rangle}} \sum_{\gamma=1}^{c^{\langle k\rangle}} \Xi(k,\gamma)$$

over the variables $(\bar{p}_x)_{x\in\mathbb{S}[\Lambda]}$, where for

$$1 \le k \le \ell \; , \quad 1 \le \gamma \le c^{\langle k\rangle} \; , \quad y\in\mathbb{S}[\Upsilon_\gamma^{\langle k\rangle}] \; , \quad z\in\mathbb{S}[\Psi_\gamma^{\langle k\rangle}] \; ,$$

we define

$$\bar{p}_{\gamma,y}^{\langle k\rangle} \triangleq \sum_{x\in\mathbb{S}[\Lambda]\,:\,x[\Upsilon_\gamma^{\langle k\rangle}]=y} \bar{p}_x \; , \quad \bar{p}_{\gamma,z}^{\langle k\rangle} \triangleq \sum_{x\in\mathbb{S}[\Lambda]\,:\,x[\Psi_\gamma^{\langle k\rangle}]=z} \bar{p}_x \; ,$$

$$\Xi(k,\gamma) \triangleq -\sum_{y\in\mathbb{S}[\Upsilon_\gamma^{\langle k\rangle}]} \bar{p}_{\gamma,y}^{\langle k\rangle} \log_2 \bar{p}_{\gamma,y}^{\langle k\rangle} + \sum_{z\in\mathbb{S}[\Psi_\gamma^{\langle k\rangle}]} \bar{p}_{\gamma,z}^{\langle k\rangle} \log_2 \bar{p}_{\gamma,z}^{\langle k\rangle} \; ,$$

and the variables $\bar{p}_x$ are subject to the following requirements:

(i)
$$\sum_{x\in\mathbb{S}[\Lambda]} \bar{p}_x = 1 \; .$$

(ii) For all $x\in\mathbb{S}[\Lambda]$,
$$\bar{p}_x \ge 0 \; .$$

(iii) For all $x'\in\mathbb{S}[\Lambda']$,
$$\sum_{x\,:\,x[\Lambda']=x'} \bar{p}_x = \sum_{x\,:\,x[\sigma_{1,0}(\Lambda')]=\sigma_{1,0}(x')} \bar{p}_x \; .$$

(iv) For all $x''\in\mathbb{S}[\Lambda'']$,
$$\sum_{x\,:\,x[\Lambda'']=x''} \bar{p}_x = \sum_{x\,:\,x[\sigma_{0,1}(\Lambda'')]=\sigma_{0,1}(x'')} \bar{p}_x \; .$$

(v) (If $\mathbb{S}$ is reflection (resp. complementation) invariant) For all $x\in\mathbb{S}[\Lambda]$,
$$\bar{p}_x = \bar{p}_{v_r(x)} = \bar{p}_{h_s(x)} \quad (\text{resp. } \bar{p}_x = \bar{p}_{\mathrm{comp}(x)}) \; .$$

(vi) (If $\mathbb{S}$ is transposition invariant) For all $\chi\in\mathbb{S}[\tilde{\Lambda}]$,
$$\sum_{x\,:\,x[\tilde{\Lambda}]=\chi} \bar{p}_x = \sum_{x\,:\,x[\tilde{\Lambda}]=\tau(\chi)} \bar{p}_x \; .$$

Fig. 3. Optimization program over the variables $\bar{p}_x$ (assuming w.l.o.g. that $r \le s$). The optimum is an upper bound on $\mathsf{cap}(\mathbb{S})$.

and
$$\bar{p}_z^{(3)} = \xi\bar{p}_z^{(1)} + (1-\xi)\bar{p}_z^{(2)} \; , \quad z\in\mathbb{S}[\Psi_\gamma^{\langle k\rangle}] \; .$$

We must show that for all $y\in\mathbb{S}[\Upsilon_\gamma^{\langle k\rangle}]$, $z = y[\Psi_\gamma^{\langle k\rangle}]$

$$\bar{p}_y^{(3)} \log_2 \frac{\bar{p}_y^{(3)}}{\bar{p}_z^{(3)}} \le \xi\bar{p}_y^{(1)} \log_2 \frac{\bar{p}_y^{(1)}}{\bar{p}_z^{(1)}} + (1-\xi)\bar{p}_y^{(2)} \log_2 \frac{\bar{p}_y^{(2)}}{\bar{p}_z^{(2)}} \; .$$

This is indeed the case, by the log sum inequality [9, p. 29]. ∎

We make the following observations in passing. Suppose that all the parameters that define the optimization problem in Figure 3 have already been set, apart from the values of $\rho^{\langle k\rangle}$, $1 \le k \le \ell$. Of course, we would like to set $(\rho^{\langle k\rangle})_{k=1}^{\ell}$

such that our bound is as tight as possible. Namely, we would like to find the value of $(\rho^{\langle k\rangle})_{k=1}^{\ell}$ which minimizes the value computed in Figure 3. By [10, Section 3.2.3], this is a convex optimization problem. In fact, more is true: By [10, Exercise 5.25], instead of searching for the $(\rho^{\langle k\rangle})_{k=1}^{\ell}$ that minimizes the maximization over all valid probability distributions (min-max), we could instead — for the sake of argument — switch the optimization order (max-min).

As can be seen in Appendix B, there are cases in which it is advantageous to have $(\rho^{\langle k\rangle})_{k=1}^{\ell}$ contain more than one non-zero entry. Intuitively, this stems from the following: Consider two indices $1 \leq k_1 < k_2 \leq \ell$, and assume for simplicity of notation that $c^{\langle k_1\rangle} = c^{\langle k_2\rangle} = 1$. Suppose that we were to take $\rho^{\langle k_1\rangle}$ equal to 1, and all other entries of ρ equal to 0. In that case, by definition, the probability distribution $(\bar{p}_x)_{x\in\mathbb{S}[\Lambda]}$ calculated in Figure 3 would be the one that maximizes $\Xi(k_1, 1)$. Intuitively speaking, it might be the case that the calculated probability distribution $(\bar{p}_x)_{x\in\mathbb{S}[\Lambda]}$ manages to give the measured entropy function $\Xi(k_1, 1)$ the largest possible value, by giving the unmeasured entropy function $\Xi(k_2, 1)$ an unrealistically small value. Indeed, we might very well have $\Xi(k_2, 1) < \mathsf{cap}(\mathbb{S})$ (compare with Lemma 2). This anomaly can be fixed by properly representing both $\Xi(k_1, 1)$ and $\Xi(k_2, 1)$ in the objective function. That is, by giving both $\rho^{\langle k_1\rangle}$ and $\rho^{\langle k_2\rangle}$ sufficiently large values.

VI. Computational results

At this point, we have formulated a concave optimization problem, and wish to solve it. There are quite a few programs, termed *solvers*, that enable one to do so. Many such solvers — most of them proprietary — are hosted on the servers of the NEOS project [11][12][13], and the public may submit moderately sized optimization problems to them. We have coded our optimization problems in the AMPL modeling language [14], and submitted them to NEOS.

Essentially, a solver starts with some initial guess as to the optimizing value of $(\bar{p}_x)_{x\in\mathbb{S}[\Lambda]}$, and then iteratively improves the value of the objective function. This process is terminated when the solver decides that it is "close enough" to the optimum. Denote by $\widetilde{\boldsymbol{p}} = (\widetilde{p}_x)_{x\in\mathbb{S}[\Lambda]}$ this "close enough" assignment to the variables. Of course, we must supply an upper bound on $\mathsf{cap}(\mathbb{S})$, not an approximation to one. Thus, let $\widetilde{f}$ and

$$\widetilde{\boldsymbol{g}} = (\widetilde{g}_x)_x \; , \quad x \in \mathbb{S}[\Lambda] \; ,$$

be the value of the objective function and its gradient at $\widetilde{\boldsymbol{p}}$, respectively. Obviously, $\widetilde{f}$ is a lower bound on the value of our optimization problem. For an upper bound[3], we replace the objective function in Figure 3 by

$$\text{maximize} \quad \left(\widetilde{f} + \sum_{x\in\mathbb{S}[\Lambda]} \widetilde{g}_x \cdot (\bar{p}_x - \widetilde{p}_x) \right) \; ,$$

[3]We remark in passing that if we had chosen to optimize the *dual problem* [10, p. 215], then the "dual of" $\widetilde{f}$ would already have been an upper bound. However, we have not managed to state the dual problem in closed form.

1	5	2	6	3	7	4
8	12	9	13	10	14	11
15	19	16	20	17	21	18

$\prec_{\text{skip}}$

Fig. 4. An entry labeled i in the configuration precedes an entry labeled j according to $\prec_{\text{skip}}$ iff $i < j$.

and get a *linear program* (the value of which can be calculated exactly). By concavity, the value of this linear program is indeed an upper bound. So, we use NEOS yet again to solve it. For the sake of double-checking, we submitted the above optimization problems to two solvers: IPOPT [15] and MOSEK.

Before stating our computational results, let us first define one more strict total order, which we have termed the "skip" order, $\prec_{\text{skip}}$ (see Figure 4). We have that $(i_1, j_1) \prec_{\text{skip}} (i_2, j_2)$ iff

- $i_1 < i_2$, or
- ($i_1 = i_2$ and $j_1 \equiv 0 \pmod 2$ and $j_2 \equiv 1 \pmod 2$), or
- ($i_1 = i_2$ and $j_1 \equiv j_2 \pmod 2$ and $j_1 < j_2$)

Our computational results appear in Table I. To the best of our knowledge, they are presently the tightest. The penultimate column contains upper bounds obtained by the method described in [6]. When available, these compared-to bounds are taken from previously published work, as indicated to the right of them. The rest are the result of our implementation of [6]. For reference, the last column contains corresponding lower bounds. We note that the indexes $(a_\gamma^{\langle k\rangle}, b_\gamma^{\langle k\rangle})$ and coefficients $\rho^{\langle k\rangle}$ used for each constraint were optimized by hand, through trial and error. Also, we note that when applying our method to the 2-D $(1,\infty)$-RLL constraint, our bound was inferior to the one presented in [2] (utilizing the method of [7]). In order to make the results in Table I reproducible, we completely specify how they were obtained in Appendix B.

VII. Asymptotic analysis

For a given constraint $\mathbb{S}$ and positive integers r and s, let t be an integer such that $0 \leq t < s$. Denote by $\mu(r, s, t)$ the value of the optimization program in Figure 3, where the parameters are as in (8). In this section, we show that even if we restrict ourselves to this simple case, we get an upper bound which is asymptotically tight, in the following sense.

Theorem 9: For all $\epsilon > 0$, there exist

$$r_0 > 0 \; , \quad s_0 > 0 \; , \quad 0 \leq t_0 < s_0$$

such that for all

$$r \geq r_0 \; , \quad s \geq s_0 \; , \quad t_0 \leq t \leq s - (s_0 - t_0) \; ,$$

we have that

$$\mu(r, s, t) - \mathsf{cap}(\mathbb{S}) \leq \epsilon \; .$$

In order to prove Theorem 9, we need the following lemma.

TABLE I
UPPER-BOUNDS ON THE CAPACITY OF SOME 2-D CONSTRAINTS.

Constraint	r	s	ℓ	$\prec$ used	Upper bound	Comparison	Lower bound
$(2,\infty)$-RLL	3	8	7	$\prec_{\rm lex}$, $\prec_{\rm skip}$	0.4457	0.4459 [16]	0.444202 [17]
$(3,\infty)$-RLL	4	8	5	$\prec_{\rm lex}$	0.36821	0.3686 [16]	0.365623 [18]
$(0,2)$-RLL	3	5	2	$\prec_{\rm lex}$	0.816731	0.817053	0.816007 [18]
n.i.b.	3	4	1	$\prec_{\rm skip}$	0.92472	0.927855	0.922640 [17]

Lemma 10: For all $\epsilon > 0$, there exist

$$r_0 > 0 \ , \quad s_0 > 0 \ , \quad 0 \le t_0 < s_0$$

such that

$$\mu(r_0, s_0, t_0) - \mathsf{cap}(\mathbb{S}) < \epsilon \ .$$

Proof: Another well known method for bounding $\mathsf{cap}(\mathbb{S})$ from above is the so called "stripe method," mentioned in the introduction. Namely, for some given θ, consider the 1-D constraint $S = S(\theta)$ defined as follows. The alphabet of the constraint is Σ^θ. A word of length r' satisfies S if and only if when we write its entries as rows of length θ, one below the other, we get an $r' \times \theta$ configuration which satisfies the 2-D constraint $\mathbb{S}$.

Define the normalized capacity of S as

$$\widehat{\mathsf{cap}}(S) = \frac{1}{\theta}\mathsf{cap}(S) \ .$$

By the definition of $\mathsf{cap}(\mathbb{S})$, the normalized capacity approaches $\mathsf{cap}(\mathbb{S})$ as $\theta \to \infty$. Thus, fix a θ such that

$$\widehat{\mathsf{cap}}(S) - \mathsf{cap}(\mathbb{S}) \le \epsilon/2 \ .$$

We say that a 1-D constraint has memory m if there exists a graph representing it, and all paths in the graph of length m with the same labeling terminate in the same vertex. By [1, Theorem 3.17] and its proof, there exists a series of 1-D constraints $\{S_\mathsf{m}\}_{\mathsf{m}=1}^\infty$ such that $S \subseteq S_\mathsf{m}$, the memory of S_m is m, and $\lim_{\mathsf{m}\to\infty} \mathsf{cap}(S_\mathsf{m}) = \mathsf{cap}(S)$. Thus, fix m such that

$$\widehat{\mathsf{cap}}(S_\mathsf{m}) - \widehat{\mathsf{cap}}(S) \le \epsilon/2 \ .$$

To finish the proof, we now show that

$$\mu(r_0, s_0, t_0) \le \widehat{\mathsf{cap}}(S_\mathsf{m}) \ ,$$

where

$$r_0 = \mathsf{m} + 1 \ , \quad s_0 = 2 \cdot \theta \ , \quad t_0 = \theta - 1 \ .$$

Note that $\mu(r_0, s_0, t_0)$ is the maximum of

$$H(\bar{X}_{\mathsf{m},\theta-1} | \bar{X}[\mathsf{T}^{(\prec_{\rm lex})}_{\mathsf{m},\theta-1} \cap \mathsf{B}_{\mathsf{m}+1,2\cdot\theta}]) \ , \tag{13}$$

over all random variables $\bar{X} \in \mathbb{S}_{\mathsf{m}+1,2\cdot\theta}$ with a probability distribution satisfying our linear requirements.

For all $0 \le \phi < \theta$ we get by the (imposed) stationarity of $\bar{X}$ that (13) is bounded from above by

$$H_\phi = H(\bar{X}_{\mathsf{m},\phi} | \bar{X}[\mathsf{T}^{(\prec_{\rm lex})}_{\mathsf{m},\phi} \cap \mathsf{B}_{\mathsf{m}+1,\theta}]) \ .$$

So, (13) is also bounded from above by

$$\frac{1}{\theta} \sum_{\phi=0}^{\theta-1} H_\phi \ . \tag{14}$$

The first θ columns of $\bar{X}$ form a configuration with index set $\mathsf{B}_{\mathsf{m}+1,\theta}$. By our linear requirements, stationarity (specifically, vertical stationarity) holds for this configuration as well. So, we may define a stationary 1-D Markov chain [1, §3.2.3] on S_m, with entropy given by (14). That entropy, in turn, is at most $\widehat{\mathsf{cap}}(S_\mathsf{m})$. ■

Proof of Theorem 9: The following inequalities are easily verified:

$$\mu(r, s, t) \ge \mu(r+1, s, t) \ .$$

$$\mu(r, s, t) \ge \mu(r, s+1, t) \ .$$

$$\mu(r, s, t) \ge \mu(r, s+1, t+1) \ .$$

The proof follows from them and Lemma 10. ■

VIII. ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for helpful comments.

APPENDIX A
PROOF OF THEOREM 1

Our goal in this appendix is to prove Theorem 1. Essentially, Theorem 1 will turn out to be a corollary of [8, Theorem 1.4]. However, [8, Theorem 1.4] deals with configurations in which the index set is $\mathbb{Z}^2$. So, some definitions and auxiliary lemmas are in order.

Recall that $(G_{\rm row}, G_{\rm col})$ is the pair of vertex-labeled graphs through which $\mathbb{S} = \mathbb{S}(G_{\rm row}, G_{\rm col})$ is defined. Also, recall that each member of $\mathbb{S}$ is a configuration with a rectangular index set. Namely, the index set of a configuration in $\mathbb{S}$ is $\sigma_{i,j}(\mathsf{B}_{M,N})$, for some i, j, M, and N. We now give a very similar definition to that of $\mathbb{S}$, only now we require that the index set of each configuration is $\mathbb{Z}^2$. Namely, define $\mathcal{S} = \mathcal{S}(G_{\rm row}, G_{\rm col})$ as follows: A configuration $(w_{i,j})_{(i,j)\in\mathbb{Z}^2}$ over Σ is in $\mathcal{S}(G_{\rm row}, G_{\rm col})$ iff there exists a configuration $(u_{i,j})_{(i,j)\in\mathbb{Z}^2}$ over the vertex set V with the following properties: for all $(i,j) \in \mathbb{Z}^2$, (a) the labeling of $u_{i,j}$ satisfies $L(u_{i,j}) = w_{i,j}$; (b) there exists an edge from $u_{i,j}$ to $u_{i,j+1}$ in $G_{\rm row}$; (c) there exists an edge from $u_{i,j}$ to $u_{i+1,j}$ in $G_{\rm col}$.

For positive integers $M, N > 0$, define $\mathcal{S}_{M,N}$ as the restriction of $\mathcal{S}$ to $\mathsf{B}_{M,N}$. Namely,

$$\mathcal{S}_{M,N} = \mathcal{S}[\mathsf{B}_{M,N}] \ ,$$

where the definition of the restriction operation is as in (3). Also, for M equal to N, define

$$\mathcal{S}_M = \mathcal{S}_{M,M} \ .$$

Note that for all $M, N > 0$ we have

$$\mathcal{S}_{M,N} \subseteq \mathbb{S}_{M,N} \ , \tag{15}$$

and there are cases in which the inclusion is strict. Next, define the capacity of $\mathcal{S}$ as

$$\mathsf{cap}(\mathcal{S}) = \lim_{M \to \infty} \frac{1}{M^2} \cdot \log_2 |\mathcal{S}_M| \ .$$

The limit indeed exists, by sub-additivity (see [3, Appendix], and references therein).

For integers $M, N > 0$ and $\delta \geq 0$, denote

$$\mathsf{C}_{M,N,\delta} = \sigma_{-\delta,-\delta}(\mathsf{B}_{M+2\delta,N+2\delta})$$

and let

$$\mathbb{S}_{M,N,\delta} = \mathbb{S}[\mathsf{C}_{M,N,\delta}] \ .$$

Note that the index set $\mathsf{C}_{M,N,\delta}$ of each element of $\mathbb{S}_{M,N,\delta}$ is simply $\mathsf{B}_{M,N}$, padded with δ columns to the right and left and δ rows to the top and bottom. The following lemma will help us bridge the gap between finite and infinite index sets. It is essentially a restatement of [19, Lemma 2.6].

Lemma 11: Let w be a configuration over the finite alphabet Σ with index set $\mathsf{B}_{M,N}$. If for all $\delta \geq 0$ we have that

$$w \in \mathbb{S}_{M,N,\delta}[\mathsf{B}_{M,N}] \ , \tag{16}$$

then we must have that

$$w \in \mathcal{S}_{M,N} \ .$$

Proof: Define the following auxiliary directed graph. The vertex set is

$$\bigcup_{\delta \geq 0} \{\hat{w} \in \mathbb{S}_{M,N,\delta} \ : \ \hat{w}[\mathsf{B}_{M,N}] = w\} \ .$$

For every $\delta \geq 0$, there is a directed edge from $w_1 \in \mathbb{S}_{M,N,\delta}$ to $w_2 \in \mathbb{S}_{M,N,\delta+1}$ iff $w_1 = w_2[\mathsf{C}_{M,N,\delta}]$. It is easily seen that this graph is a directed tree with root w, as defined in [20, §2.4]. Since (16) holds for all $\delta \geq 0$, the vertex set of the tree is infinite (and countable). On the other hand, since the alphabet size $|\Sigma|$ is finite, the out-degree of each vertex is finite. Thus, by König's Infinity Lemma [20, Theorem 2.8], we must have an infinite path in the tree starting from the root w.

Denote the vertices of the above-mentioned infinite path as

$$w = w^{[0]}, w^{[1]}, w^{[2]}, \ldots \ .$$

We now show how to find a configuration $(w'_{i,j})_{(i,j)\in\mathbb{Z}^2}$ such that $w' \in \mathcal{S}$ and $w = w'[\mathsf{B}_{M,N}]$. For each $(i,j) \in \mathbb{Z}^2$, define $w'_{i,j}$ as follows: let $\delta \geq 0$ be such that $(i,j) \in \mathsf{C}_{M,N,\delta}$, and take $w'_{i,j} = w^{[\delta]}_{i,j}$. It is easily seen that w' is well defined and contained in $\mathcal{S}$. ■

The following lemma states that although the inclusion in (15) may be strict, the capacities of $\mathbb{S}$ and $\mathcal{S}$ are equal. The lemma is stated in [19] as Theorem 2.5, for the case of finite type constraints. For the sake of completeness, and also since our setting is more general, we give a proof here as well.

Lemma 12: Let $\mathbb{S}$ and $\mathcal{S}$ be as previously defined. Then,

$$\mathsf{cap}(\mathcal{S}) = \mathsf{cap}(\mathbb{S}) \ . \tag{17}$$

Proof: By (15), we must have that $\mathsf{cap}(\mathcal{S}) \leq \mathsf{cap}(\mathbb{S})$. For the other direction, it suffices to prove that for all $M > 0$,

$$\mathsf{cap}(\mathbb{S}) \leq \frac{1}{M^2} \log_2 |\mathcal{S}_M| \ . \tag{18}$$

So, let us fix M and prove the above. By Lemma 11, there exists $\delta \geq 0$ such that for all $w \in \Sigma^{M \times M}$,

$$w \notin \mathcal{S}_M \quad \Longrightarrow \quad w \notin \mathbb{S}_{M,M,\delta}[\mathsf{B}_M] \ .$$

For $t > 0$, let M' be shorthand for

$$M' = t \cdot M \ .$$

By the definition of capacity, we have that

$$\mathsf{cap}(\mathbb{S}) = \lim_{t \to \infty} \frac{1}{(M')^2} \log_2 |\mathbb{S}_{M'}| \ . \tag{19}$$

Now, let us partition $\mathsf{B}_{M'}$ into the following disjoint sub-sets of indexes: for $0 \leq i, j < t$, define the set

$$\mathsf{D}_{i,j} = \sigma_{i \cdot M, j \cdot M}(\mathsf{B}_M) \ .$$

Let $w' \in \mathbb{S}_{M'}$. Notice that for all $0 \leq i, j < t$ for which

$$\sigma_{i \cdot M, j \cdot M}(\mathsf{C}_{M,M,\delta}) \subseteq \mathsf{B}_{M'} \ , \tag{20}$$

we must have that $w'[\mathsf{D}_{i,j}]$ is equal to some correspondingly shifted element of $\mathcal{S}_M$. On the other hand, for M and δ fixed, the number of pairs (i,j) for which (20) does not hold is $O(t)$. Thus, a simple calculation gives us that

$$\frac{1}{(M')^2} \log_2 |\mathbb{S}_{M'}| \leq \frac{1}{M^2} \log_2 |\mathcal{S}_M| + O(1/t) \ .$$

This, together with (19), proves (18). ■

For a given $M > 0$, define the set $\mathcal{F}(M)$ of configurations with index set $\mathbb{Z}^2$ as follows: a configuration $(w_{i,j})_{(i,j)\in\mathbb{Z}^2}$ is in $\mathcal{F}(M)$ iff for all $(i,j) \in \mathbb{Z}^2$,

$$w[\sigma_{i,j}(\mathsf{B}_M)] \in \mathcal{S}_M \ .$$

Namely, each $M \times M$ "patch" is a correspondingly shifted element of $\mathcal{S}_M$.

Note that there exist vertex-labeled graphs $G_{\mathrm{row}}(M)$ and $G_{\mathrm{col}}(M)$ such that $\mathcal{F}(M) = \mathcal{S}(G_{\mathrm{row}}(M), G_{\mathrm{col}}(M))$. Specifically, the vertex set of both graphs is equal to $\mathcal{S}_M$; the label of each such vertex is its lower-left entry; there is an edge from $w_1 \in \mathcal{S}_M$ to $w_2 \in \mathcal{S}_M$ in $G_{\mathrm{row}}(M)$ ($G_{\mathrm{col}}(M)$) iff the first $M-1$ rows (columns) of w_1 are equal to the last $M-1$ (rows) columns of w_2. Thus, $\mathsf{cap}(\mathcal{F}(M))$ exists. Also, since $w \in \mathcal{S}$ implies $w \in \mathcal{F}(M)$, we have

$$\mathsf{cap}(\mathcal{S}) \leq \mathsf{cap}(\mathcal{F}(M)) \ . \tag{21}$$

The following is a direct corollary of [8, Theorem 1.4].

Corollary 13: For all $M > 0$, there exists a stationary random variable $W^{(M)}$ taking values on $\mathcal{F}(M)[\mathsf{B}_M]$ such that

$$\mathsf{cap}(\mathcal{F}(M)) \leq \frac{1}{M^2} H(W^{(M)}) \ . \tag{22}$$

Proof of Theorem 1: Notice that

$$\mathcal{F}(M)[\mathsf{B}_M] = \mathcal{S}_M \subseteq \mathbb{S}_M \ .$$

Thus, take $W^{(M)}$ as in Corollary 13 and notice that it satisfies conditions (i) and (ii) in Theorem 1. From (17), (21), and (22) we get that

$$\mathsf{cap}(\mathbb{S}) \leq \lim_{M \to \infty} \frac{1}{M^2} \cdot H(W^{(M)}) \ .$$

But since $W^{(M)}$ takes values on $\mathbb{S}_M$, we have by [9, p. 19] that the above inequality is in fact an equality. Thus, condition (iii) is proved. ■

Appendix B
Full specification of results in Table I

This appendix is devoted to completely specifying how the results given in the sixth column of Table I were obtained.

Recall that the first four columns of Table I specify the constraint considered, along with the values of r, s, and ℓ used. We will repeat these here, for convenience. Next, for $k = 1, 2, \ldots, \ell$, we must specify $\rho^{\langle k \rangle}$, $\prec^{\langle k \rangle}$, $c = c^{\langle k \rangle}$, $f^{\langle k \rangle}$, and $(a_\gamma^{\langle k \rangle}, b_\gamma^{\langle k \rangle})_{\gamma=1}^{c}$. Note that we have omitted $(\Psi_\gamma^{\langle k \rangle})_{\gamma=1}^{c}$ in the above, since it can be deduced from (6).

We relate the following facts about our calculations, in order to make the specification more concise later on. We stress that the reader may choose to preform calculations in which the following do not hold.

- For a fixed $1 \leq k \leq \ell$, we have that the pair $(a_\gamma^{\langle k \rangle}, b_\gamma^{\langle k \rangle})$ is independent of $1 \leq \gamma \leq c^{\langle k \rangle}$. Hence, even if $c^{\langle k \rangle} > 1$, we will specify only one pair (denoted $(a_*^{\langle k \rangle}, b_*^{\langle k \rangle})$ in the following tables).
- Recall from Table I that the two possible values for $\prec^{\langle k \rangle}$ are $\prec_{\mathrm{lex}}$ and $\prec_{\mathrm{skip}}$. In our case, specifying $\prec^{\langle k \rangle}$ specifies $c^{\langle k \rangle}$ and $f^{\langle k \rangle}$ as well: If $\prec^{\langle k \rangle} = \prec_{\mathrm{lex}}$, then $c^{\langle k \rangle} = 1$ and $f^{\langle k \rangle}$ is identically 1. If $\prec^{\langle k \rangle} = \prec_{\mathrm{skip}}$, then $c^{\langle k \rangle} = 2$ and $f^{\langle k \rangle}$ is given by

$$f_{\mathrm{skip}}(i, j) = \begin{cases} 1 & j \equiv 0 \pmod 2 \\ 2 & j \equiv 1 \pmod 2 \end{cases} .$$

  Thus, we have omitted $c^{\langle k \rangle}$ and $f^{\langle k \rangle}$ from Tables II–V, which specify the parameters that yielded the results in Table I.

TABLE II
First row in Table I — $(2, \infty)$-RLL: $r = 3$, $s = 8$, $\ell = 7$.

k	$\rho^{\langle k \rangle}$	$\prec^{\langle k \rangle}$	$(a_*^{\langle k \rangle}, b_*^{\langle k \rangle})$
1	0.32	$\prec_{\mathrm{lex}}$	$(2, 2)$
2	0.20	$\prec_{\mathrm{lex}}$	$(2, 3)$
3	0.35	$\prec_{\mathrm{lex}}$	$(2, 4)$
4	0.08	$\prec_{\mathrm{lex}}$	$(2, 5)$
5	0.01	$\prec_{\mathrm{lex}}$	$(2, 7)$
6	0.03	$\prec_{\mathrm{skip}}$	$(2, 2)$
7	0.01	$\prec_{\mathrm{skip}}$	$(2, 3)$

TABLE III
Second row in Table I — $(3, \infty)$-RLL: $r = 4$, $s = 8$, $\ell = 5$.

k	$\rho^{\langle k \rangle}$	$\prec^{\langle k \rangle}$	$(a_*^{\langle k \rangle}, b_*^{\langle k \rangle})$
1	0.56	$\prec_{\mathrm{lex}}$	$(3, 3)$
2	0.23	$\prec_{\mathrm{lex}}$	$(3, 4)$
3	0.13	$\prec_{\mathrm{lex}}$	$(3, 5)$
4	0.04	$\prec_{\mathrm{lex}}$	$(3, 6)$
5	0.04	$\prec_{\mathrm{lex}}$	$(3, 7)$

References

[1] 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, 1998, pp. 1635–1764.

[2] W. Weeks and R. E. Blahut, "The capacity and coding gain of certain checkerboard codes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1193–1203, 1998.

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

[4] T. Etzion, "Cascading methods for runlength-limited arrays," *IEEE Trans. Inform. Theory*, vol. 43, pp. 319–324, 1997.

[5] S. Halevy, J. Chen, R. M. Roth, P. H. Siegel, and J. K. Wolf, "Improved bit-stuffing bounds on two-dimensional constraints," *IEEE Trans. Inform. Theory*, vol. 50, pp. 824–838, 2004.

[6] S. Forchhammer and J. Justesen, "Bounds on the capacity of constrained two-dimensional codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2659–2666, 2000.

[7] N. Calkin and H. S. Wilf, "The number of independent sets in a grid graph," *SIAM J. Discrete Math.*, vol. 11, pp. 54–60, 1997.

[8] R. Burton and J. E. Steif, "Non-uniqueness of measures of maximal entropy for subshifts of finite type," *Ergod. Th. Dynam. Sys.*, vol. 14, pp. 213–235, 1994.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[11] J. Czyzyk, M. P. Mesnier, and J. J. Moré, "The NEOS server," *IEEE Computational Science & Engineering*, vol. 5, no. 3, pp. 68–75, 1998.

[12] W. Gropp and J. J. Moré, "Optimization environments and the NEOS server," in *Approximation Theory and Optimization*. Cambridge University Press, 1997, pp. 167–182.

[13] E. D. Dolan, R. Fourer, J. J. Moré, and T. S. Munson, "The NEOS server for optimization: Version 4 and beyond," Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, Tech. Rep., 2002.

[14] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*, 2nd ed. Duxbury Press, 2002.

[15] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Math. Program.*, vol. 106, pp. 25–57, 2006.

[16] S. Forchhammer and T. V. Laursen, "Entropy of bit-stuffing-induced measures for two-dimensional checkerboard constraints," *IEEE Trans. Inform. Theory*, vol. 53, pp. 1537–1546, 2007.

[17] I. Tal and R. M. Roth, "Bounds on the rate of 2-D bit-stuffing encoders," *IEEE Trans. Inform. Theory*, vol. 56, pp. 2561–2567, 2010.

[18] A. Sharov and R. M. Roth, "Two-dimensional constrained coding based on tiling," *IEEE Trans. Inform. Theory*, vol. 56, pp. 1800–1807.

[19] S. Friedland, "On the entropy of Z^d subshifts of finite type," *Lin. Alg. Appls.*, vol. 252, pp. 199–220, 1997.

[20] S. Even, *Graph Algorithms*. Computer Science Press, 1979.

TABLE IV
THIRD ROW IN TABLE I — $(0,2)$-RLL: $r = 3$, $s = 5$, $\ell = 2$.

k	$\rho^{\langle k \rangle}$	$\prec^{\langle k \rangle}$	$(a_*^{\langle k \rangle}, b_*^{\langle k \rangle})$
1	0.54	$\prec_{\mathrm{lex}}$	$(2, 2)$
2	0.46	$\prec_{\mathrm{lex}}$	$(2, 3)$

TABLE V
FOURTH ROW — N.I.B.: $r = 3$, $s = 4$, $\ell = 1$.

k	$\rho^{\langle k \rangle}$	$\prec^{\langle k \rangle}$	$(a_*^{\langle k \rangle}, b_*^{\langle k \rangle})$
1	1	$\prec_{\mathrm{lex}}$	$(2, 2)$

Biographies:

Ido Tal was born in Haifa, Israel, in 1975. He received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Technion—Israel Institute of Technology, Haifa, Israel, in 1998, 2003 and 2009, respectively.

He is a Postdoctoral Scholar at the Information Theory and Applications (ITA) Center and is affiliated with the Center for Magnetic Recording Research (CMRR), both at the University of California at San Diego, La Jolla, CA, USA. His research interests include constrained coding and error-control coding.

Ron M. Roth was born in Ramat Gan, Israel, in 1958. He received the B.Sc. degree in computer engineering, the M.Sc. in electrical engineering and the D.Sc. in computer science from Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1984 and 1988, respectively. Since 1988 he has been with the Computer Science Department at Technion, where he now holds the General Yaakov Dori Chair in Engineering. During the academic years 1989–91 he was a Visiting Scientist at IBM Research Division, Almaden Research Center, San Jose, California, and during 1996–97 and 2004–05 he was on sabbatical leave at Hewlett–Packard Laboratories, Palo Alto, California. He is the author of the book *Introduction to Coding Theory* published by Cambridge University Press in 2006. Dr. Roth was an associate editor for coding theory in IEEE TRANSACTIONS ON INFORMATION THEORY from 1998 till 2001, and he is now serving as an associate editor in *SIAM Journal on Discrete Mathematics*. His research interests include coding theory, information theory, and their application to the theory of complexity.