Improved Bit-Stuffing Bounds on Two-Dimensional Constraints

Shirley Halevy, Jiangxin Chen, Ron M. Roth, *Fellow, IEEE*, Paul H. Siegel, *Fellow, IEEE*, and Jack Keil Wolf, *Life Fellow, IEEE*

***Abstract*—We derive lower bounds on the capacity of certain two-dimensional (2-D) constraints by considering bounds on the entropy of measures induced by bit-stuffing encoders. A more detailed analysis of a previously proposed bit-stuffing encoder for (d, ∞)-runlength-limited (RLL) constraints on the square lattice yields improved lower bounds on the capacity for all $d \geq 2$. This encoding approach is extended to (d, ∞)-RLL constraints on the hexagonal lattice, and a similar analysis yields lower bounds on the capacity for $d \geq 2$. For the hexagonal $(1, \infty)$-RLL constraint, the exact coding ratio of the bit-stuffing encoder is calculated and is shown to be within 0.5% of the (known) capacity. Finally, a lower bound is presented on the coding ratio of a bit-stuffing encoder for the constraint on the square lattice where each bit is equal to at least one of its four closest neighbors, thereby providing a lower bound on the capacity of this constraint.**

***Index Terms*—Bit-stuffing encoder, hexagonal constraint, runlength-limited (RLL) constraints, two-dimensional (2-D) constraints.**

I. Introduction

MANY data storage systems, such as those based upon magnetic and optical recording technology, require the use of constrained modulation codes. These codes transform, in a lossless manner, streams of arbitrary binary data into binary sequences that satisfy certain prescribed constraints. The set of words from which the code sequences may be drawn is referred to as a *constrained system*, or simply a *constraint*.

Historically, many digital recording applications have required that the binary recorded sequences belong to a (d, k)-runlength-limited (RLL) constraint. The parameters (d, k) represent, respectively, the minimum and maximum admissible number of 0's separating consecutive 1's in any allowable sequence. With the advent of page-oriented storage technologies, such as holographic storage, interest in constrained arrays in two or more dimensions has arisen; see Brady and Psaltis [5], Heanue, Bashaw, and Hesselink [11], [12], and Psaltis and Mok [25]. Among the constraints of theoretical and possible practical interest are two-dimensional (2-D) (d, k)-RLL constraints. When defined over the square lattice, each such constraint consists of all binary arrays in which the one-dimensional (1-D) (d, k)-RLL constraint is satisfied along each row and column. In both one and two dimensions, the relevant range of parameters is $0 \leq d < k \leq \infty$.

Runlength constraints can be defined also over the hexagonal lattice [1]. Using a simple transformation from the hexagonal lattice into the square lattice [17], one can define the 2-D (d, k)-RLL hexagonal constraint as the set of all binary arrays in which the 1-D (d, k)-RLL constraint is satisfied along each row, each column, and each "northeast-to-southwest" (i.e., upper-right to lower-left) diagonal.

Another example is the 2-D "no isolated bits" *constraint* (in short, the n.i.b. constraint), which consists of all binary arrays that contain neither the pattern

	0	
0	1	0
	0	

nor its complement. Observe that the n.i.b. constraint is the natural generalization to two dimensions of the 1-D constraint that consists of all binary sequences in which every bit—except possibly for the first and last bits—is equal to at least one of its adjacent bits; this constraint, in turn, can be described as a *precoding* of the 1-D $(1, \infty)$-RLL constraint; see [20, Section 1].

The n.i.b. constraint (and generalizations thereof) may be found in future optical disks. Attempts to increase the recording density have been made recently by exploiting the fact that the recording device is typically a *surface*: the recorded data is regarded as 2-D, as opposed to the track-oriented 1-D recording model. When recording on the disk, "pits" and "lands" on the recording surface must be large enough so that they can be detected from the reflection beam [24, Ch. 3]. This, in turn, dictates that the recorded data belongs to the n.i.b. constraint. See also Psaltis *et al.* [26] and Weeks and Blahut [34].

In general, a 2-D constraint $\mathbb{S}$ over an alphabet Σ is defined by two state-labeled finite directed graphs, $G = (V, E_G, L)$ and $H = (V, E_H, L)$, with the same set of states V and the same state labeling $L : V \to \Sigma$. The constraint $\mathbb{S}$ consists of all finite rectangular arrays $x = (x_{i,j})$ over Σ for which one can associate arrays $\Psi(x) = (v_{i,j})$ over V that satisfy the following three conditions: a) $L(v_{i,j}) = x_{i,j}$ for all i and j; b) each row in $\Psi(x)$ is a path in G; and c) each column in $\Psi(x)$ is a path in H.

Manuscript received October 4, 2002; revised October 15, 2003. This work was supported in part by the United-States-Israel Binational Science Foundation (BSF), Jerusalem, Israel, under Grants 98-199 and 2002197. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

S. Halevy and R. M. Roth are with the Computer Science Department, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: shirleyh@cs.technion.ac.il; ronny@cs.technion.ac.il).

J. Chen, P. H. Siegel, and J. K. Wolf are with the Department of Electrical and Computer Engineering and the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: chenj@qualcomm.com; psiegel@ucsd.edu; jwolf @ucsd.edu).

Communicated by K. Abdel-Ghaffar, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2004.826633

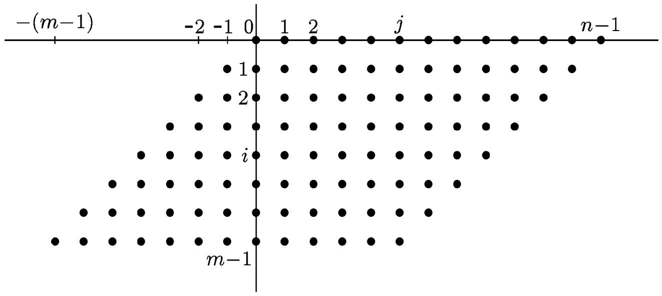


Fig. 1. Parallelogram $\Delta_{m,n}$.

The (d,k)-RLL constraints on the square and hexagonal lattices will be denoted by $\mathbb{S}_{\rm sq}^{d,k}$ and $\mathbb{S}_{\rm hex}^{d,k}$, respectively, while the n.i.b. constraint will be denoted by $\mathbb{S}_{\rm nib}$.

Let U be a finite subset of $\mathbb{Z}^2$ and let Σ be a finite alphabet. A *U-configuration* is a mapping $\varphi : U \to \Sigma$. Given a 2-D constraint $\mathbb{S}$ over Σ, we denote by $\mathbb{S}(U)$ the set of all U-configurations φ for which there exists an array $(x_{i,j}) \in \mathbb{S}$ such that

$$\varphi(i,j) = x_{i,j}, \qquad \text{for every } (i,j) \in U$$

that is, the images of φ at the elements $(i,j) \in U$ can be extended to an array in $\mathbb{S}$.

For a U-configuration x, we denote by $x_{i,j}$ the value of x at location $(i,j) \in U$. Given two finite subsets $U' \subseteq U$ of $\mathbb{Z}^2$ and a U-configuration x, we denote by $x(U')$ the U'-configuration x' which is the restriction of x to U'; namely, $x_{i,j} = x'_{i,j}$ for every $(i,j) \in U'$.

The subsets $U \subset \mathbb{Z}^2$ considered in this work will mainly be rectangles

$$B_{m,n} = \{(i,j) \in \mathbb{Z}^2 : 0 \le i < m,\ 0 \le j < n\}$$

or parallelograms

$$\Delta_{m,n} = \{(i,j) \in \mathbb{Z}^2 : 0 \le i < m,\ 0 \le i+j < n\} \qquad (1)$$

(see Fig. 1).

The capacity of the 2-D constraint $\mathbb{S}$ is defined by

$$\mathsf{cap}(\mathbb{S}) = \lim_{m,n\to\infty} \frac{1}{mn} \log_2 |\mathbb{S}(B_{m,n})|$$

i.e., it measures the growth rate of the number of $m \times n$ arrays in $\mathbb{S}$. By subadditivity, the limit indeed exists (see [6], [14], [15], [21], [30]). It is easy to see that we also have

$$\mathsf{cap}(\mathbb{S}) = \lim_{m,n\to\infty} \frac{1}{mn} \log_2 |\mathbb{S}(\Delta_{m,n})|.$$

One can readily verify that $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,k}\right) \le \mathsf{cap}(S^{d,k})$, where $\mathsf{cap}(S^{d,k})$ stands for the capacity of the 1-D (d,k)-RLL constraint [20, p. 1672]. However, the corresponding 1- and 2-D capacities may be quite different. For example [2], the capacity of the 1-D (1,2)-RLL constraint satisfies $\mathsf{cap}(S^{1,2}) \approx 0.4057$, whereas the capacity of the corresponding 2-D constraint is $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{1,2}\right) = 0$. This result has been generalized to a complete characterization of the (d,k)-RLL constraints in two dimensions and higher with zero capacity [13], [15]. Specifically, for every $r \ge 2$, the capacity of the r-dimensional (d,k)-RLL constraint is zero if and only if $d > 0$ and $k = d+1$. Partial characterizations exist also when the horizontal and vertical runlength constraints are not necessarily the same [16].

The determination of whether $\mathsf{cap}(\mathbb{S}) \ge 0$ for a given 2-D constraint $\mathbb{S}$ is known to be an undecidable problem [4], [27]. As for the special case of 2-D (d,k)-RLL constraints, no efficient algorithms are known for approximating their capacity. The case $(d,k) = (1,\infty)$ (or equivalently, $(d,k) = (0,1)$) has arisen in various forms in statistical mechanics and combinatorics, as well as in the information-theoretic context. Engel [9] and Calkin and Wilf [7] used an adjacency matrix method to derive a technique for obtaining close lower and upper bounds for this constraint. Using this technique, it has been shown that $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{1,\infty}\right)$ agrees with 0.587891161 up to the first nine decimal places [22], [34].

Kato and Zeger [15] used the bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{1,\infty}\right)$ to derive lower bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,\infty}\right)$, for $d \ge 2$, and $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{0,k}\right)$ for $k \ge 2$. (They noted that Talyansky [32] and Talyansky, *et al.* [33] described a construction that yields a lower bound on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{0,k}\right)$ that is stronger than the Kato–Zeger bound for all $k \ge 8$.) The lower bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{0,k}\right)$ were then used to derive lower bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,k}\right)$ for the remaining cases where $k \ne d+1$. Upper bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,\infty}\right)$ and $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{0,k}\right)$ were also derived [15]. Together with the lower bounds, they imply that, as d grows, $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,\infty}\right)$ converges to 0 exactly at the rate $(\log_2 d)/d$, and they give asymptotic bounds on how fast, as k grows, $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{0,k}\right)$ converges to 1.

Siegel and Wolf [31] (see also [28]) used a different approach to derive lower bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,\infty}\right)$, for $d \ge 1$. They computed a simple lower bound on the average coding ratio of a variable-rate, bit-stuffing encoding algorithm that creates 2-D (d,∞)-RLL constrained arrays from a 1-D sequence produced by a possibly biased binary source. These lower bounds were then optimized with respect to the 1-D binary source probability. We review the technique of Siegel and Wolf in Section II and fill in some details of the proof that were missing from [28] and [31]. The bit-stuffing approach is closely related to one introduced by Lee [18] and Bender and Wolf [3] for 1-D, RLL, charge-constrained $(d,k;c)$ sequences [23]. Roth, Siegel, and Wolf [29] have recently improved the results of [31] for the constraint $\mathbb{S}_{\rm sq}^{1,\infty}$. The improvement has been obtained by applying a more generalized model of a bit-stuffing encoder and by a refinement of the analysis; the coding ratio thus obtained is approximately 0.587277, i.e., only 0.1% below $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{1,\infty}\right)$.

In Section III, we present improved lower bounds on the coding ratio of a bit-stuffing encoder for $\mathbb{S}_{\rm sq}^{d,\infty}$. Then, in Section IV, we adapt the bit-stuffing encoder to the $\mathbb{S}_{\rm hex}^{d,\infty}$ and use a similar analysis to derive lower bounds on its coding ratio, for $d \ge 2$. For $\mathbb{S}_{\rm hex}^{1,\infty}$, we compute the exact coding ratio of the bit-stuffing encoder using results in [29] and we show that it lies within 0.5% of the (known) capacity.

Finally, in Section V, we present a bit-stuffing coding scheme for $\mathbb{S}_{\rm nib}$. We show that the encoding ratio of this scheme—and hence the value $\mathsf{cap}(\mathbb{S}_{\rm nib})$—is at least 0.91276; this value is fairly close to our empirical estimate ($\approx$ 0.917) on the true coding ratio of this scheme (yet $\mathsf{cap}(\mathbb{S}_{\rm nib})$ is believed to be bigger; see the discussion at the end of Section V). In comparison, Ashley and Marcus estimated in [2] the value $\mathsf{cap}(\mathbb{S}_{\rm nib})$ to be around 0.86.

As was the case in [29], the bit-stuffing encoders that we consider induce probability measures $\mu_{m,n}$ on the corresponding constraints. A lower bound on the coding ratio is then obtained by computing a lower bound on the (measure-theoretic) entropy of $\mu_{m,n}$

$$\mathsf{H}(\mu_{m,n}) = -\frac{1}{mn} \sum_{x \in \mathbb{S}(\Delta_{m,n})} \mu_{m,n}(x) \log_2 \mu_{m,n}(x) \tag{2}$$

in the limit when both m and n go to infinity. Recall, however, that the analysis of the encoder in [29] was based on the very strong Markovian stationary properties of the measure induced by the encoder upon $\mathbb{S}_{\rm sq}^{1,\infty}$. Unfortunately, such properties do not seem to hold in general; in particular, they do not hold for the measures $\mu_{m,n}$ induced by the encoders discussed here. Nevertheless, we can still derive lower bounds on the coding ratio based upon much weaker *global* properties of $\mu_{m,n}$. (These properties support the "stationarity" assumption underlying the analysis in [31].) Specifically, we obtain lower bounds as a function of expected values of the number of occurrences of certain events in the generated outputs for the constraints considered. By further establishing relations between those expected values, we have ended up with numerical lower bounds.

As an alternative approach, one can guarantee a "quasi-stationary" induced measure by a proper initialization of the boundary values in the generated output. We discuss this in Section VI.

II. Bit-Stuffing Lower Bounds on $\mathsf{cap}\left(\mathbb{S}_{\rm sq}^{d,\infty}\right)$

We describe next a bit-stuffing encoder that maps unconstrained data into $\mathbb{S}_{\rm sq}^{d,\infty}(\Delta_{m,n})$, where $\Delta_{m,n}$ is the parallelogram defined in (1) and shown in Fig. 1.

We will use the following terms. *Row* i in $\Delta_{m,n}$ consists of all locations (i,j) such that $-i \le j < n-i$. *Diagonal* r consists of all locations $(i, r-i)$ such that $0 \le i < m$. The first M rows and diagonals in $\Delta_{m,n}$ form its boundary of width M and will be denoted by $\partial\Delta_{m,n}^{(M)}$; that is,

$$\partial\Delta_{m,n}^{(M)} = \{(i,j) \in \Delta_{m,n} : i < M \quad \text{or} \quad i+j < M\}.$$

The bit-stuffing encoder first applies a distribution transformer $\mathcal{E}$ that bijectively converts the binary data sequence into a sequence of statistically independent bits which is p*-biased* for some real $p \in [0,1]$: the probability of a 1 equals p and the probability of a 0 equals $1-p$. This conversion occurs at a rate penalty of $h(p)$, where $h(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy function. The purpose of creating a p-biased sequence will be to write more 0's than 1's. The optimal value of p will be chosen later. We now write the p-biased sequence (without further coding) down successive diagonals, skipping all positions that contain "stuffed" 0's, which arise in a manner which will now be explained. Whenever a 1 in the p-biased source sequence is written, d 0's are inserted—or "stuffed"—in the d positions to the right of it and in the d positions below it. It will sometimes occur that a 0 has already been stuffed in some of the positions to the right of the 1 or (when $d > 1$) below it, in which case it is not necessary to stuff another 0. In writing the p-biased sequence down diagonals, any position already filled by a previously stuffed 0 is skipped.

To define the encoding process also for the boundary of $\Delta_{m,n}$ of width d, we assume identically-zero entries at all locations (i,j) such that $i < 0$ or $i+j < 0$.

(We mention that the coding can alternatively be done into elements of $\mathbb{S}_{\rm sq}^{d,\infty}(B_{m,n})$, where entries are generated row by row or column by column.)

Decoding the array is accomplished by reading down diagonals in a similar manner. The bits of the p-biased sequence are read successively from the array, with certain 0 bits being ignored. Specifically, whenever a 1 is read from the array, the stuffed 0's to the right of it and below it are normally deleted. It may occur that stuffed 0's to the right of the 1 or below it have already been deleted, in which case only the remaining stuffed 0's to the right and the stuffed 0's below it are deleted. This procedure reproduces the encoded p-biased sequence. The original binary data is then obtained from the p-biased stream by the inverse of the mapping used to create the p-biased stream of bits.

The bit-stuffing encoder induces a probability measure on $\mathbb{S}_{\rm sq}^{d,\infty}(\Delta_{m,n})$. We denote this measure by $\mu_{m,n} = \mu_{{\rm sq};m,n}^{d,\infty}$ and we have

$$\mu_{m,n}(x) = \prod_{i,j \in \Delta_{m,n}} \vartheta_{\rm sq}\left(x_{i,j} \mid x_{i,j-1}, x_{i,j-2}, \ldots, x_{i,j-d},\right.$$
$$\left. x_{i-1,j}, x_{i-2,j}, \ldots, x_{i-d,j}\right) \tag{3}$$

where the function $\vartheta_{\rm sq} : \{0,1\}^{2d+1} \to [0,1]$ is defined by

$$\vartheta_{\rm sq}(1 \mid y_1, y_2, \ldots, y_{2d}) = \begin{cases} p, & \text{if } y_1 = y_2 = \cdots = y_{2d} = 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$\vartheta_{\rm sq}(0 \mid y_1, y_2, \ldots, y_{2d}) = 1 - \vartheta_{\rm sq}(1 \mid y_1, y_2, \ldots, y_{2d})$$

and $x_{i,j}$ is assumed to be zero whenever $i < 0$ or $i+j < 0$.

Given a random element $X \in \mathbb{S}_{\rm sq}^{d,\infty}(\Delta_{m,n})$ that is generated by the bit-stuffing encoder (according to the probability measure $\mu_{m,n}$), denote by $\mathcal{X}_{i,j}$ the event "$X_{i,j} = 1$" (this event never occurs when $i < 0$ or $i+j < 0$) and let the event $\mathcal{C}_{i,j}$ be defined for $(i,j) \in \Delta_{m,n}$ by

$$\mathcal{C}_{i,j} = \bigcup\nolimits_{s=1}^{d} \mathcal{X}_{i-s,j}$$

namely, $\mathcal{C}_{i,j}$ stands for the event, "location (i,j) in X is a stuffed 0 as a result of one of the d locations above (i,j) being equal to 1." Similarly, define

$$\mathcal{R}_{i,j} = \bigcup\nolimits_{t=1}^{d} \mathcal{X}_{i,j-t}$$

and

$$\mathcal{B}_{i,j} = \overline{\mathcal{C}_{i,j} \cup \mathcal{R}_{i,j}}$$

where the overbar stands for complementation; that is, $\mathcal{B}_{i,j}$ is the event that location (i,j) in X is filled by a bit of the p-biased sequence. We denote by $\mathcal{C}$ (respectively, $\mathcal{R}$ and $\mathcal{B}$) the random variable that stands for the number of indexes $(i,j) \in \Delta_{m,n}$ where the event $\mathcal{C}_{i,j}$ (respectively, $\mathcal{R}_{i,j}$ and $\mathcal{B}_{i,j}$) occurs. Clearly

$$\mathsf{E}\{\mathcal{C}\} = \sum_{(i,j)\in\Delta_{m,n}} \mathsf{Pr}\{\mathcal{C}_{i,j}\}, \quad \mathsf{E}\{\mathcal{R}\} = \sum_{(i,j)\in\Delta_{m,n}} \mathsf{Pr}\{\mathcal{R}_{i,j}\}$$

and

$$\mathsf{E}\{\mathcal{B}\} = \sum_{(i,j)\in\Delta_{m,n}} \mathsf{Pr}\{\mathcal{B}_{i,j}\}.$$

The following lemma easily follows from (2) and (3).

Lemma 2.1:

$$\mathsf{H}(\mu_{m,n}) = \frac{1}{mn}\cdot h(p)\cdot \mathsf{E}\{\mathcal{B}\}.$$

Lemma 2.2:

$$\mathsf{E}\{\mathcal{C}\} \le dp\cdot\mathsf{E}\{\mathcal{B}\} \quad \textit{and} \quad \mathsf{E}\{\mathcal{R}\} \le dp\cdot\mathsf{E}\{\mathcal{B}\}.$$

Proof: For every $(i,j) \in \Delta_{m,n}$

$$\mathsf{Pr}\{\mathcal{C}_{i,j}\} \le \sum_{s=1}^{\min\{d,i\}} \mathsf{Pr}\{\mathcal{X}_{i-s,j}\}$$
$$\text{and} \quad \mathsf{Pr}\{\mathcal{R}_{i,j}\} \le \sum_{t=1}^{\min\{d,i+j\}} \mathsf{Pr}\{\mathcal{X}_{i,j-t}\}$$

and

$$\mathsf{Pr}\{\mathcal{X}_{i,j}\} = \mathsf{Pr}\{\mathcal{X}_{i,j} \mid \mathcal{B}_{i,j}\}\cdot\mathsf{Pr}\{\mathcal{B}_{i,j}\} = p\cdot\mathsf{Pr}\{\mathcal{B}_{i,j}\}.$$

Therefore, for every $(i,j) \in \Delta_{m,n}$

$$\mathsf{Pr}\{\mathcal{C}_{i,j}\} \le p\cdot \sum_{s=1}^{\min\{d,i\}} \mathsf{Pr}\{\mathcal{B}_{i-s,j}\}$$
$$\text{and} \quad \mathsf{Pr}\{\mathcal{R}_{i,j}\} \le p\cdot \sum_{t=1}^{\min\{d,i+j\}} \mathsf{Pr}\{\mathcal{B}_{i,j-t}\}.$$

Summing over all $(i,j) \in \Delta_{m,n}$ we have

$$\begin{aligned}\mathsf{E}\{\mathcal{C}\} &= \sum_{(i,j)\in\Delta_{m,n}} \mathsf{Pr}\{\mathcal{C}_{i,j}\} \\ &\le \sum_{(i,j)\in\Delta_{m,n}} p\cdot \sum_{s=1}^{\min\{d,i\}} \mathsf{Pr}\{\mathcal{B}_{i-s,j}\} \\ &\le dp \sum_{(i,j)\in\Delta_{m,n}} \mathsf{Pr}\{\mathcal{B}_{i,j}\} = dp\cdot\mathsf{E}\{\mathcal{B}\}.\end{aligned}$$

The inequality $\mathsf{E}\{\mathcal{R}\} \le dp\cdot\mathsf{E}\{\mathcal{B}\}$ is obtained in a similar manner. □

Lemma 2.3: For $d > 0$ and any $0 \le p \le 1$

$$\mathsf{H}(\mu_{m,n}) \ge \frac{h(p)}{1+2dp}.$$

Proof: Taking expectations of both sides of the inequality

$$\mathcal{B} \ge mn - \mathcal{R} - \mathcal{C}$$

we obtain by Lemma 2.2

$$\mathsf{E}\{\mathcal{B}\} \ge mn - 2dp\cdot\mathsf{E}\{\mathcal{B}\}$$

and so

$$\frac{\mathsf{E}\{\mathcal{B}\}}{mn} \ge \frac{1}{1+2dp}.$$

The proof now follows from Lemma 2.1. □

It was shown in [28], [31] that when ranging over $p \in [0,1]$, the maximum value of $h(p)/(1+2dp)$ equals the capacity of the 1-D constraint $S^{2d,\infty}$. Thus, we obtain the following lower bound.

Proposition 2.4: For $d > 0$

$$\mathsf{H}(\mu_{m,n}) \ge \mathsf{cap}(S^{2d,\infty}).$$

Table I shows the lower bound of Proposition 2.4 for small values of d. Also shown are lower bounds computed numerically using [15, Theorems 5 and 6].

TABLE I
LOWER BOUNDS ON THE RATE OF A BIT-STUFFING ENCODER FOR $\mathbb{S}_{\mathrm{sq}}^{d,\infty}$

d	From [15]	Proposition 2.4	Proposition 3.3	Improved Bounds
2	0.3509	0.4057	0.4190	0.4267
3	0.2939	0.3282	0.3347	0.3402
4	0.2408	0.2788	0.2825	0.2858
5	0.2118	0.2440	0.2464	–

III. IMPROVED BOUNDS FOR $\mathbb{S}_{\mathrm{sq}}^{d,\infty}$

Our improvement on the results of Section II will be obtained by accounting for some of the patterns that give rise to an "overlap" of stuffed 0's in the encoded output.

Let $\mu_{m,n}$ denote the probability measure induced on $\mathbb{S}_{\mathrm{sq}}^{d,\infty}(\Delta_{m,n})$ by the bit-stuffing encoder of Section II. Also, for a random element $X \in \mathbb{S}_{\mathrm{sq}}^{d,\infty}(\Delta_{m,n})$, we let the notations $\mathcal{X}_{i,j}$, $\mathcal{C}$, $\mathcal{R}$, $\mathcal{B}_{i,j}$, and $\mathcal{B}$ be as in Section II.

Lemma 3.1: For a random element $X \in \mathbb{S}_{\mathrm{sq}}^{d,\infty}(\Delta_{m,n})$ and an index $(i,j) \in \Delta_{m,n}$, let $\mathcal{A}$ denote an event that is a function of the random variables $X_{s,t}$, where $s < i$ or $t < j$. Then

$$\mathsf{Pr}\{\mathcal{X}_{i,j} \mid \mathcal{B}_{i,j} \cap \mathcal{A}\} = \mathsf{Pr}\{\mathcal{X}_{i,j} \mid \mathcal{B}_{i,j}\} = p.$$

Proof: Define the sets

$$U_{i,j} = \Delta_{m,n} \cap \{(s,t)\in\mathbb{Z}^2 : s \le i \ \underline{\text{and}}\ t \le j \ \underline{\text{and}}\ (s,t) \ne (i,j)\}$$

and

$$V_{i,j} = \Delta_{m,n} \cap \{(s,t)\in\mathbb{Z}^2 : s < i \ \underline{\text{or}}\ t < j\}.$$

In Fig. 2, the set $V_{i,j}$ corresponds to the lattice points marked with dots (filled or unfilled), while $U_{i,j}$ corresponds to the subset of $V_{i,j}$ marked with filled dots. The following two facts follow from the particular encoding process applied by the bit-stuffing algorithm.

1. The $U_{i,j}$-configuration $X(U_{i,j})$ (which is the restriction of X to $U_{i,j}$) completely determines whether the event $\mathcal{B}_{i,j}$ occurs; so, we can regard $\mathcal{B}_{i,j}$ as the set of all $U_{i,j}$-configurations y that imply $\mathcal{B}_{i,j}$.
2. Given the configuration $X(U_{i,j})$, the entries in $X(V_{i,j} \setminus U_{i,j})$ (corresponding to the locations marked with unfilled dots in Fig. 2) are statistically independent of the

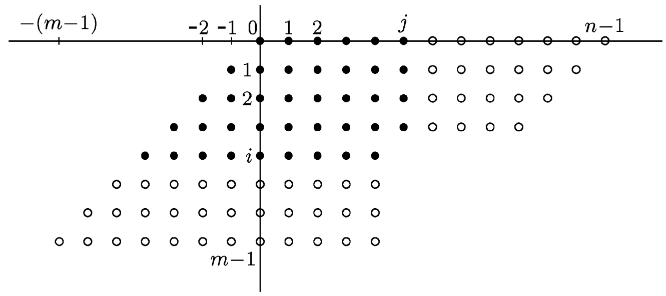


Fig. 2. Subsets $U_{i,j}$ and $V_{i,j}$ in $\Delta_{m,n}$.

event $\mathcal{X}_{i,j}$ (regardless of whether any of those values is encoded prior to location (i,j)).

Therefore,

$$\begin{aligned}\mathsf{Pr}\{\mathcal{X}_{i,j}\cap\mathcal{A}\cap\mathcal{B}_{i,j}\} &= \sum_{y\in\mathcal{B}_{i,j}}\mathsf{Pr}\{\mathcal{X}_{i,j}\cap\mathcal{A}\cap(X(U_{i,j})=y)\}\\ &= \sum_{y\in\mathcal{B}_{i,j}}\big(\mathsf{Pr}\{\mathcal{X}_{i,j}\mid X(U_{i,j})=y\}\\ &\qquad\cdot\mathsf{Pr}\{\mathcal{A}\mid X(U_{i,j})=y\}\\ &\qquad\cdot\mathsf{Pr}\{X(U_{i,j})=y\}\big)\\ &= p\cdot\sum_{y\in\mathcal{B}_{i,j}}\big(\mathsf{Pr}\{\mathcal{A}\mid X(U_{i,j})=y\}\\ &\qquad\cdot\mathsf{Pr}\{X(U_{i,j})=y\}\big)\\ &= p\cdot\mathsf{Pr}\{\mathcal{A}\cap\mathcal{B}_{i,j}\}\end{aligned}$$

where the first equality follows from Fact 1 and the second equality from Fact 2. □

Let $\Lambda(i,j)$ and $\Gamma(i,j)$ be subsets of $\mathbb{Z}^2$ defined by

$$\Lambda(i,j)=\Big\{(i-a,j-a)\Big\}_{a=1}^{d-1}\bigcup\Big\{(i-d-a,j+d+1-a)\Big\}_{a=1}^{d} \tag{4}$$

and

$$\Gamma(i,j)=\Big\{(s,t)\in\mathbb{Z}^2 : i-d\le s<i \text{ and } j<t\le j+d\Big\}$$

respectively, and let $\mathcal{S}_{i,j}$ for $(i,j)\in\Delta_{m,n}$ denote the event

$$\mathcal{S}_{i,j}=\bigcap\nolimits_{(s,t)\in\Lambda(i,j)}\mathcal{X}_{s,t}.$$

Fig. 3(a) depicts the event $\mathcal{S}_{i,j}$ for $d=3$. In the figure, the entries indexed by $\Lambda(i,j)$ are marked by 1's and the entries indexed by $\Gamma(i,j)$ are marked by thick dots.

Lemma 3.2: For $(i,j)\in\Delta_{m,n}\setminus\partial\Delta_{m,n}^{(2d)}$

$$(\mathcal{B}_{i,j}\setminus\mathcal{S}_{i,j})=\mathcal{B}_{i,j}\cap\left(\bigcup\nolimits_{(s,t)\in\Gamma(i,j)}\mathcal{B}_{s,t}\right).$$

Proof: We need to show that

$$(\mathcal{B}_{i,j}\cap\mathcal{S}_{i,j})=\mathcal{B}_{i,j}\cap\overline{\bigcup\nolimits_{(s,t)\in\Gamma(i,j)}\mathcal{B}_{s,t}}. \tag{5}$$

The event $\mathcal{B}_{i,j}\cap\mathcal{S}_{i,j}$ in the left-hand side of (5) is shown in Fig. 3(b) for $d=3$; the event in the right-hand side of (5) states

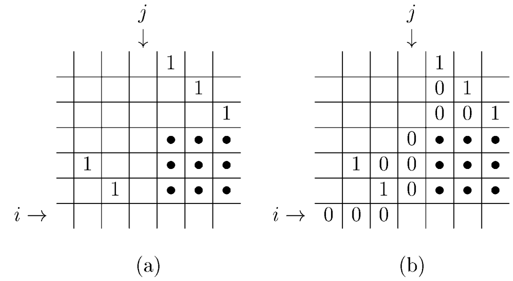


Fig. 3. (a) Event $\mathcal{S}_{i,j}$ and (b) event $\mathcal{B}_{i,j}\cap\mathcal{S}_{i,j}$ for $d=3$.

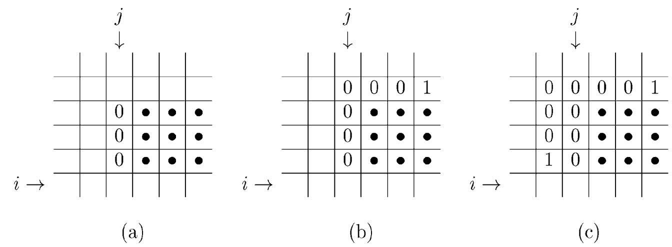


Fig. 4. Patterns for the proof of Lemma 3.2 for $d=3$.

that location (i,j) contains a p-biased bit, while all locations $(s,t)\in\Gamma(i,j)$ are stuffed with 0's.

It is easy to see that $\mathcal{B}_{i,j}\cap\mathcal{S}_{i,j}$ is a subset of the event in the right-hand side of (5). Next we show the inclusion in the other direction.

Let X be an array that belongs to the event in the right-hand side of (5). In particular, X contains 0's in all locations that are indexed by the $d\times(d+1)$ rectangle

$$\hat{\Gamma}(i,j)=\{(s,t)\in\mathbb{Z}^2 : i-d\le s<i \text{ and } j\le t\le j+d\}$$

as shown in Fig. 4(a) for the case $d=3$ (the thick dots, which represent the entries that are indexed by $\Gamma(i,j)$, are all 0). The entry $X_{i-1,j+d}$ can be stuffed only if $X_{i-d-1,j+d}=1$. This, in turn, implies that $X_{i-d-1,t}=0$ for all $j\le t<j+d$, thereby reaching the pattern shown in Fig. 4(b) for the case $d=3$.

Next we turn to location $(i-1,j+d-1)$. Since $X_{s,t}=0$ for all $i-d-1\le s<i-1$ and $j\le t<j+d-2$, the entry $X_{i-1,j+d-1}$ can be stuffed only if $X_{i-1,j-1}=1$; this brings us to the pattern in Fig. 4(c).

We conclude that $X_{s,t}=0$ for all locations (s,t) within the $d\times(d+1)$ rectangle $\hat{\Gamma}(i-1,j-1)$. The proof now reiterates for this shifted rectangle. □

In the sequel, we assume some fixed linear ordering on the elements of $\mathbb{Z}^2$ that satisfies the following condition: if (ς,τ) precedes (s,t) (denoted $(\varsigma,\tau)\prec(s,t)$), then $\varsigma<s$ or $\tau<t$. For example, the standard lexicographic ordering

$$(\varsigma,\tau)\prec(s,t)\Longleftrightarrow((\varsigma<s)\text{ or }(\varsigma=s\text{ and }\tau<t))$$

satisfies this condition.

Hereafter, the notation $O(t)$ stands for a real expression $f(t)$ such that

$$\overline{\lim}_{t\to\infty}|f(t)|/t<\infty$$

and $o_t(1)$ stands for an expression $f(t)$ such that

$$\overline{\lim}_{t\to\infty}|f(t)|=0.$$

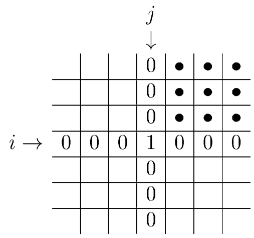


Fig. 5. Neighborhood of an entry set to 1 for the case $d = 3$.

Proposition 3.3:

$$\mathsf{H}(\mu_{m,n}) \geq \max_{0<p<1} \frac{h(p)}{1+2dp-p^2(1-p^{2d-1})} - o_{(\min\{m,n\})/d}(1).$$

Proof: Consider a portion of X in the neighborhood of an index $(i,j) \in \Delta_{m,n} \backslash \partial\Delta_{m,n}^{(2d)}$ where $X_{i,j} = 1$, as shown in Fig. 5 for the case $d = 3$. Clearly, the value $X_{i,j} = 1$ requires that the entries indexed by $\{(i,t)\}_{t=j+1}^{j+d}$ be stuffed with 0's (and so be the entries indexed by $\{(s,j)\}_{s=i+1}^{i+d}$). Now, entries indexed by $\{(i,t)\}_{t=j+1}^{j+d}$ may be stuffed with 0's also because of some other entries in X that are set to 1; yet, the latter entries are limited to the locations indexed by $\Gamma(i,j)$. Our improvement of the lower bound on $\mathsf{H}(\mu_{m,n})$ will be obtained by taking into account such a "double-stuffing."

As a first step, we bound from below the probability of the event $(\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j})$. Denote by $(u,v) = (u_{i,j}, v_{i,j})$ the smallest element (with respect to the ordering $\prec$) in the set $\Lambda(i,j)$ defined by (4). The probability of the event $\mathcal{S}_{i,j}$ can be written as a product of $2d-1$ terms

$$\mathsf{Pr}\{\mathcal{S}_{i,j}\} = \mathsf{Pr}\{\mathcal{X}_{u,v}\} \cdot \prod_{(s,t)\in\Lambda(i,j)\setminus\{(u,v)\}} \mathsf{Pr}\left\{\mathcal{X}_{s,t} \mid \bigcap_{(\varsigma,\tau)\in\Lambda(i,j):(\varsigma,\tau)\prec(s,t)} \mathcal{X}_{\varsigma,\tau}\right\}.$$

The first term in the right-hand side of the preceding equation is given by

$$\mathsf{Pr}\{\mathcal{X}_{u,v}\} = \mathsf{Pr}\{\mathcal{X}_{u,v} \mid \mathcal{B}_{u,v}\} \cdot \mathsf{Pr}\{\mathcal{B}_{u,v}\} = p \cdot \mathsf{Pr}\{\mathcal{B}_{u,v}\}$$

and for any other term therein we have

$$\begin{aligned}
&\mathsf{Pr}\left\{\mathcal{X}_{s,t} \middle| \bigcap_{(\varsigma,\tau)\prec(s,t)} \mathcal{X}_{\varsigma,\tau}\right\} \\
&\quad = \mathsf{Pr}\left\{\mathcal{X}_{s,t} \middle| \mathcal{B}_{s,t} \cap \bigcap_{(\varsigma,\tau)\prec(s,t)} \mathcal{X}_{\varsigma,\tau}\right\} \\
&\qquad \cdot \mathsf{Pr}\left\{\mathcal{B}_{s,t} \middle| \bigcap_{(\varsigma,\tau)\prec(s,t)} \mathcal{X}_{\varsigma,\tau}\right\} \\
&\quad \leq \mathsf{Pr}\left\{\mathcal{X}_{s,t} \middle| \mathcal{B}_{s,t} \cap \bigcap_{(\varsigma,\tau)\prec(s,t)} \mathcal{X}_{\varsigma,\tau}\right\} \\
&\quad = p,
\end{aligned}$$

where the last step follows from Lemma 3.1. Hence,

$$\mathsf{Pr}\{\mathcal{S}_{i,j}\} \leq p^{2d-1} \cdot \mathsf{Pr}\{\mathcal{B}_{u,v}\} \tag{6}$$

and so

$$\begin{aligned}
\mathsf{Pr}\{\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j}\} &\geq \mathsf{Pr}\{\mathcal{B}_{i,j}\} - \mathsf{Pr}\{\mathcal{S}_{i,j}\} \\
&\geq \mathsf{Pr}\{\mathcal{B}_{i,j}\} - p^{2d-1} \cdot \mathsf{Pr}\{\mathcal{B}_{u,v}\}.
\end{aligned} \tag{7}$$

We now turn to bounding from below the probability of the "double-stuffing" event

$$\mathcal{M}_{i,j} = \mathcal{X}_{i,j} \cap \left(\bigcup_{(s,t)\in\Gamma(i,j)} \mathcal{X}_{s,t}\right)$$

namely, the event of having 1 at location (i,j) and at one or more of the d^2 locations that are indexed by $\Gamma(i,j)$. Since

$$\mathcal{M}_{i,j} \subseteq (\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j})$$

it follows that

$$\begin{aligned}
\mathsf{Pr}\{\mathcal{M}_{i,j}\} &= \mathsf{Pr}\{\mathcal{M}_{i,j} | (\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j})\} \cdot \mathsf{Pr}\{\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j}\} \\
&\geq \mathsf{Pr}\{\mathcal{M}_{i,j} | (\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j})\} \\
&\quad \cdot (\mathsf{Pr}\{\mathcal{B}_{i,j}\} - p^{2d-1} \cdot \mathsf{Pr}\{\mathcal{B}_{u,v}\})
\end{aligned} \tag{8}$$

where the inequality follows from (7).

As our next step, we show that

$$\mathsf{Pr}\{\mathcal{M}_{i,j} | (\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j})\} \geq p^2. \tag{9}$$

By Lemma 3.2 and de Morgan laws [19, Sec. 1.2] we have

$$(\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j}) = \bigcup_{(s,t)\in\Gamma(i,j)} (\mathcal{B}_{i,j} \cap \mathcal{B}_{s,t}).$$

So, we can partition the event $\mathcal{B}_{i,j} \setminus \mathcal{S}_{i,j}$ into d^2 disjoint events $\{\mathcal{A}_{s,t}\}_{(s,t)\in\Gamma(i,j)}$ that are defined inductively for successive indexes $(s,t) \in \Gamma(i,j)$ (according to the ordering $\prec$) as follows:

$$\mathcal{A}_{s,t} = (\mathcal{B}_{i,j} \cap \mathcal{B}_{s,t}) \setminus \left(\bigcup_{(\varsigma,\tau)\in\Gamma(i,j):(\varsigma,\tau)\prec(s,t)} \mathcal{A}_{\varsigma,\tau}\right), \quad (s,t) \in \Gamma(i,j).$$

To prove (9), it suffices to show that

$$\mathsf{Pr}\{\mathcal{M}_{i,j} | \mathcal{A}_{s,t}\} \geq p^2, \qquad (s,t) \in \Gamma(i,j).$$

And, indeed, for every $(s,t) \in \Gamma(i,j)$ we have

$$\begin{aligned}
\mathsf{Pr}\{\mathcal{M}_{i,j} | \mathcal{A}_{s,t}\} &\geq \mathsf{Pr}\{(\mathcal{X}_{i,j} \cap \mathcal{X}_{s,t}) | \mathcal{A}_{s,t}\} \\
&= \mathsf{Pr}\{\mathcal{X}_{i,j} | \mathcal{A}_{s,t}\} \cdot \mathsf{Pr}\{\mathcal{X}_{s,t} | (\mathcal{X}_{i,j} \cap \mathcal{A}_{s,t})\} \\
&= \mathsf{Pr}\{\mathcal{X}_{i,j} | \mathcal{B}_{i,j}\} \cdot \mathsf{Pr}\{\mathcal{X}_{s,t} | \mathcal{B}_{s,t}\} \\
&= p^2
\end{aligned}$$

where the penultimate equality follows from Lemma 3.1. This proves (9).

Combining (8) and (9) we thus obtain

$$\mathsf{Pr}\{\mathcal{M}_{i,j}\} \geq p^2 \cdot (\mathsf{Pr}\{\mathcal{B}_{i,j}\} - p^{2d-1} \cdot \mathsf{Pr}\{\mathcal{B}_{u,v}\}).$$

Summing the latter inequality over all $(i,j) \in \Delta_{m,n} \setminus \partial\Delta_{m,n}^{(2d)}$ yields

$$\mathsf{E}\{\mathcal{M}\} \geq p^2(1-p^{2d-1}) \cdot \mathsf{E}\{\mathcal{B}\} - O(d(m+n)) \tag{10}$$

with $\mathcal{M}$ standing for the number of locations $(i,j) \in \Delta_{m,n}$ where the event $\mathcal{M}_{i,j}$ occurs.

We now recall that

$$\mathcal{B} \geq mn - \mathcal{C} - \mathcal{R} + \mathcal{M}.$$

Taking expectations, by Lemma 2.2 and (10) we obtain

$$\mathsf{E}\{\mathcal{B}\} \geq mn - \left(2dp - p^2\left(1-p^{2d-1}\right)\right) \cdot \mathsf{E}\{\mathcal{B}\} - O(d(m+n)).$$

Hence,

$$\frac{\mathsf{E}\{\mathcal{B}\}}{mn} \geq \frac{1}{1+2dp-p^2(1-p^{2d-1})} - o_{(\min\{m,n\})/d}(1)$$

and the result follows from Lemma 2.1. □

The probability p that maximizes the coding ratio can be found numerically. Table I presents the improved bit-stuffing lower bounds for $2 \leq d \leq 5$.

For small d's, we can further improve the bound by an even more precise enumeration of patterns that cause an "overlap." The numerical results are in the right column of Table I, and the detailed derivation can be found in the Appendix.

IV. Bit-Stuffing Bounds for $\mathbb{S}_{\mathrm{hex}}^{d,\infty}$

The constraint $\mathbb{S}_{\mathrm{hex}}^{d,\infty}$ consists of all binary arrays in which all rows, columns, and (northeast-to-southwest) diagonals belong to the 1-D (d,∞)-RLL constraint; see [1], [17].

A bit-stuffing encoder that maps unconstrained data into $\mathbb{S}_{\mathrm{hex}}^{d,\infty}(\Delta_{m,n})$ (or into $\mathbb{S}_{\mathrm{hex}}^{d,\infty}(B_{m,n})$) operates similarly to that used for $\mathbb{S}_{\mathrm{sq}}^{d,\infty}$. When a p-biased bit 1 is written into $\Delta_{m,n}$ (or $B_{m,n}$), we should stuff 0's into the d positions to the right, the d positions below, and the d positions along the diagonal below the bit 1.

We note that the bit-stuffing encoder for $\mathbb{S}_{\mathrm{hex}}^{1,\infty}$ is a special case of the encoder analyzed in [29]. We can apply the result in [29] directly to compute the maximum coding ratio of 0.47868. Baxter [1] has derived the exact value of $\mathsf{cap}(\mathbb{S}_{\mathrm{hex}}^{1,\infty})$, which, to nine decimal places, is 0.480767622. The maximum coding ratio of the bit-stuffing encoder is only 0.5% below the capacity.

For $d \geq 2$, we can obtain a simple lower bound on the rate of a bit-stuffing encoder for $\mathbb{S}_{\mathrm{hex}}^{d,\infty}$ following a similar procedure as in Section II. Specifically, denote by $\mu_{m,n} = \mu_{\mathrm{hex};m,n}^{d,\infty}$ the probability measure induced by the bit-stuffing encoder on $\mathbb{S}_{\mathrm{hex}}^{d,\infty}(\Delta_{m,n})$; the next lemma is then a counterpart of Lemma 2.3 for the constraint $\mathbb{S}_{\mathrm{hex}}^{d,\infty}$.

Lemma 4.1: For $d > 0$ and any $0 \leq p \leq 1$

$$\mathsf{H}(\mu_{m,n}) \geq \frac{h(p)}{1+3dp}.$$

We can further improve the lower bound by accounting for the "double-stuffing" event. Let the notations $\mathcal{X}_{i,j}$, $\mathcal{C}_{i,j}$, $\mathcal{R}_{i,j}$, be as in Section II. Similarly, define

$$\mathcal{D}_{i,j} = \bigcup_{s=1}^{d} \mathcal{X}_{i-s,j+s}$$

and

$$\mathcal{B}_{i,j}^{\mathrm{hex}} = \overline{\mathcal{C}_{i,j} \cup \mathcal{R}_{i,j} \cup \mathcal{D}_{i,j}}.$$

That is, $\mathcal{B}_{i,j}^{\mathrm{hex}}$ is the event that location (i,j) is filled by a bit of the p-biased sequence on a hexagonal lattice. We also define

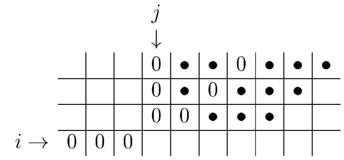


Fig. 6. Event $\mathcal{B}_{i,j}^{\mathrm{hex}}$ and the region $\Gamma_{\mathrm{hex}}(i,j)$ for $d = 3$.

the region $\Gamma_{\mathrm{hex}}(i,j)$ at the bottom of the page. Fig. 6 depicts the event $\mathcal{B}_{i,j}^{\mathrm{hex}}$ for $d = 3$, and the entries indexed by $\Gamma_{\mathrm{hex}}(i,j)$ are marked in the figure by thick dots.

The following lemma plays the role of Lemma 3.2 for the constraint $\mathbb{S}_{\mathrm{hex}}^{d,\infty}$.

Lemma 4.2: For $(i,j) \in \Delta_{m,n} \setminus \partial\Delta_{m,n}^{(2d)}$

$$\mathcal{B}_{i,j}^{\mathrm{hex}} = \mathcal{B}_{i,j}^{\mathrm{hex}} \cap \left(\bigcup_{(s,t)\in\Gamma_{\mathrm{hex}}(i,j)} \mathcal{B}_{s,t}^{\mathrm{hex}}\right).$$

Proof: We will prove the following equivalent relationship:

$$\mathcal{B}_{i,j}^{\mathrm{hex}} \cap \overline{\left(\bigcup_{(s,t)\in\Gamma_{\mathrm{hex}}(i,j)} \mathcal{B}_{s,t}^{\mathrm{hex}}\right)} = \emptyset.$$

Let X be an array that belongs to the left-hand side of the above equation. Region $\Gamma_{\mathrm{hex}}(i,j)$ in X should be all stuffed 0's. Let us consider entry $X_{i-1,j+d}$. It can be stuffed either due to the event $\mathcal{X}_{i-d-1,j+d}$ or the event $\mathcal{X}_{i-d-1,j+2d}$.

If $X_{i-1,j+d}$ is stuffed due to $\mathcal{X}_{i-d-1,j+d}$, then $X_{i-1,j+d+1}$ can only be stuffed due to $\mathcal{X}_{i-d-1,j+2d+1}$, as shown in Fig. 7(a). However, $X_{i-2,j+d+1}$ cannot be stuffed in this pattern. On the other hand, if $X_{i-1,j+d}$ is stuffed due to $\mathcal{X}_{i-d-1,j+2d}$, then $X_{i-1,j+d+1}$ cannot be stuffed, as shown in Fig. 7(b). Therefore, we conclude that

$$\mathcal{B}_{i,j}^{\mathrm{hex}} \cap \overline{\left(\bigcup_{(s,t)\in\Gamma_{\mathrm{hex}}(i,j)} \mathcal{B}_{s,t}^{\mathrm{hex}}\right)} = \emptyset. \quad \square$$

Proposition 4.3:

$$\mathsf{H}(\mu_{m,n}) \geq \max_{0<p<1} \frac{h(p)}{1+3dp-p^2} - o(\min\{m,n\})/d(1).$$

Proof: Define the "double-stuffing" event

$$\mathcal{M}_{i,j}^{\mathrm{hex}} = \mathcal{X}_{i,j} \cap \left(\bigcup_{(s,t)\in\Gamma_{\mathrm{hex}}(i,j)} \mathcal{X}_{s,t}\right).$$

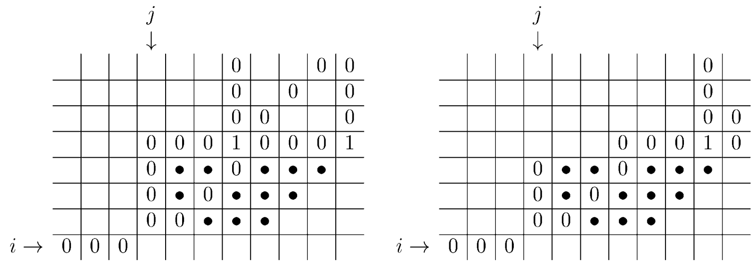
It is clear by definition that $\mathcal{M}_{i,j}^{\mathrm{hex}} \subseteq \mathcal{B}_{i,j}^{\mathrm{hex}}$ and so

$$\mathsf{Pr}\left\{\mathcal{M}_{i,j}^{\mathrm{hex}}\right\} = \mathsf{Pr}\left\{\mathcal{M}_{i,j}^{\mathrm{hex}} | \mathcal{B}_{i,j}^{\mathrm{hex}}\right\} \cdot \mathsf{Pr}\left\{\mathcal{B}_{i,j}^{\mathrm{hex}}\right\}$$

(compare with (8)). Applying Lemma 4.2 and proceeding as in the proof of Proposition 3.3, we can obtain the following counterpart of (9):

$$\mathsf{Pr}\left\{\mathcal{M}_{i,j}^{\mathrm{hex}} | \mathcal{B}_{i,j}^{\mathrm{hex}}\right\} \geq p^2.$$

$$\Gamma_{\mathrm{hex}}(i,j) = \left\{(s,t) \in \mathbb{Z}^2 : i-d \leq s < i \text{ and } j < t \leq j+d+i-s\right\} \setminus \left\{(i-a, j+a)\right\}_{a=1}^{d}.$$



(a) (b)

Fig. 7. Patterns for the proof of Lemma 4.2 for $d = 3$.

Hence,

$$\mathsf{Pr}\left\{\mathcal{M}_{i,j}^{\text{hex}}\right\} \geq p^2 \cdot \mathsf{Pr}\left\{\mathcal{B}_{i,j}^{\text{hex}}\right\}$$

and by summing the latter inequality over all $(i,j) \in \Delta_{m,n} \setminus \partial\Delta_{m,n}^{(2d)}$ we obtain

$$\mathsf{E}\left\{\mathcal{M}^{\text{hex}}\right\} \geq p^2 \cdot \mathsf{E}\left\{\mathcal{B}^{\text{hex}}\right\} - O(d(m+n))$$

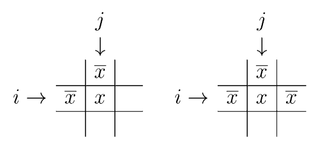
with $\mathcal{B}^{\text{hex}}$ (respectively, $\mathcal{M}^{\text{hex}}$) standing for the number of locations $(i,j) \in \Delta_{m,n}$ where the event $\mathcal{B}_{i,j}^{\text{hex}}$ (respectively, $\mathcal{M}_{i,j}^{\text{hex}}$) occurs (compare with (10)). The result is finally deduced by following along the remaining lines of the proof of Proposition 3.3. □

The probability p that maximizes the coding ratio is found numerically and the resulting lower bounds on the capacity are summarized in Table II for $2 \leq d \leq 5$. Also shown in the table are the numerically computed values of the lower bounds presented in [17].

As in the case of the constraint $\mathbb{S}_{\text{sq}}^{d,\infty}$, we can tighten the lower bound by enumerating certain patterns that cause an overlap of stuffed bits. The numerical result for $d = 2$ is shown in Table II, and details of the derivation are presented in the Appendix.

TABLE II
LOWER BOUNDS ON THE RATE OF A BIT-STUFFING ENCODER FOR $\mathbb{S}_{\text{hex}}^{d,\infty}$

d	From [17]	Proposition 4.3	Improved Bound
2	0.3333	0.3347	0.3387
3	0.2000	0.2630	-
4	0.2000	0.2196	-
5	0.1428	0.1901	-



(a) (b)

Fig. 8. (a) Event $\mathcal{F}_{i,j}$ and (b) event $\mathcal{N}_{i,j}$.

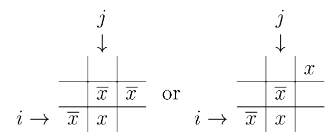


Fig. 9. Event $\mathcal{L}_{i,j}$.

V. BIT-STUFFING BOUNDS FOR $\mathbb{S}_{\text{nib}}$

The description of our bit-stuffing encoder for $\mathbb{S}_{\text{nib}}$ makes use of the following definitions.

Let X be a random element in $\mathbb{S}_{\text{nib}}(\Delta_{m,n})$. For $(i,j) \in \Delta_{m,n}$, denote by $\mathcal{H}_{i,j}$ the event $X_{i,j} = X_{i,j-1}$ and by $\mathcal{V}_{i,j}$ the event $X_{i,j} = X_{i-1,j}$. We hereafter assume that $X_{i,j} = 0$ whenever $i < 0$ or $i+j < 0$; hence, for $i < 0$ (respectively, $i+j < 0$), the event $\mathcal{V}_{i,j}$ (respectively, $\mathcal{H}_{i,j}$) holds for each element $X \in \mathbb{S}_{\text{nib}}(\Delta_{m,n})$.

Also define

$$\mathcal{F}_{i,j} = \overline{\mathcal{H}_{i,j} \cup \mathcal{V}_{i,j}}$$
$$\mathcal{N}_{i,j} = \overline{\mathcal{H}_{i,j} \cup \mathcal{V}_{i,j} \cup \mathcal{H}_{i,j+1}}$$

and

$$\mathcal{L}_{i,j} = \mathcal{F}_{i,j} \setminus \mathcal{F}_{i-1,j+1}.$$

The events $\mathcal{F}_{i,j}$ and $\mathcal{N}_{i,j}$ correspond to the patterns in Fig. 8(a) and (b), respectively, where $x \in \{0,1\}$ and $\overline{x}$ stands for the complement of x. The event $\mathcal{L}_{i,j}$ stands for any of the two patterns in Fig. 9.

The bit-stuffing encoder is fed by two streams of independent Bernoulli random variables: the first stream consists of unbiased bits (fair coins), while in the second stream, the probability of having 0 is $\frac{2}{3}$ (the latter stream is generated by applying a respective distribution transformer on bits of the user data sequence). The bit-stuffing encoder generates an output $X \in \mathbb{S}_{\text{nib}}(\Delta_{m,n})$, diagonal by diagonal (or row by row), by assuming that $X_{i,j} = 0$ whenever $i < 0$ or $i+j < 0$ and applying the following rule to every location $(i,j) \in \Delta_{m,n}$.

NIB-1: If $\mathcal{L}_{i,j-1}$ occurs then $X_{i,j}$ is set to $X_{i,j-1}$ with probability $\frac{2}{3}$; i.e., the event $\mathcal{H}_{i,j}$ will occur depending on the biased stream.

NIB-2: If $\mathcal{N}_{i-1,j}$ occurs then $X_{i,j}$ is set to the value of $X_{i-1,j}$; i.e., the event $\mathcal{V}_{i,j}$ is forced.

NIB-3: Otherwise, $X_{i,j}$ is set to 1 with probability $\frac{1}{2}$; i.e., the event $\mathcal{H}_{i,j}$ (alternatively, $\mathcal{V}_{i,j}$) will occur depending on the unbiased stream.

Denote by $\mu_{m,n} = \mu_{\text{nib};m,n}$ the probability measure on $\mathbb{S}_{\text{nib}}(\Delta_{m,n})$ that is induced by the bit-stuffing encoder. Similarly to what we had in Section II, the measure $\mu_{m,n}(x)$ takes for every $x \in \mathbb{S}_{\text{nib}}(\Delta_{m,n})$ the form

$$\mu_{m,n}(x) = \prod_{i,j \in \Delta_{m,n}} \vartheta_{\text{nib}}\left(x_{i,j} | x_{i,j-1}, x_{i,j-2}, x_{i-1,j}, x_{i-1,j-1}, x_{i-2,j}, x_{i-1,j+1}\right) \quad (11)$$

where $\vartheta_{\rm nib} : \{0,1\}^7 \to [0,1]$ is given by

$$\vartheta_{\rm nib}(x|x,y,z,u,v,w) = \begin{cases} \frac{2}{3}, & \text{if } u=y\neq x \text{ and either } u\neq v \text{ or } u=z \\ 1, & \text{if } u=v=w\neq z \\ \frac{1}{2}, & \text{otherwise} \end{cases}$$

and $\vartheta_{\rm nib}(\overline{x}|x,\cdot) = 1-\vartheta_{\rm nib}(x|x,\cdot)$. Furthermore, similarly to the encoders of Sections II–IV, the coding rule of the encoder here (yet not necessarily the measure $\mu_{m,n}$!) is *shift-invariant* in the sense that it does not depend on the particular location $(i,j)\in\Delta_{m,n}$.

The rest of this section is devoted to proving the following lower bound on $\mathsf{H}(\mu_{m,n})$.

Proposition 5.1:

$$\mathsf{H}(\mu_{m,n}) > 0.9017 - o_{\min\{m,n\}}(1).$$

Let $\mathcal{F}$ (respectively, $\mathcal{L}$) be the number of locations $(i,j)\in\Delta_{m,n}$ in which the event $\mathcal{F}_{i,j}$ (respectively, $\mathcal{L}_{i,j}$) occurs within a random element $X\in\mathbb{S}_{\rm nib}(\Delta_{m,n})$. Proposition 5.1 will be proved by first showing a lower bound on $\mathsf{H}(\mu_{m,n})$ in terms of $\mathsf{E}\{\mathcal{F}\}$ and $\mathsf{E}\{\mathcal{L}\}$. The proof will then continue with obtaining upper bounds on $\mathsf{E}\{\mathcal{F}\}$ and $\mathsf{E}\{\mathcal{L}\}$, using simple properties of the probability measure $\mu_{m,n}$.

We mention that the technique presented here can be refined to obtain the stronger bound

$$\mathsf{H}(\mu_{m,n}) \geq 0.91276 - o_{\min\{m,n\}}(1) \tag{12}$$

yet the proof of this inequality is rather long and hence omitted; the full proof can be found in [10].

For a nonnegative integer $r<n$, let

$$U_r = \Delta_{m,n}\cap\left\{(s,t)\in\mathbb{Z}^2 : s+t\leq r\right\}.$$

The random U_{i+j}-configuration $X(U_{i+j})$ (namely, the restriction of X to U_{i+j}) completely determines whether the event $\mathcal{F}_{i,j}$ occurs; hence, we can regard the latter event as the set of all U_{i+j}-configurations y that imply $\mathcal{F}_{i,j}$.

The following sequence of lemmas presents several properties of the probability measure $\mu_{m,n}$; these properties will lead to the proof of Proposition 5.1.

Lemma 5.2: The following holds for every location $(i,j)\in\Delta_{m-1,n-1}$:

a) $\mathsf{Pr}\{\mathcal{H}_{i,j+1} \mid X(U_{i+j})\in\mathcal{F}_{i,j}\} = \frac{2}{3}$;

b) $\mathsf{Pr}\{\mathcal{V}_{i+1,j} \mid X(U_{i+j})\in\mathcal{F}_{i,j}\} = \frac{2}{3}$;

c) $\mathsf{Pr}\left\{\mathcal{H}_{i,j+1} \mid X(U_{i+j})\in\overline{\mathcal{F}_{i,j}\cup\mathcal{F}_{i-1,j+1}}\right\} = \frac{1}{2}$;

d) $\mathsf{Pr}\{\mathcal{H}_{i,j+1}\cap\mathcal{V}_{i+1,j} \mid X(U_{i+j})\in\mathcal{F}_{i,j}\} = \frac{1}{3}$;

e) $\mathsf{Pr}\{\mathcal{H}_{i,j+1}\cup\mathcal{V}_{i+1,j} \mid X(U_{i+j})\in\mathcal{F}_{i,j}\} = 1$.

Proof: We start with part a) and fix a diagonal r. We prove by induction on i that $\mathsf{Pr}\{\mathcal{H}_{i,r+1-i} \mid X(U_r)=y\} = \frac{2}{3}$ for every U_r-configuration $y\in\mathcal{F}_{i,r-i}$. The proof distinguishes between two cases, where the first case serves also as the induction basis; hereafter Y_r stands for $X(U_r)$.

Case 1. $y\in\mathcal{F}_{i,r-i}\cap\overline{\mathcal{F}}_{i-1,r+1-i}$. In this case, y belongs to $\mathcal{L}_{i,r-i}$; so, by Step NIB-1 of the encoding process we obtain $\mathsf{Pr}\{\mathcal{H}_{i,r+1-i} \mid Y_r=y\} = \frac{2}{3}$.

Case 2. $y\in\mathcal{F}_{i,r-i}\cap\mathcal{F}_{i-1,r+1-i}$. Write

$$\begin{aligned}\mathsf{Pr}\{\mathcal{H}_{i,r+1-i} \mid Y_r=y\} &= \mathsf{Pr}\{\mathcal{H}_{i,r+1-i} \mid (Y_r=y)\cap\mathcal{H}_{i-1,r+2-i}\} \\ &\quad\cdot\mathsf{Pr}\{\mathcal{H}_{i-1,r+2-i} \mid Y_r=y\} \\ &\quad+\mathsf{Pr}\left\{\mathcal{H}_{i,r+1-i} \mid (Y_r=y)\cap\overline{\mathcal{H}}_{i-1,r+2-i}\right\} \\ &\quad\cdot\mathsf{Pr}\left\{\overline{\mathcal{H}}_{i-1,r+2-i} \mid Y_r=y\right\}.\end{aligned}$$

By the induction hypothesis we have

$$\mathsf{Pr}\{\mathcal{H}_{i-1,r+2-i} \mid Y_r=y\} = 1-\mathsf{Pr}\left\{\overline{\mathcal{H}}_{i-1,r+2-i} \mid Y_r=y\right\} = \frac{2}{3}.$$

Observing that $y\in\mathcal{F}_{i,r-i}\cap\mathcal{F}_{i-1,r+1-i}$ implies $y_{i,r-i} = y_{i-1,r+1-i}$, we obtain by Step NIB-2 of the encoding process that

$$\mathsf{Pr}\left\{\mathcal{H}_{i,r+1-i} \mid (Y_r=y)\cap\overline{\mathcal{H}}_{i-1,r+2-i}\right\} = 1$$

and by Step NIB-3 that

$$\mathsf{Pr}\{\mathcal{H}_{i,r+1-i} \mid (Y_r=y)\cap\mathcal{H}_{i-1,r+2-i}\} = \frac{1}{2}.$$

It follows from the last four equations that

$$\mathsf{Pr}\{\mathcal{H}_{i,r+1-i} \mid Y_r=y\} = \frac{1}{2}\cdot\frac{2}{3}+1\cdot\frac{1}{3} = \frac{2}{3}$$

thus completing the proof of part a).

Now, from part a) and Step NIB-3 we have

$$\begin{aligned}\mathsf{Pr}\{\mathcal{V}_{i+1,j}\cap\mathcal{H}_{i,j+1} \mid Y_{i+j}\in\mathcal{F}_{i,j}\} &= \mathsf{Pr}\{\mathcal{V}_{i+1,j} \mid (Y_{i+j}\in\mathcal{F}_{i,j})\cap\mathcal{H}_{i,j+1}\} \\ &\quad\cdot\mathsf{Pr}\{\mathcal{H}_{i,j+1} \mid Y_{i+j}\in\mathcal{F}_{i,j}\} = \frac{1}{2}\cdot\frac{2}{3} = \frac{1}{3}\end{aligned}$$

and from Step NIB-2

$$\begin{aligned}\mathsf{Pr}\left\{\mathcal{V}_{i+1,j}\cap\overline{\mathcal{H}}_{i,j+1} \mid Y_{i+j}\in\mathcal{F}_{i,j}\right\} &= \mathsf{Pr}\left\{\mathcal{V}_{i+1,j} \mid (Y_{i+j}\in\mathcal{F}_{i,j})\cap\overline{\mathcal{H}}_{i,j+1}\right\} \\ &\quad\cdot\mathsf{Pr}\left\{\overline{\mathcal{H}}_{i,j+1} \mid Y_{i+j}\in\mathcal{F}_{i,j}\right\} = 1\cdot\frac{1}{3} = \frac{1}{3}.\end{aligned}$$

The last two equations yield parts b), d), and e). Finally, part c) follows immediately from Step NIB-3. □

Lemma 5.3: For every location $(i,j)\in\Delta_{m-1,n-1}$ and U_{i+j}-configuration $y\in\overline{\mathcal{F}}_{i,j}$

$$\begin{aligned}\mathsf{Pr}\{X_{i,j+1}=X_{i+1,j} \mid X(U_{i+j})=y\} &= \sum_{x\in\{0,1\}}\big(\mathsf{Pr}\{X_{i,j+1}=x \mid X(U_{i+j})=y\} \\ &\qquad\cdot\mathsf{Pr}\{X_{i+1,j}=x \mid X(U_{i+j})=y\}\big).\end{aligned}$$

Proof: Given that $X(U_{i+j})=y$ for some fixed $y\in\overline{\mathcal{F}}_{i,j}$, the value $X_{i+1,j}$ is determined by one of the steps, NIB-1 or

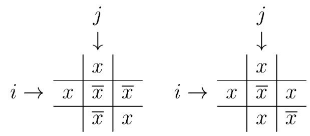


Fig. 10. Two events in Lemma 5.4.

NIB-3, of the encoding process. In either case, the value of $X_{i+1,j}$ is statistically independent of $X_{i,j+1}$. □

Lemma 5.4: For every $(i,j) \in \Delta_{m-1,n-2}$

$$\mathsf{Pr}\left\{\mathcal{F}_{i+1,j+1} \mid \mathcal{F}_{i,j}\right\} = \frac{1}{6}.$$

Proof: We first observe that $\mathcal{F}_{i+1,j+1} \cap \mathcal{F}_{i,j}$ is a union of two events

$$\mathcal{H}_{i,j+1} \cap \mathcal{V}_{i+1,j} \cap \overline{\mathcal{H}}_{i+1,j+1} \cap \mathcal{F}_{i,j}$$

and

$$(\overline{\mathcal{H}_{i,j+1} \cup \mathcal{V}_{i+1,j}}) \cap \overline{\mathcal{H}}_{i+1,j+1} \cap \mathcal{F}_{i,j}$$

(see Fig. 10). Yet, by Lemma 5.2, part e), the second event has probability 0. Therefore,

$$\begin{aligned}\mathsf{Pr}\left\{\mathcal{F}_{i+1,j+1} \mid \mathcal{F}_{i,j}\right\} &= \mathsf{Pr}\left\{\overline{\mathcal{H}}_{i+1,j+1} \mid \mathcal{H}_{i,j+1} \cap \mathcal{V}_{i+1,j} \cap \mathcal{F}_{i,j}\right\}\\ &\quad\cdot \mathsf{Pr}\left\{\mathcal{H}_{i,j+1} \cap \mathcal{V}_{i+1,j} \mid \mathcal{F}_{i,j}\right\}.\end{aligned}$$

Now, by Lemma 5.2, part d) we have

$$\mathsf{Pr}\left\{\mathcal{H}_{i,j+1} \cap \mathcal{V}_{i+1,j} \mid \mathcal{F}_{i,j}\right\} = \frac{1}{3}$$

and from the inclusion $(\mathcal{H}_{i,j+1} \cap \mathcal{V}_{i+1,j}) \subseteq \overline{\mathcal{F}_{i,j+1} \cup \mathcal{F}_{i+1,j}}$ and Lemma 5.2, part c) we get

$$\mathsf{Pr}\left\{\overline{\mathcal{H}}_{i+1,j+1} \mid \mathcal{H}_{i,j+1} \cap \mathcal{V}_{i+1,j} \cap \mathcal{F}_{i,j}\right\} = \frac{1}{2}.$$

The last three equations imply the result. □

Proposition 5.5:

$$\mathsf{H}(\mu_{m,n}) = 1 - \frac{\left(1-h\left(\frac{2}{3}\right)\right)\cdot \mathsf{E}\{\mathcal{L}\} + \frac{1}{3}\cdot \mathsf{E}\{\mathcal{F}\}}{mn} - o_{\min\{m,n\}}(1).$$

Proof: By the expression for $\mu_{m,n}$ in (11) we have

$$\begin{aligned}mn\cdot \mathsf{H}(\mu_{m,n}) &= h\left(\frac{2}{3}\right)\cdot\left(\sum_{i,j}\mathsf{Pr}\{\mathcal{L}_{i,j-1}\}\right)\\ &\quad + h\left(\frac{1}{2}\right)\cdot\left(\sum_{i,j}\mathsf{Pr}\left\{\overline{\mathcal{L}_{i,j-1}\cup\mathcal{N}_{i-1,j}}\right\}\right)\\ &= h\left(\frac{2}{3}\right)\cdot\left(\sum_{i,j}\mathsf{Pr}\{\mathcal{L}_{i,j-1}\}\right)\\ &\quad + \left(\sum_{i,j}\left(1-\mathsf{Pr}\{\mathcal{L}_{i,j-1}\}-\mathsf{Pr}\{\mathcal{N}_{i-1,j}\}\right)\right)\end{aligned}$$

where (i,j) ranges in the summations over the elements of $\Delta_{m,n}$. Now, from Lemma 5.2, part a) we get

$$\begin{aligned}\mathsf{Pr}\{\mathcal{N}_{i,j}\} &= \mathsf{Pr}\{\mathcal{F}_{i,j}\cap\overline{\mathcal{H}}_{i,j+1}\}\\ &= \mathsf{Pr}\left\{\overline{\mathcal{H}}_{i,j+1}\mid\mathcal{F}_{i,j}\right\}\cdot\mathsf{Pr}\{\mathcal{F}_{i,j}\}\\ &= \mathsf{Pr}\{\mathcal{F}_{i,j}\}\cdot\frac{1}{3}.\end{aligned}$$

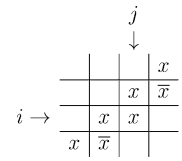


Fig. 11. Event $\mathcal{Q}_{i,j}$.

Hence,

$$\begin{aligned}mn\cdot\mathsf{H}(\mu_{m,n}) &= h\left(\frac{2}{3}\right)\cdot\left(\sum_{i,j}\mathsf{Pr}\{\mathcal{L}_{i,j-1}\}\right)\\ &\quad + \left(\sum_{i,j}\left(1-\mathsf{Pr}\{\mathcal{L}_{i,j-1}\}-\frac{1}{3}\cdot\mathsf{Pr}\{\mathcal{F}_{i-1,j}\}\right)\right)\\ &= mn-\left(1-h\left(\frac{2}{3}\right)\right)\cdot\mathsf{E}\{\mathcal{L}\}\\ &\quad -\frac{1}{3}\cdot\mathsf{E}\{\mathcal{F}\}-O(m+n)\end{aligned}$$

thereby yielding the result. □

We next turn to obtaining an upper bound on $\mathsf{E}\{\mathcal{F}\}$; such a bound, combined with the inequality $\mathcal{L} \le \mathcal{F}$ and with Proposition 5.5, will lead to a lower bound on $\mathsf{H}(\mu_{m,n})$.

For $(i,j) \in \Delta_{m-1,n}$ let the event $\mathcal{Q}_{i,j}$ be defined by

$$\mathcal{Q}_{i,j} = \mathcal{H}_{i,j}\cap\mathcal{V}_{i,j}\cap\mathcal{F}_{i-1,j+1}\cap\mathcal{F}_{i+1,j-1}$$

(see Fig. 11)

Lemma 5.6: For $(i,j) \in \Delta_{m-1,n-2}$

$$\mathsf{Pr}\{\mathcal{F}_{i+1,j+1}\mid\mathcal{Q}_{i,j}\} = \frac{5}{18}.$$

Proof: Let y be a particular U_{i+j}-configuration in $\mathcal{Q}_{i,j}$. By Lemma 5.2, parts a) and b), we have

$$\begin{aligned}\mathsf{Pr}\left\{X_{i,j+1}=y_{i-1,j+1}\mid X(U_{i+j})=y\right\} &= \mathsf{Pr}\left\{X_{i+1,j}=y_{i+1,j-1}\mid X(U_{i+j})=y\right\}\\ &= \frac{2}{3}.\end{aligned}$$

Noting that $y \in \overline{\mathcal{F}}_{i,j}$ and $y_{i-1,j+1} = y_{i+1,j-1}$, it follows by Lemma 5.3 that

$$\mathsf{Pr}\left\{X_{i,j+1}=X_{i+1,j}\mid X(U_{i+j})=y\right\} = \left(\frac{2}{3}\right)^2+\left(\frac{1}{3}\right)^2 = \frac{5}{9}.$$

In addition, $X(U_{i+j}) \in \mathcal{Q}_{i,j}$ implies that

$$X(U_{i+j+1}) \in \overline{\mathcal{F}_{i+1,j}\cup\mathcal{F}_{i,j+1}};$$

hence, by applying Lemma 5.2, part c) to $X(U_{i+j+1})$, we get

$$\mathsf{Pr}\left\{\overline{\mathcal{H}}_{i+1,j+1}\mid(X_{i,j+1}=X_{i+1,j})\cap(X(U_{i+j})=y)\right\} = \frac{1}{2}.$$

The result follows from the last two equations and by recalling that the event $\mathcal{F}_{i+1,j+1}$ is identical to

$$(X_{i,j+1}=X_{i+1,j})\cap\overline{\mathcal{H}}_{i+1,j+1}.$$ □

Lemma 5.7: For $(i,j) \in \Delta_{m-1,n-1}$

$$\mathsf{Pr}\left\{X_{i,j+1}=X_{i+1,j}\mid\overline{\mathcal{F}_{i,j}\cup\mathcal{Q}_{i,j}}\right\} \le \frac{1}{2}.$$

Proof: Let y be a particular U_{i+j}-configuration in $\overline{\mathcal{F}_{i,j}\cup\mathcal{Q}_{i,j}}$. We distinguish between three cases.

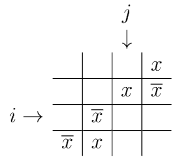


Fig. 12. Event $(\overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}) \cap \mathcal{F}_{i-1,j+1} \cap \mathcal{F}_{i+1,j-1}$.

Case 1. $y \in \overline{\mathcal{F}}_{i-1,j+1}$. By Lemma 5.2 part c) we get

$$\mathsf{Pr}\{X_{i,j+1} = 0 \mid X(U_{i+j}) = y\} = \frac{1}{2}$$

and by Lemma 5.3 we obtain

$$\mathsf{Pr}\{X_{i,j+1} = X_{i+1,j} \mid X(U_{i+j}) = y\} = \frac{1}{2}.$$

Case 2. $y \in \overline{\mathcal{F}}_{i+1,j-1}$. Here

$$\mathsf{Pr}\{X_{i+1,j} = 0 \mid X(U_{i+j}) = y\} = \frac{1}{2}$$

and we get the same result as in Case 1.

Case 3. $y \in \mathcal{F}_{i-1,j+1} \cap \mathcal{F}_{i+1,j-1}$. By Lemma 5.2, parts a) and b)

$$\begin{aligned}\mathsf{Pr}\{X_{i+1,j} = y_{i+1,j-1} \mid X(U_{i+j}) = y\} \\ = \mathsf{Pr}\{X_{i,j+1} = y_{i-1,j+1} \mid X(U_{i+j}) = y\} \\ = \frac{2}{3}.\end{aligned}$$

On the other hand, $y \in \overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}$ implies $y_{i+1,j-1} \neq y_{i-1,j+1}$ (see Fig. 12); we thus obtain from Lemma 5.3 that

$$\mathsf{Pr}\{X_{i,j+1} = X_{i+1,j} \mid X(U_{i+j}) = y\} = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9} < \frac{1}{2}$$

thus completing the proof. □

Lemma 5.8: For $(i,j) \in \Delta_{m-1,n-2}$

$$\mathsf{Pr}\{\mathcal{F}_{i+1,j+1} \mid \overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}\} \le \frac{1}{4}.$$

Proof: Recall that $\mathcal{F}_{i+1,j+1}$ equals $(X_{i,j+1} = X_{i+1,j}) \cap \overline{\mathcal{H}}_{i+1,j+1}$ and write

$$\begin{aligned}&\mathsf{Pr}\{\mathcal{F}_{i+1,j+1} \mid \overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}\} \\ &\quad = \mathsf{Pr}\{\overline{\mathcal{H}}_{i+1,j+1} \mid (X_{i,j+1} = X_{i+1,j}) \cap (\overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}})\} \\ &\qquad \cdot \mathsf{Pr}\{X_{i,j+1} = X_{i+1,j} \mid \overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}\} \\ &\quad \le \frac{1}{2} \cdot \mathsf{Pr}\{\overline{\mathcal{H}}_{i+1,j+1} \mid (X_{i,j+1} = X_{i+1,j}) \cap (\overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}})\} \\ &\quad \le \frac{1}{2} \cdot \max_y \mathsf{Pr}\{\overline{\mathcal{H}}_{i+1,j+1} \mid X(U_{i+j+1}) = y\}\end{aligned}$$

where the first inequality follows from Lemma 5.7 and y ranges over all U_{i+j+1}-configurations in $\overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}$ such that $y_{i,j+1} = y_{i+1,j}$. By Lemma 5.2, parts a)–c) we have for every such y

$$\begin{aligned}&\mathsf{Pr}\{\overline{\mathcal{H}}_{i+1,j+1} \mid X(U_{i+j+1}) = y\} \\ &\qquad = \mathsf{Pr}\{\overline{\mathcal{V}}_{i+1,j+1} \mid X(U_{i+j+1}) = y\} \in \left\{\frac{1}{3}, \frac{1}{2}\right\}\end{aligned}$$

thereby implying the result. □

We are now ready to prove an upper bound on $\mathsf{E}\{\mathcal{F}\}$.

Proposition 5.9:

$$\mathsf{E}\{\mathcal{F}\} \le \frac{9}{38} \cdot mn + O(m+n).$$

Proof: Noting that $\mathcal{F}_{i,j} \cap \mathcal{Q}_{i,j} = \emptyset$, for every $(i,j) \in \Delta_{m-1,n-2}$ we have

$$\begin{aligned}\mathsf{Pr}\{\mathcal{F}_{i+1,j+1}\} &= \sum\nolimits_{\mathcal{A} \in \{\mathcal{F}_{i,j}, \mathcal{Q}_{i,j}, \overline{\mathcal{F}_{i,j} \cup \mathcal{Q}_{i,j}}\}} \\ &\quad \cdot \left(\mathsf{Pr}\{\mathcal{F}_{i+1,j+1} \mid \mathcal{A}\} \cdot \mathsf{Pr}\{\mathcal{A}\}\right) \\ &\le \frac{1}{6} \cdot \mathsf{Pr}\{\mathcal{F}_{i,j}\} + \frac{5}{18} \cdot \mathsf{Pr}\{\mathcal{Q}_{i,j}\} \\ &\quad + \frac{1}{4} \cdot (1 - \mathsf{Pr}\{\mathcal{F}_{i,j}\} - \mathsf{Pr}\{\mathcal{Q}_{i,j}\}) \\ &= \frac{1}{36} \cdot \mathsf{Pr}\{\mathcal{Q}_{i,j}\} - \frac{1}{12} \cdot \mathsf{Pr}\{\mathcal{F}_{i,j}\} + \frac{1}{4} \\ &\le \frac{1}{36} \cdot \mathsf{Pr}\{\mathcal{F}_{i+1,j-1}\} - \frac{1}{12} \cdot \mathsf{Pr}\{\mathcal{F}_{i,j}\} + \frac{1}{4}\end{aligned}$$

where the first inequality follows from Lemmas 5.4, 5.6, and 5.8, and the second inequality follows from the inclusion $\mathcal{Q}_{i,j} \subseteq \mathcal{F}_{i+1,j-1}$. Summing over $(i,j) \in \Delta_{m-1,n-2}$ we obtain

$$\mathsf{E}\{\mathcal{F}\} \le \frac{1}{36} \cdot \mathsf{E}\{\mathcal{F}\} - \frac{1}{12} \cdot \mathsf{E}\{\mathcal{F}\} + \frac{1}{4} \cdot mn + O(m+n).$$

The result follows. □

Proof of Proposition 5.1: Combine Propositions 5.5 and 5.9 with the inequality $\mathcal{L} \le \mathcal{F}$. □

The stronger bound (12) is obtained in [10] through Proposition 5.5, using an improved version of Proposition 5.9 and showing that $\mathsf{E}\{\mathcal{L}\}$ is bounded from above by $\frac{73}{81}\mathsf{E}\{\mathcal{F}\}$.

By looking at the growth rate of $|\mathbb{S}(B_{m,n})|$ for fixed m while n increases, one can easily obtain an upper bound on the capacity of a 2-D constraint; see Weeks and Blahut [34]. Applying this method to $\mathbb{S}_{\rm nib}$ with $m = 9$ yields an upper bound of 0.93965. Similarly, one can obtain lower bounds on $\mathsf{cap}(\mathbb{S}_{\rm nib})$ by considering the growth rate (with n) of the number of elements in $\mathbb{S}(B_{m,n})$ that can be freely concatenated vertically while satisfying the constraint [2]. Thus, we can obtain a sequence of upper bounds and a sequence of lower bounds on the capacity of the constraint, as a function of m. By computing those sequences for $\mathbb{S}_{\rm nib}$ up to $m = 9$ and applying a first-order Richardson extrapolation to each sequence [34], we have discovered that the extrapolated values agree in their first ten decimal places with $0.9238294367\ldots$, and we conjecture that so does the value $\mathsf{cap}(\mathbb{S}_{\rm nib})$.

VI. Quasi-Stationary Measures

The probability measure $\mu_{m,n}$ induced by the bit-stuffing encoders in Sections II–V does not seem to possess any stationary (shift-invariant) properties. Still, since the *coding rule* is shift-invariant, one can guarantee a "quasi-stationary"-induced measure by a proper initialization of the boundary entries in the generated output array. We show this next.

Given any subset $U \subseteq \mathbb{Z}^2$, denote by $\sigma_{r,s}(U)$ the shifted subset

$$\sigma_{r,s}(U) = \{(i+r, j+s) : (i,j) \in U\}$$

and by $-U$ the inverted subset

$$-U = \{(-i,-j) : (i,j) \in U\}.$$

For a U-configuration x we let $\sigma_{r,s}(x)$ denote the shifted $\sigma_{r,s}(U)$-configuration y, where $y_{i+r,j+s} = x_{i,j}$ for every $(i,j) \in U$.

Let $\mathbb{S}$ be a 2-D constraint and $(\pi_{m,n})_{m,n=1}^{\infty}$ be a (2-D) sequence of probability measures, where each measure $\pi_{m,n}$ is defined on $\mathbb{S}(\Delta_{m,n})$ (the use of parallelograms here is arbitrary; we could use rectangles $B_{m,n}$ instead). We say that the sequence $(\pi_{m,n})$ is *nested* if for every $1 \le m \le m'$, $1 \le n \le n'$, and $x \in \mathbb{S}(\Delta_{m,n})$

$$\pi_{m,n}(x) = \sum_{\substack{y \in \mathbb{S}(\Delta_{m',n'}): \\ y(\Delta_{m,n})=x}} \pi_{m',n'}(y).$$

One can verify that the probability measures that are induced by the bit-stuffing encoders in this paper form nested sequences (see (3) and (11)).

Let Δ^+ denote the set $\{(i,j) \in \mathbb{Z}^2 : i \ge 0, i+j \ge 0\}$. Also, denote by $\mathbb{S}^*$ the union $\cup_U \mathbb{S}(U)$, taken over all finite subsets $U \subset \Delta^+$. The nesting property allows to associate with $(\pi_{m,n})_{m,n=1}^{\infty}$ a function $\pi : \mathbb{S}^* \to [0,1]$, which is defined for every $U \subset \Delta^+$ and $x \in \mathbb{S}(U)$ by

$$\pi(x) = \sum_{\substack{y \in \mathbb{S}(\Delta_{m,n}): \\ y(U)=x}} \pi_{m,n}(y)$$

where (m,n) is such that $U \subseteq \Delta_{m,n}$; indeed, the nesting property guarantees that the value $\pi(x)$ is independent of the choice of m or n, as long as $U \subseteq \Delta_{m,n}$. In addition, π defines a probability measure on $\mathbb{S}(U)$ for every finite subset $U \subset \Delta^+$. We will hereafter represent the sequence $(\pi_{m,n})_{m,n=1}^{\infty}$ by the function π and call the latter a nested probability function on $\mathbb{S}^*$.

Given a nested probability function π on $\mathbb{S}^*$ and a positive number N, define the function $\pi^N : \mathbb{S}^* \to [0,1]$ by

$$\pi^N(x) = \frac{1}{N^2} \sum_{(r,s) \in B_{N,N}} \pi(\sigma_{r,s}(x)).$$

It can be easily seen that π^N also defines a nested probability function on $\mathbb{S}^*$. The next result establishes the "*quasi-stationary*" property of $\pi^N(x)$.

Proposition 6.1: For fixed nonnegative integers m, m', n, and n', and every $x \in \mathbb{S}^*$

$$\left|\pi^N(\sigma_{m,n}(x)) - \pi^N(\sigma_{m',n'}(x))\right| \le \frac{2}{N}\Big(|m-m'| + |n-n'|\Big) = o_N(1).$$

Proof: By definition

$$\pi^N(\sigma_{m,n}(x)) = \frac{1}{N^2} \sum_{(r,s) \in B_{N,N}} \pi(\sigma_{r,s}(\sigma_{m,n}(x)) = \frac{1}{N^2} \sum_{(r,s) \in B_{N,N}} \pi(\sigma_{m+r,n+s}(x)).$$

Similarly

$$\pi^N(\sigma_{m',n'}(x)) = \frac{1}{N^2} \sum_{(r,s) \in B_{N,N}} \pi(\sigma_{m'+r,n'+s}(x)).$$

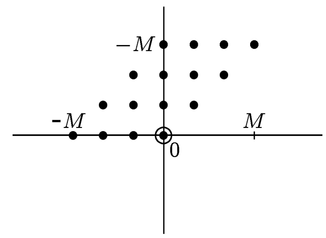


Fig. 13. Set $-\Delta_{M+1,M+1}$ which contains W.

Therefore,

$$\left|\pi^N(\sigma_{m,n}(x)) - \pi^N(\sigma_{m',n'}(x))\right| = \frac{1}{N^2}\left|\left(\sum_{(r,s) \in B_{N,N}} \pi(\sigma_{m+r,n+s}(x))\right) - \left(\sum_{(r',s') \in B_{N,N}} \pi(\sigma_{m'+r',n'+s'}(x))\right)\right|.$$

The two sums in the right-hand side include $2N^2$ terms, yet all but at most $2N(|m-m'|+|n-n'|)$ terms cancel out. The result follows. □

Let μ be a nested probability function on $\mathbb{S}^*$. We say that μ is *local* if there exist a finite subset $W \subset \mathbb{Z}^2$ and a conditional probability function $\vartheta : \Sigma^{|W|} \to [0,1]$ such that the following three conditions hold.

L-1: $W \subset -\Delta^+$ and $(0,0) \in W$. Hereafter, we let the integer M be such that $W \subseteq -\Delta_{M+1,M+1}$ and define $W' = W \setminus \{(0,0)\}$ (see Fig. 13).

L-2: For every $y \in \mathbb{S}(W')$

$$\sum_{x \in \mathbb{S}(W): x(W')=y} \vartheta\left(x \mid (x_{u,v})_{(u,v) \in W'}\right) = 1.$$

L-3: For every $m, n \ge M$ and $x \in \mathbb{S}(\Delta_{m,n})$, the value $\mu(x)(= \mu_{m,n}(x))$ takes the form

$$\mu(x) = \mu\Big(x\left(\partial\Delta_{m,n}^{(M)}\right)\Big) \cdot \prod_{(i,j) \in \Delta_{m,n} \setminus \partial\Delta_{m,n}^{(M)}} \vartheta\left(x_{i,j} \mid (x_{u,v})_{(u,v) \in \sigma_{i,j}(W')}\right).$$

Observe that L-2 implies that every element $y \in \mathbb{S}(W')$ can be extended to at least one element $x \in \mathbb{S}(W)$ such that $x(W') = y$.

A local probability function is also *causal* [8]: it can be simulated (e.g., during encoding) by first setting the entries at the boundary $\partial\Delta_{m,n}^{(M)}$, and then scanning $\Delta_{m,n}$ diagonal by diagonal (or row by row, or alternating between diagonal and row scans) and setting the entries according to the output of biased sequences. The measure (3) is local with $M = d$, and the measure (11) is local with $M = 2$. While we do make here assumptions about the shape of W, we point out that these assumptions can be relaxed by, say, using parallelograms whose diagonals have slopes other than 1.

Proposition 6.2: Let μ be a local probability function on $\mathbb{S}^*$ with a respective set W and conditional probability function ϑ.

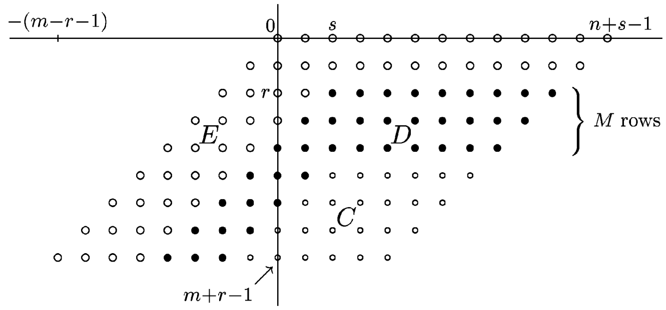


Fig. 14. Sets C, D, and E.

Given a positive integer N, define the nested probability function $\mu^N : \mathbb{S}^* \to [0,1]$ by

$$\mu^N(x) = \frac{1}{N^2} \sum_{(r,s)\in B_{N,N}} \mu(\sigma_{r,s}(x)).$$

Then μ^N is local with the same set W and function ϑ.

Proof: Let M be such that $W \subseteq -\Delta_{M+1,M+1}$. We need to show that for every $m, n \geq M$ and $x \in \mathbb{S}(\Delta_{m,n})$

$$\mu^N(x) = \mu^N\left(x\left(\partial\Delta_{m,n}^{(M)}\right)\right) \cdot \prod_{(i,j)\in\Delta_{m,n}\setminus\partial\Delta_{m,n}^{(M)}} \vartheta\left(x_{i,j} \mid (x_{u,v})_{(u,v)\in\sigma_{i,j}(W')}\right). \quad (13)$$

Fix $x \in \mathbb{S}(\Delta_{m,n})$ and $(r,s) \in B_{N,N}$ and define the sets

$$C = \sigma_{r,s}\left(\Delta_{m,n} \setminus \partial\Delta_{m,n}^{(M)}\right)$$
$$D = \sigma_{r,s}\left(\partial\Delta_{m,n}^{(M)}\right)$$

and

$$E = \Delta_{m+r,n+s} \setminus (C \cup D)$$

(see Fig. 14). As μ satisfies L-3, it follows that for every $y \in \mathbb{S}(\Delta_{m+r,n+s})$

$$\mu(y) = \mu(y(D\cup E)) \cdot \prod_{(i,j)\in C} \vartheta\left(y_{i,j} \mid (y_{u,v})_{(u,v)\in\sigma_{i,j}(W')}\right). \quad (14)$$

Now

$$\mu(\sigma_{r,s}(x)) = \sum_{y\in\mathcal{J}} \mu(y)$$

where

$$\mathcal{J} = \left\{y \in \mathbb{S}(\Delta_{m+r,n+s}) : y(C\cup D) = \sigma_{r,s}(x)\right\}.$$

Hence,

$$\mu(\sigma_{r,s}(x)) = \sum_{y\in\mathcal{J}} \Big(\mu(y(D\cup E)) \cdot \underbrace{\prod_{(i,j)\in C} \vartheta\left(y_{i,j} \mid (y_{u,v})_{(u,v)\in\sigma_{i,j}(W')}\right)}_{\alpha}\Big) \quad (15)$$

and one can verify that

$$\alpha = \prod_{(i,j)\in\Delta_{m,n}\setminus\partial\Delta_{m,n}^{(M)}} \vartheta\left(x_{i,j} \mid (x_{u,v})_{(u,v)\in\sigma_{i,j}(W')}\right)$$

namely, α depends neither on the particular element $y \in \mathcal{J}$ nor on $(r,s) \in B_{N,N}$.

Assume first that $\alpha > 0$. Let z belong to the set

$$\mathcal{K} = \left\{z \in \mathbb{S}(D\cup E) : z(D) = \sigma_{r,s}\left(x\left(\partial\Delta_{m,n}^{(M)}\right)\right)\right\}$$

and suppose further that $\mu(z) > 0$. Let y be a $\Delta_{m+r,n+s}$-configuration such that $y(D\cup E) = z$ and $y(C\cup D) = \sigma_{r,s}(x)$. The right-hand side of (14), being equal to $\mu(z)\cdot\alpha$, is strictly positive; hence, y is necessarily in $\mathbb{S}(\Delta_{m+r,n+s})$. It follows that every element $z \in \mathcal{K}$ with $\mu(z) > 0$ can be extended to an element $y \in \mathcal{J}$. Therefore,

$$\sum_{y\in\mathcal{J}} \mu(y(D\cup E)) = \sum_{z\in\mathcal{K}} \mu(z) = \mu\Big(\sigma_{r,s}\left(x\left(\partial\Delta_{m,n}^{(M)}\right)\right)\Big)$$

and so, from (15), we get

$$\mu(\sigma_{r,s}(x)) = \mu\Big(\sigma_{r,s}\left(x\left(\partial\Delta_{m,n}^{(M)}\right)\right)\Big) \cdot \alpha; \quad (16)$$

furthermore, (15) implies that (16) holds also when $\alpha = 0$. The equality (13) is finally obtained by averaging both sides of (16) over $(r,s) \in B_{N,N}$. □

Given a local probability function μ on $\mathbb{S}^*$, it follows from Propositions 6.1 and 6.2 that the function μ^N is both quasi-stationary and local, and it shares with μ the same set W and conditional probability function ϑ; that is, μ^N differs from μ only due to the measure on the boundary $\partial\Delta_{m,n}^{(M)}$. Note that in (3) or (11), the conditional probability function ϑ is determined by the coding rule; this means that once we set the entries at the boundary, the probability function μ^N can be simulated by the very same coding rule which induces μ.

The proofs in Sections III–V were based on global properties of μ, such as bounds on the expected values of the number of occurrences of a given event throughout $\Delta_{m,n}$. The quasi-stationary property allows to obtain local properties as well. For example, by using μ^N instead of μ, we can strengthen Lemma 2.2 to

$$\Pr\{\mathcal{C}_{i,j}\} \leq dp \cdot \Pr\{\mathcal{B}_{i,j}\} + O(d^2/N)$$

and

$$\Pr\{\mathcal{R}_{i,j}\} \leq dp \cdot \Pr\{\mathcal{B}_{i,j}\} + O(d^2/N)$$

both inequalities holding for every $(i,j) \in \Delta_{m,n} \setminus \partial\Delta_{m,n}^{(d)}$.

Appendix
Improved Lower Bounds for Small d

We can further improve the bounds in Sections III and IV by bounding the "double-stuffing" probability $\Pr\{\mathcal{M}_{i,j}\}$ ($\Pr\{\mathcal{M}_{i,j}^{\text{hex}}\}$) more carefully. The following derivation is based on the square lattice, and we use the same notations as in Section III; similar results can be obtained for the hexagonal lattice.

For $r = 1, 2, \ldots, d^2$, denote by $\mathcal{T}_{i,j}^{(r)}$ the event that $\mathcal{B}_{i,j}$ occurs and there are (exactly) r locations $(s,t) \in \Gamma(i,j)$ which are not stuffed from the locations outside of $\Gamma(i,j)$. Since $\mathcal{M}_{i,j} \subseteq \left(\bigcup_{r\geq 1} \mathcal{T}_{i,j}^{(r)}\right)$, it follows that

$$\Pr\{\mathcal{M}_{i,j}\} = \sum_{r\geq 1} \Pr\{\mathcal{M}_{i,j} \mid \mathcal{T}_{i,j}^{(r)}\} \cdot \Pr\{\mathcal{T}_{i,j}^{(r)}\}.$$

Given the event $\mathcal{T}_{i,j}^{(r)}$, there will be no "double-stuffing" from $\Gamma(i,j)$ only when all the r locations there are set to 0; this, in turn, occurs with probability $(1-p)^r$. Thus,

$$\Pr\left\{\mathcal{M}_{i,j} \mid \mathcal{T}_{i,j}^{(r)}\right\} = p\left(1 - (1-p)^r\right)$$

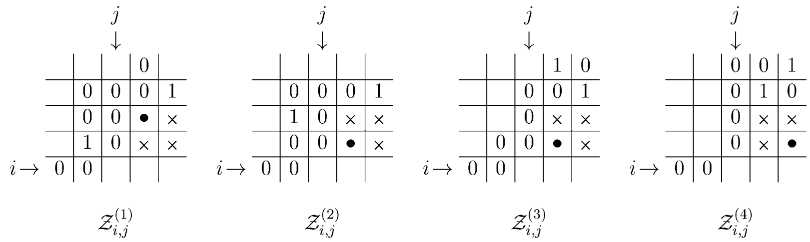


Fig. 15. Events $\mathcal{Z}_{i,j}^{(\ell)}$; entries indexed by $\Gamma(i,j)$ that are stuffed are marked by "$\times$."

and

$$\begin{aligned}\mathsf{Pr}\{\mathcal{M}_{i,j}\} &= \sum_{r\geq 1} p\left(1-(1-p)^r\right)\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\}\\ &\geq p^2\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(1)}\right\} + p\left(1-(1-p)^2\right)\sum_{r\geq 2}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\}\\ &= p^2\sum_{r\geq 1}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\}\\ &\quad + \left(p\left(1-(1-p)^2\right)-p^2\right)\sum_{r\geq 2}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\}\\ &= p^2\sum_{r\geq 1}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\}\\ &\quad + p^2(1-p)\left(\left(\sum_{r\geq 1}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\}\right)-\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(1)}\right\}\right)\end{aligned}$$

where the inequality comes from the fact that

$$p\left(1-(1-p)^r\right) \geq p\left(1-(1-p)^2\right), \qquad \text{for } r\geq 2.$$

Now, by Lemma 3.2 and (7) we have

$$\begin{aligned}\sum_{r\geq 1}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(r)}\right\} &= \mathsf{Pr}\left\{\mathcal{B}_{i,j}\setminus\mathcal{S}_{i,j}\right\}\\ &\geq \mathsf{Pr}\{\mathcal{B}_{i,j}\} - p^{2d-1}\cdot\mathsf{Pr}\{\mathcal{B}_{u,v}\}\end{aligned}$$

(see the definition of (u,v) in Section III). Therefore, we conclude that

$$\begin{aligned}\mathsf{Pr}\{\mathcal{M}_{i,j}\} \geq\, & p^2\left(\mathsf{Pr}\{\mathcal{B}_{i,j}\}-p^{2d-1}\cdot\mathsf{Pr}\{\mathcal{B}_{u,v}\}\right)+p^2(1-p)\\ &\cdot\left(\left(\mathsf{Pr}\{\mathcal{B}_{i,j}\}-p^{2d-1}\cdot\mathsf{Pr}\{\mathcal{B}_{u,v}\}\right)-\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(1)}\right\}\right).\end{aligned} \tag{17}$$

We next derive an upper bound on $\mathsf{Pr}\{\mathcal{T}_{i,j}^{(1)}\}$ by identifying the patterns that result in the event $\mathcal{T}_{i,j}^{(1)}$. In what follows, we confine ourselves to the special case $d=2$ and define the following four events:

$$\begin{aligned}\mathcal{Z}_{i,j}^{(1)} &= (\mathcal{B}_{i,j}\cap\mathcal{X}_{i-1,j-1}\cap\mathcal{X}_{i-3,j+2})\\ \mathcal{Z}_{i,j}^{(2)} &= (\mathcal{B}_{i,j}\cap\mathcal{X}_{i-2,j-1}\cap\mathcal{X}_{i-3,j+2})\\ \mathcal{Z}_{i,j}^{(3)} &= (\mathcal{B}_{i,j}\cap\mathcal{X}_{i-4,j+1}\cap\mathcal{X}_{i-3,j+2})\\ \mathcal{Z}_{i,j}^{(4)} &= (\mathcal{B}_{i,j}\cap\mathcal{X}_{i-3,j+1}\cap\mathcal{X}_{i-4,j+2})\end{aligned}$$

(see Fig. 15).

The next lemma can be easily verified.

Lemma A.1: The following holds when $d=2$ for all $(i,j)\in\Delta_{m,n}\setminus\partial\Delta_{m,n}^{(4)}$:

a) $\mathcal{Z}_{i,j}^{(1)} = \mathcal{B}_{i,j}\cap\overline{\mathcal{B}}_{i-2,j+1}\cap\overline{\mathcal{B}}_{i-2,j+2}\cap\overline{\mathcal{B}}_{i-1,j+1}\cap\overline{\mathcal{B}}_{i-1,j+2}$;

b) $\mathcal{Z}_{i,j}^{(2)}\cup\mathcal{Z}_{i,j}^{(3)} = \mathcal{B}_{i,j}\cap\overline{\mathcal{B}}_{i-2,j+1}\cap\overline{\mathcal{B}}_{i-2,j+2}\cap\mathcal{B}_{i-1,j+1}\cap\overline{\mathcal{B}}_{i-1,j+2}$;

c) $\mathcal{Z}_{i,j}^{(4)} = \mathcal{B}_{i,j}\cap\overline{\mathcal{B}}_{i-2,j+1}\cap\overline{\mathcal{B}}_{i-2,j+2}\cap\overline{\mathcal{B}}_{i-1,j+1}\cap\mathcal{B}_{i-1,j+2}$;

d) $\mathcal{B}_{i,j}\cap\overline{\mathcal{B}}_{i-2,j+1}\cap\mathcal{B}_{i-2,j+2}\cap\overline{\mathcal{B}}_{i-1,j+1}\cap\overline{\mathcal{B}}_{i-1,j+2}=\emptyset$.

It follows from Lemma A.1 that $\mathcal{T}_{i,j}^{(1)} = \bigcup_{\ell=1}^{4}\mathcal{Z}_{i,j}^{(\ell)}$. By a procedure similar to the one that develops inequality (6), we get for $\ell=1,2,3,4$ that

$$\mathsf{Pr}\left\{\mathcal{Z}_{i,j}^{(\ell)}\right\} \leq p^2\cdot\mathsf{Pr}\{\mathcal{B}_{u^{(\ell)},v^{(\ell)}}\}$$

for some $(u^{(\ell)},v^{(\ell)}) = (u_{i,j}^{(\ell)},v_{i,j}^{(\ell)})\in\Delta_{m,n}$. Therefore,

$$\begin{aligned}\mathsf{Pr}\left\{\mathcal{T}_{i,j}^{(1)}\right\} &= \mathsf{Pr}\left\{\bigcup_{\ell=1}^{4}\mathcal{Z}_{i,j}^{(\ell)}\right\}\\ &\leq \sum_{\ell=1}^{4}\mathsf{Pr}\left\{\mathcal{Z}_{i,j}^{(\ell)}\right\} \leq p^2\sum_{\ell=1}^{4}\mathsf{Pr}\{\mathcal{B}_{u^{(\ell)},v^{(\ell)}}\}.\end{aligned}$$

Plugging this bound into (17) and setting $d=2$, we obtain

$$\begin{aligned}\Pr\{\mathcal{M}_{i,j}\} \geq\, & p^2\left(\Pr\{\mathcal{B}_{i,j}\}-p^3\cdot\Pr\{\mathcal{B}_{u,v}\}\right)+p^2(1-p)\\ &\cdot\Bigg(\left(\Pr\{\mathcal{B}_{i,j}\}-p^3\cdot\Pr\{\mathcal{B}_{u,v}\}\right)\\ &\qquad - p^2\sum_{\ell=1}^{4}\Pr\{\mathcal{B}_{u^{(\ell)},v^{(\ell)}}\}\Bigg).\end{aligned}$$

Summing over all $(i,j)\in\Delta_{m,n}\setminus\partial\Delta_{m,n}^{(4)}$ yields

$$\mathsf{E}\{\mathcal{M}\} \geq p^2[(1-p^3)+(1-p)(1-4p^2-p^3)]\cdot\mathsf{E}\{\mathcal{B}\}-O(m+n)$$

which leads to the improved lower bound for $d=2$

$$\begin{aligned}&\mathsf{H}(\mu_{\mathrm{sq};m,n}^{2,\infty})\\ &\quad\geq \max_{0<p<1}\frac{h(p)}{1+4p-p^2[(1-p^3)+(1-p)(1-4p^2-p^3)]}\\ &\qquad - o_{\min\{m,n\}}(1).\end{aligned}$$

Similarly, we can derive improved lower bounds for $d=3$ and $d=4$ on the square lattice

$$\begin{aligned}&\mathsf{H}(\mu_{\mathrm{sq};m,n}^{3,\infty})\\ &\quad\geq \max_{0<p<1}\frac{h(p)}{1+6p-p^2[(1-p^5)+(1-p)(1-4p^4-5p^5)]}\\ &\qquad - o_{\min\{m,n\}}(1)\end{aligned}$$

and

$$\mathsf{H}(\mu^{4,\infty}_{\mathrm{sq};m,n}) \geq \max_{0<p<1} \frac{h(p)}{1+8p-p^2[(1-p^7)+(1-p)(1-4p^6-9p^7)]} - o_{\min\{m,n\}}(1).$$

The procedure is also applicable to the hexagonal lattice. For $d=2$, it yields the following improved lower bound:

$$\mathsf{H}(\mu^{2,\infty}_{\mathrm{hex};m,n}) \geq \max_{0<p<1} \frac{h(p)}{1+6p-p^2[1+(1-p)(1-2p^2-16p^3-3p^4)]} - o_{\min\{m,n\}}(1).$$

Acknowledgment

The authors thank Søren Forchhammer, Jørn Justesen, Erik Ordentlich, Alon Orlitsky, and Ken Zeger for very helpful comments.

References

[1] R. J. Baxter, "Hard hexagons: Exact solution," *J. Phys. A*, vol. 13, pp. 61–70, 1980.
[2] J. J. Ashley and B. H. Marcus, "Two-dimensional low-pass filtering codes," *IEEE Trans. Commun.*, vol. 46, pp. 724–727, June 1998.
[3] P. Bender and J. K. Wolf, "A universal algorithm for generating optimal and nearly optimal run-length-limited, charge constrained binary sequences," in *Proc. 1993 IEEE Int. Symp. Information Theory*, San Antonio, TX, 1997, p. 6.
[4] R. Berger, "The undecidability of the domino problem," *Mem. Amer. Math. Soc.*, vol. 66, 1966.
[5] D. Brady and D. Psaltis, "Control of volume holograms," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1167–1182, 1992.
[6] R. Burton and J. E. Steif, "Non-uniqueness of measures of maximal entropy for subshifts of finite type," *Ergod. Theory Dynam. Syst.*, vol. 14, pp. 213–235, 1994.
[7] N. J. Calkin and H. S. Wilf, "The number of independent sets in a grid graph," *SIAM J. Discr. Math.*, vol. 11, pp. 54–60, 1998.
[8] F. Champagnat, J. Idier, and Y. Goussard, "Stationary Markov random fields on a finite rectangular lattice," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2901–2916, Nov. 1998.
[9] K. Engel, "On the fibonacci number of an $m \times n$ lattice," *Fibonacci Quart.*, vol. 28, pp. 72–78, 1990.
[10] S. Halevy, "Two-dimensional coding for recording channels," M.Sc. thesis, Computer Sci. Dept., Technion, Haifa, Israel, 2001.
[11] J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science*, vol. 265, pp. 749–752, 1994.
[12] ——, "Channel codes for digital holographic data storage," *J. Opt. Soc. Amer. A*, vol. 12, pp. 2432–2439, 1995.
[13] H. Ito, A. Kato, Z. Nagy, and K. Zeger, "Zero capacity region of multidimensional run length constraints," *Electr. J. Combin.*, vol. 6, p. R33, 1999.
[14] J. Justesen and Y. M. Shtarkov, "The combinatorial entropy of images" (in Russian), *Probl. Pered. Inform.*, vol. 33, pp. 3–11, 1997.
[15] A. Kato and K. Zeger, "On the capacity of two-dimensional run-length constrained channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1527–1540, July 1999.
[16] ——, "Partial characterization of the positive capacity region of two-dimensional asymmetric run length constrained channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2666–2670, Nov. 2000.
[17] Z. Kukorelly and K. Zeger, "The capacity of some hexagonal (d,k)-constraints," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 64.
[18] P. Lee, "Combined error-correcting/modulation recording codes," Ph.D. dissertation, Univ. California, San Diego , La Jolla, 1988.
[19] M. Marcus, *A Survey of Finite Mathematics*. New York: Dover, 1969.
[20] 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. C. Huffman, Eds. Amsterdam, The Netherlands: Elsevier Scientific, 1998, pp. 1635–1764.
[21] N. Markley and M. Paul, "Maximal measures and entropy for Z^ν subshifts of finite type," in *Classical Mechanics and Dynamical Systems*. ser. Lecture Notes in Pure and Applied Mathematics, R. Devaney and Z. Nitecki, Eds. New York: Marcel Dekker, 1981, vol. 70, pp. 135–157.
[22] Z. Nagy and K. Zeger, "Capacity bounds for the 3-dimensional $(0,1)$ runlength limited channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1030–1033, May 2000.
[23] K. Norris and D. S. Bloomberg, "Channel capacity of charge-constrained run-length limited systems," *IEEE Trans. Magn.*, vol. MAG-17, pp. 3452–3455, Nov. 1981.
[24] K. C. Pohlmann, *The Compact Disc Handbook*, 2nd ed. Madison, WI: A-R Editions, 1992.
[25] D. Psaltis and F. Mok, "Holographic memories," *Scientific Amer.*, vol. 273, no. 5, pp. 70–76, 1995.
[26] D. Psaltis, M. A. Neifeld, A. Yamamura, and S. Kobayashi, "Optical memory disks in optical information processing," *Appl. Opt.*, vol. 29, pp. 2038–2057, 1990.
[27] R. M. Robinson, "Undecidability and nonperiodicity for tilings of the plane," *Inventiones Math.*, vol. 12, pp. 177–209, 1971.
[28] R. M. Roth, P. H. Siegel, and J. K. Wolf, "Efficient coding for a two-dimensional runlength-limited constraint," in *Proc. SPIE Int. Symp. Optical Science, Engineering, and Instrumentation*, Denver, CO, July 1999, pp. 8–17.
[29] ——, "Efficient coding schemes for the two-dimensional runlength-limited $(1,\infty)$ constraint," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1166–1176, Mar. 2001.
[30] Y. M. Shtarkov, S. Forchhammer, and J. Justesen, "On Definitions and Existence of Combinatorial Entropy of 2d Fields," Tech. Univ. Denmark, Lyngby, Tech. Rep. TR-TELE-10, 1998.
[31] P. H. Siegel and J. K. Wolf, "Bit-stuffing bounds on the capacity of 2-dimensional constrained arrays," in *Proc. 1998 IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 323.
[32] R. Talyansky, "Coding for two-dimensional constraints," M.Sc. thesis (in Hebrew), Computer Sci. Dept., Technion, Haifa, Israel, 1997.
[33] R. Talyansky, T. Etzion, and R. M. Roth, "Efficient constructions for certain two-dimensional constraints," *IEEE Trans. Inform. Theory*, vol. 45, pp. 794–799, Mar. 1999.
[34] W. Weeks IV and R. E. Blahut, "The capacity and coding gain of certain checkerboard codes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1193–1203, May 1998.