Efficient Coding Schemes for the Hard-Square Model*

Ron M. Roth† Paul H. Siegel‡ Jack K. Wolf§

Abstract

The hard-square model, also known as the two-dimensional $(1,\infty)$-RLL constraint, consists of all binary arrays in which the 1's are isolated both horizontally and vertically. Based on a certain probability measure defined on those arrays, an efficient variable-to-fixed encoder scheme is presented that maps unconstrained binary words into arrays that satisfy the hard-square model. For sufficiently large arrays, the average rate of the encoder approaches a value which is only 0.1% below the capacity of the constraint. A second, fixed-rate encoder is presented whose rate for large arrays is within 1.2% of the capacity value.

Keywords: Constrained codes; Enumerative coding; Hard-square model; Max-entropic probability measure; Permutation codes; Variable-to-fixed encoders; Two-dimensional runlength-limited constraints.

1 Introduction

In current digital optical and magnetic recording systems, such as disks and tapes, the data is written along *tracks,* thus visualized as a one-dimensional long sequence. To ensure reliability, the raw data typically undergoes lossless coding into a binary sequence that satisfies certain constraints. One of the most commonly used constraints is the (one-dimensional) (d,k)-runlength-limited (RLL) constraint, which consists of all finite binary words in which the

*This work was supported in part by Grants Nos. 95-00522 and 99-00199 from the United-States–Israel Binational Science Foundation (BSF), Jerusalem, Israel, by Grant No. NCR-9612802 of the National Science Foundation (NSF), and by the Center for Magnetic Recording Research at the University of California, San Diego.

†Computer Science Department, Technion—Israel Institute of Technology, Haifa 32000, Israel. e-mail: ronny@cs.technion.ac.il.

‡Department of Electrical and Computer Engineering, 0407, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0407, USA. e-mail: psiegel@ucsd.edu.

§Center for Magnetic Recording Research, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0407, USA. e-mail: jwolf@ucsd.edu.

runlengths of 0's are at least d, and runlengths of 0's between two consecutive 1's do not exceed k [18],[19],[25].

Recent developments in optical storage—especially in the area of holographic memory—are attempting to increase the recording density by exploiting the fact that the recording device is a *surface*. Under this new model, the recorded data is regarded as two-dimensional, as opposed to the track-oriented one-dimensional recording paradigm. The new approach, however, introduces new types of error patterns and constraints—those now become two-dimensional rather than one-dimensional. See [2],[4],[12],[13],[22],[23].

The treatment of two-dimensional constraints seems to be much more difficult than the one-dimensional case. This is, in part, due to the fact that in the general constrained setting, there are problems that are easy to solve in the one-dimensional case, yet they become undecidable when we shift to two dimensions [3],[24].

One important example of a two-dimensional constraint is the two-dimensional extension of the $(1, \infty)$-RLL constraint. This constraint, which is also referred to as the hard-square model, has been treated in quite a few papers in the past several years; see, for example, [6], [10], [11], [20], [29]. This constraint will also be the focus of this work. We define next the hard-square model, borrowing terms from [5].

Let U be a finite subset of the integer plane $\mathbb{Z}^2$ and let Σ be a finite set, referred to as an *alphabet*. A *U-configuration* is a mapping $x : U \to \Sigma$. The value of x at location $(i, j) \in U$ will be denoted by $x_{i,j}$.

We say that a U-configuration x *satisfies the hard-square model* if $\Sigma = \{0, 1\}$ and for every two distinct locations $(i, j), (i', j') \in U$,

$$|i - i'| + |j - j'| \le 1 \quad \Longrightarrow \quad (x_{i,j} = 0 \quad \text{or} \quad x_{i',j'} = 0) \ .$$

Equivalently, if we write down the values of the U-configuration in the integer plane, then the 1's are isolated both horizontally and vertically (either by 0's or by unassigned locations). The set of all U-configurations that satisfy the hard-square model will be denoted by $\mathcal{S}(U)$.

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

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

or parallelograms

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

(see Figure 1). We will be mainly concentrating on the hard-square model, as the known literature, as well as the results obtained herein, are elaborate enough already for this special case.

The *capacity*, or the *topological entropy*, of the hard-square model is given by

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

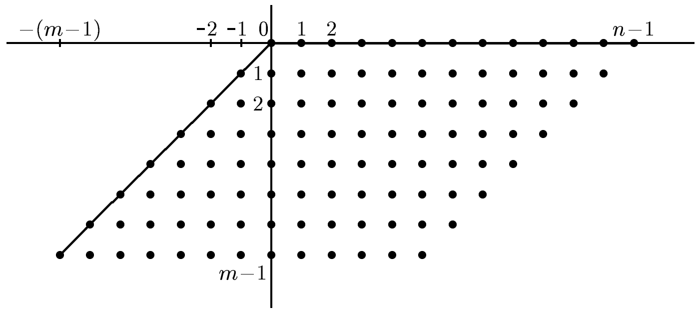

Figure 1: Parallelogram $\Delta_{m,n}$.

The limits indeed exist and are equal [5],[16]. The value of $\mathsf{cap}(\mathcal{S})$ is known to be approximately 0.5878911162; see [6], [10], [11], [29].

Much less is known about efficient (i.e., polynomial-time, or low-complexity) high-rate coding schemes for this constraint. In [26], the idea of two-dimensional bit-stuffing was introduced, resulting in a variable-to-fixed encoder whose expected rate was bounded from below in [26] by approximately 0.5515. Note that in a variable-to-fixed scheme, the set of pre-images, denoted D, consists of binary words that are not necessarily of the same length; still every sufficiently long binary unconstrained word has exactly one element in D as a prefix (namely, the set D is prefix-free and complete). For the purpose of computing the rate, we define a probability measure on D, where a pre-image w of length $\ell = \ell(w)$ has probability $2^{-\ell}$. Indeed, by the properties of D it follows that $\sum_{w\in D} 2^{-\ell(w)} = 1$ (the Kraft equality). The expected rate of such a coding scheme is given by

$$R_{m,n} = \frac{1}{mn} \cdot \sum_{w\in D} 2^{-\ell(w)}\ell(w) \ .$$

A very simple coding scheme into $\mathcal{S}(\Delta_{m,n})$ at a *fixed rate* $1:2$ is implied by Lemma 1(e) in [14]: entries $(i,j) \in \Delta_{m,n}$ such that $i+j$ is even are filled with the input bit stream, while the remaining entries are set to zero. We do not know of any other published efficient fixed-rate encoders at (significantly) higher rates for the two-dimensional $(1,\infty)$-RLL constraint.

The main goal of this work is designing efficient coding schemes for mapping, in a one-to-one manner, unconstrained binary words into elements of $\mathcal{S}(B_{m,n})$ or $\mathcal{S}(\Delta_{m,n})$. Based on the idea of two-dimensional bit-stuffing introduced in [26], we present in Section 3 a variable-to-fixed encoder into $\mathcal{S}(\Delta_{m,n})$. Our coding scheme attains a rate which is approximately 0.587277, namely, only 0.1% below the value of $\mathsf{cap}(\mathcal{S})$.

Our variable-to-fixed rate encoder effectively realizes a certain probability measure $\mu_{m,n}$ on $\mathcal{S}(\Delta_{m,n})$. This measure is defined in Section 2 and its properties are proved in Section 4. In particular, we show that the marginal probability induced by $\mu_{m,n}$ at every given row—and respectively at every given diagonal—of a random $\Delta_{m,n}$-configuration in $\mathcal{S}(\Delta_{m,n})$ is a first-order Markov process.

With a slight compromise on the coding rate, we can also obtain an efficient fixed-rate encoder into $\mathcal{S}(B_{m,n})$. Such an encoder is presented in Section 5 with a rate that approaches 0.581074 for large values of m or n; this rate is within 1.2% of the value of $\mathsf{cap}(\mathcal{S})$.

2 Probability measure on parallelograms

Let $\Delta_{m,n} \subseteq \mathbb{Z}^2$ be the parallelogram defined by (2) and shown in Figure 1. *Row* i in $\Delta_{m,n}$ consists of all locations (i, j) such that $-i \le j < n-i$. *Diagonal* d consists of all locations $(i, d-i)$ such that $0 \le i < m$. Row 0 will be denoted by $\Delta_n^{(\mathrm{h})}$ and will be referred to as the *horizontal boundary* of $\Delta_{m,n}$. Similarly, Diagonal 0, denoted $\Delta_m^{(\mathrm{d})}$, will be referred to as the *diagonal boundary* of $\Delta_{m,n}$. Those boundaries are depicted as thick lines in Figure 1. The set $\Delta_{m,n} \setminus (\Delta_n^{(\mathrm{h})} \cup \Delta_m^{(\mathrm{d})})$ (i.e., the parallelogram excluding its boundaries) will be denoted by $\Delta^*_{m,n}$.

A random $\Delta_{m,n}$-configuration taking values from $\mathcal{S}(\Delta_{m,n})$ (according to some probability measure) will be denoted by X, and its value at location (i, j) will be denoted by $X_{i,j}$.

Let $\pi_{m,n}$ be a probability measure defined on $\mathcal{S}(\Delta_{m,n})$; that is, $\pi_{m,n}(x) = \mathsf{Prob}\{X = x\}$ for every $x \in \mathcal{S}(\Delta_{m,n})$. The *(measure-theoretic) entropy* of $\pi_{m,n}$ is defined by

$$H(\pi_{m,n}) = -\frac{1}{mn} \sum_{x \in \mathcal{S}(\Delta_{m,n})} \pi_{m,n}(x) \log_2 \pi_{m,n}(x) \ .$$

The value $H(\pi_{m,n})$ is the largest possible expected rate of any encoder that maps, in a one-to-one manner, a set D of input binary words into $\mathcal{S}(\Delta_{m,n})$, with a probability measure defined on D that induces the measure $\pi_{m,n}$ on $\mathcal{S}(\Delta_{m,n})$. This clearly implies the inequality

$$H(\pi_{m,n}) \le \frac{\log_2 |\mathcal{S}(\Delta_{m,n})|}{mn} \ . \tag{3}$$

Now, suppose that $\boldsymbol{\pi} = \{\pi_{m,n}\}_{m,n=1}^{\infty}$ is a (two-dimensional) sequence of probability measures, where each individual measure $\pi_{m,n}$ is defined on $\mathcal{S}(\Delta_{m,n})$. For a $\Delta_{m,n}$-configuration $x \in \mathcal{S}(\Delta_{m,n})$, let $\Lambda^{(\mathrm{h})}(x)$ be the set of all $\Delta_{m+1,n}$-configurations in $\mathcal{S}(\Delta_{m+1,n})$ obtained from x by appending an $(m+1)$st row. Similarly, let $\Lambda^{(\mathrm{d})}(x)$ be the set of all $\Delta_{m,n+1}$-configurations in $\mathcal{S}(\Delta_{m,n+1})$ obtained from x by appending an $(n+1)$st diagonal. We say that the sequence $\boldsymbol{\pi} = \{\pi_{m,n}\}_{m,n}$ is *nested* if for every $m, n \ge 1$ and $x \in \mathcal{S}(\Delta_{m,n})$,

$$\pi_{m,n}(x) = \sum_{z \in \Lambda^{(\mathrm{h})}(x)} \pi_{m+1,n}(z) = \sum_{z \in \Lambda^{(\mathrm{d})}(x)} \pi_{m,n+1}(z) \ .$$

In other words, for every $m \le m'$ and $n \le n'$, the measure $\pi_{m,n}$ is the marginal distribution on $\mathcal{S}(\Delta_{m,n})$ which is induced by the measure $\pi_{m',n'} : \mathcal{S}(\Delta_{m',n'}) \to [0, 1]$. The nesting property allows us to regard $\boldsymbol{\pi}$ as a measure which is an infinite extension of the individual measures $\pi_{m,n}$. The entropy of $\boldsymbol{\pi}$ is defined by

$$H(\boldsymbol{\pi}) = \lim_{m,n \to \infty} H(\pi_{m,n})$$

(by subadditivity the limit exists), and from (3) we have $H(\boldsymbol{\pi}) \le \mathsf{cap}(\mathcal{S})$ [5]. An (infinite extension) measure $\boldsymbol{\pi}$ for which $H(\boldsymbol{\pi}) = \mathsf{cap}(\mathcal{S})$ is called a *maxentropic measure*. Such a measure indeed exists [5].

Our coding scheme effectively defines nested measures $\mu_{m,n} : \mathcal{S}(\Delta_{m,n}) \to [0,1]$ for every $m, n \ge 1$. As we show, the sequence $\boldsymbol{\mu} = \{\mu_{m,n}\}_{m,n}$ satisfies

$$H(\boldsymbol{\mu}) = \lim_{m,n\to\infty} H(\mu_{m,n}) \approx 0.587277 \; .$$

Since the limit is very close to the known bounds on $\mathsf{cap}(\mathcal{S})$, we can say that $\boldsymbol{\mu}$ is 'almost maxentropic.' The expected rate of our coding scheme approaches, through the values of $H(\mu_{m,n})$, the value $H(\boldsymbol{\mu})$.

For every $x \in \mathcal{S}(\Delta_{m,n})$, the value $\mu_{m,n}(x) = \mathsf{Prob}\{X = x\}$ takes the following form:

$$\begin{aligned} \mu_{m,n}(x) &= \mu_0(x_{0,0}) \cdot \mu_n^{(\mathrm{h})}(x_{0,1}, x_{0,2} \ldots, x_{0,n-1} \,|\, x_{0,0}) \\ &\cdot \; \mu_m^{(\mathrm{d})}(x_{1,-1}, x_{2,-2}, \ldots, x_{m-1,-(m-1)} \,|\, x_{0,0}) \\ &\cdot \; \prod_{i=1}^{m-1} \prod_{j=-i+1}^{n-1-i} \vartheta(x_{i,j} \,|\, x_{i,j-1}, x_{i-1,j}, x_{i-1,j+1}) \; . \end{aligned} \tag{4}$$

The components μ_0, $\mu_n^{(\mathrm{h})}$, and $\mu_m^{(\mathrm{d})}$ define the measure on location $(0,0)$ and on the horizontal and diagonal boundaries, respectively, and will be specified in more detail below. The function $\vartheta : \{0,1\}^4 \to [0,1]$ is defined through two parameters, $q_0 \in [0,1)$ and $q_1 \in (0,1]$, as follows:

$$\vartheta(0 \,|\, u, y, v) = \begin{cases} q_v & \text{if } u = y = 0 \\ 1 & \text{otherwise} \end{cases} , \tag{5}$$

and $\vartheta(1 \,|\, u, y, v) = 1 - \vartheta(0 \,|\, u, y, v)$. The distribution ϑ on $X_{i,j}$ can be described verbally as follows. As dictated by the hard-square model, the value of $X_{i,j}$ is forced to be 0 unless $X_{i,j-1} = X_{i-1,j} = 0$. When the latter condition is met, then $X_{i,j}$ will be a Bernoulli random bit whose distribution depends on the value of $X_{i-1,j+1}$; if that value is 0, then $X_{i,j}$ takes the value 0 with probability q_0; otherwise, $X_{i,j}$ takes the value 0 with probability q_1. Figure 2 shows the values that determine the distribution of $X_{i,j}$; the location (i,j) is marked by a box.

The measures on the boundaries, defined by $\mu_n^{(\mathrm{h})}$ and $\mu_m^{(\mathrm{d})}$, are set so that the non-boundary values have a *stationary* distribution in the sense stated in Proposition 2.1 below. Specifically, $\mu_n^{(\mathrm{h})}$ will take the form of a first-order Markov process

$$\mu_n^{(\mathrm{h})}(w_1, w_2, \ldots, w_{n-1} \,|\, w_0) = \prod_{j=1}^{n-1} \mu^{(\mathrm{h})}(w_j \,|\, w_{j-1}) \; , \tag{6}$$

where $\mu^{(\mathrm{h})} : \{0,1\}^2 \to [0,1]$ is given by

$$\mu^{(\mathrm{h})}(0 \,|\, u) = \begin{cases} \alpha & \text{if } u = 0 \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

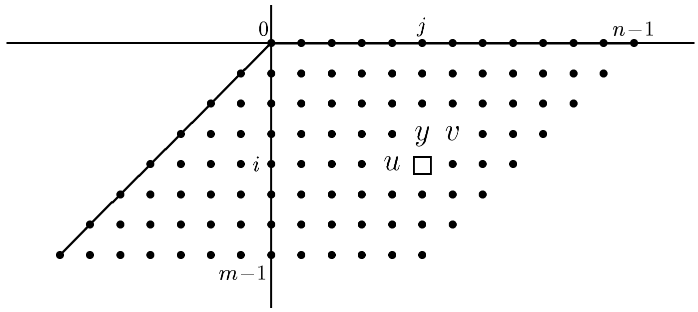

Figure 2: Location of the arguments of the function $\vartheta(\cdot \mid u, y, v)$.

for some $\alpha \in [0,1]$, and $\mu^{(\mathrm{h})}(1 \mid u) = 1 - \mu^{(\mathrm{h})}(0 \mid u)$.

The values of $\mu_0 : \{0,1\} \to [0,1]$ will be set to the stationary probabilities of the first-order Markov process $\mu^{(\mathrm{h})}$ as follows: $\mu_0(0) = \sigma$ and $\mu_0(1) = 1 - \sigma$, where

$$\mu_0(0) = \sigma = \frac{1}{2-\alpha} \ . \tag{8}$$

As for the diagonal boundary, $\mu_m^{(\mathrm{d})}$ will be a first-order Markov process of the form

$$\mu_m^{(\mathrm{d})}(w_1, w_2, \ldots, w_{m-1} \mid w_0) = \prod_{i=1}^{m-1} \mu^{(\mathrm{d})}(w_i \mid w_{i-1}) \ , \tag{9}$$

where $\mu^{(\mathrm{d})} : \{0,1\}^2 \to [0,1]$ is given by

$$\mu^{(\mathrm{d})}(0 \mid v) = \beta_v \ , \tag{10}$$

with

$$\beta_0 = \frac{\alpha}{\alpha + q_1 - \alpha q_1} \qquad \text{and} \qquad \beta_1 = \frac{q_1}{\alpha + q_1 - \alpha q_1} \ , \tag{11}$$

and $\mu^{(\mathrm{d})}(1 \mid v) = 1 - \mu^{(\mathrm{d})}(0 \mid v)$ (since $q_1 > 0$, the denominators in (11) are guaranteed to be positive). The values in (11) are consistent with the stationary distribution along the horizontal boundary: as we show in Section 4.1, (11) implies that $\mathsf{Prob}\{X_{i,-i} = 0\} = \sigma$ and, furthermore, $\mathsf{Prob}\{X_{i-1,1-i} = X_{i,-i} = 0\} = \mathsf{Prob}\{X_{0,j} = X_{1,j-1} = 0\} = \beta_0 \sigma$ for all $1 \le i < m$ and $1 \le j < n$.

The nesting property of the measures $\mu_{m,n}$ is easily verified. Next we state other properties of those measures that will be proved in Section 4. Hereafter, the notation $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ will mean that the random $\Delta_{m,n}$-configuration X is taken from the sample space $\mathcal{S}(\Delta_{m,n})$ according to the distribution $\mu_{m,n}$.

We say that row i in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ forms a first-order Markov process identical to the horizontal boundary if for every $1-i \le j < n-i$ and every nonempty word $\boldsymbol{c} = c_1 c_2 \ldots c_\ell$

of length $\ell \le i{+}j$,

$$\mathsf{Prob}\{X_{i,j} = 0 \mid X_{i,j-1}X_{i,j-2}\ldots X_{i,j-\ell} = \boldsymbol{c}\} = \begin{cases} \alpha & \text{if } c_1 = 0 \\ 1 & \text{if } c_1 = 1 \end{cases} ,$$

provided that the event we condition on has positive probability.

The main result in Section 4.1 is the following.

Proposition 2.1 *For $m, n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11). Then the entries in each row in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ form a first-order Markov process identical to the horizontal boundary if and only if*

$$\alpha = \alpha(q_0, q_1) = \frac{-q_1 + \sqrt{q_1^2 + 4q_1(1-q_0)}}{2(1-q_0)} . \tag{12}$$

As we show in Section 4.2, there exists a counterpart of Proposition 2.1 also for the diagonals in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$.

Remark. The definition of $\mu_{m,n}$ through a 'local' conditional measure $\vartheta(\cdot \mid u, y, v)$ on $\Delta_{2,2}$ as given by (5) somewhat resembles the Pickard random fields defined in [21], except that columns therein are replaced here by diagonals. Note, however, that Pickard fields assume that the measure is invariant under all the symmetries of the square, whereas we require less: the distribution along rows may (and will) differ from the distribution along diagonals. Thus, the result in Proposition 2.1 is different than the first-order Markov property of Pickard fields. □

We now turn to the measure-theoretic entropy of $\mu_{m,n}$. Define the real-valued function $h : [0,1] \to [0,1]$ by

$$h(t) = -t \log_2 t - (1{-}t) \log_2 (1{-}t) ,$$

where $h(0) = h(1) = 0$.

We show in Section 4.1 the following lower and upper bounds on $H(\mu_{m,n})$.

Proposition 2.2 *For $m, n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11) and (12). Then*

$$0 \le H(\mu_{m,n}) - \frac{(m{-}1)(n{-}1)}{mn} \cdot \beta_0 \sigma \cdot (\alpha \cdot h(q_0) + (1-\alpha) \cdot h(q_1)) \le \frac{m{+}n{-}1}{mn} .$$

By Proposition 2.2 we have

$$H(\boldsymbol{\mu}) = \lim_{m,n \to \infty} H(\mu_{m,n}) = \beta_0 \sigma \cdot (\alpha \cdot h(q_0) + (1-\alpha) \cdot h(q_1)) ,$$

which, with (8) and (11) yields

$$H(\boldsymbol{\mu}) = H(\boldsymbol{\mu}(q_0, q_1)) = \frac{\alpha \cdot (\alpha \cdot h(q_0) + (1-\alpha) \cdot h(q_1))}{(2-\alpha)(\alpha + q_1 - \alpha q_1)} ,$$

where $\alpha = \alpha(q_0, q_1)$ is given by (12). The numerator and denominator can be made linear in α by observing that (12) implies that $\alpha^2 = (1-\alpha)q_1/(1-q_0)$. This yields

$$H(\boldsymbol{\mu}(q_0, q_1)) = \frac{(1-\alpha)q_1 h(q_0) + (\alpha(1-q_0+q_1) - q_1)h(q_1)}{\alpha(2 - 2q_1 - q_1^2 - 2q_0 + 3q_0 q_1) + q_1(1 + q_1 - 2q_0)} . \qquad (13)$$

To obtain the largest rate, we maximize $H(\boldsymbol{\mu}(q_0, q_1))$ with respect to q_0 and q_1. The maximum is attained for $q_0 \approx .671833$ and $q_1 \approx 0.566932$, and the maximum value is $H(\boldsymbol{\mu}(q_0, q_1)) \approx 0.587277$.

Our analysis depends strongly on the particular structure of the measure $\mu_{m,n}$—in particular, on conditioning the probability of the event $X_{i,j} = 0$ only on the values of the three entries $X_{i-1,j}$, $X_{i-1,j+1}$, and $X_{i,j-1}$, as shown in Figure 2. We note that such conditioning is *causal* in that we can select—according to the measure $\mu_{m,n}$—an element of $\mathcal{S}(\Delta_{m,n})$ by determining the values of its entries consecutively row-by-row or diagonal-by-diagonal: in such a process, the distribution of values of the next entry to be determined is well-defined as this distribution depends on values that have already been set. Such a feature enables using the measure $\mu_{m,n}$ for encoding, as we show in Section 3.

Clearly, we may maintain causality and still approach capacity by conditioning the value of $X_{i,j}$ on more entries in the 'past.' However, it appears that the analysis thus becomes much more complex. For example, when $X_{i,j}$ is conditioned also on $X_{i-1,j+2}$, we no longer have even a second-order Markov process along rows.

3 Variable-rate encoding scheme

We describe how the estimate on $H(\mu_{m,n})$, given in Proposition 2.2, can be approached by a variable-to-fixed rate coding scheme. The objective is to realize the probability measure $\mu_{m,n}$ on $\mathcal{S}(\Delta_{m,n})$ in the output of the encoder.

The encoder consists of the following components:

1. A distribution transformer $\mathcal{E}_0$ that maps, in a one-to-one manner, sequences of fair coin flips (i.e., independent Bernoulli random bits, each equaling 0 with probability 1/2), into sequences of independent Bernoulli random bits such that each bit equals 0 with probability q_0. There are known methods [8, Section 5.12] to implement variable-to-fixed rate transformers $\mathcal{E}_0$ such that, for $\epsilon \to 0$ as the code length goes to infinity, the following holds:

   (a) the expected rate (i.e., expected number of input bits per each output bit) of $\mathcal{E}_0$ is at least $h(q_0)(1-\epsilon)$;

(b) all the words of the original Bernoulli source, except for a fraction whose probability is less than ϵ, are generated by $\mathcal{E}_0$ with probability that differs from the original probability by a factor within $1 \pm \epsilon$. Namely, the typical words of the original source are generated by $\mathcal{E}_0$ with virtually the same probability.

2. A distribution transformer $\mathcal{E}_1$, similar to $\mathcal{E}_0$, except that the output is 0 with probability q_1. The rate of $\mathcal{E}_1$ can get arbitrarily close to $h(q_1)$.

3. Probabilistic boundary generator $\mathcal{E}_2$, to be explained below.

4. Constrained coder $\mathcal{E}_3$, to be explained below.

The raw input bits are fed into the transformers $\mathcal{E}_0$ and $\mathcal{E}_1$, each input bit entering exactly one of the transformers. The coder $\mathcal{E}_3$ then queries the outputs of $\mathcal{E}_0$ and $\mathcal{E}_1$ throughout the encoding process. The order of queries determines which transformer is fed by any given input bit.

The encoding procedure starts by generating the entry $X_{0,0}$ at the origin, the entries $X_{0,j}$, $1 \leq j < n$, along the horizontal boundary, and the entries $X_{i,-i}$, $1 \leq i < m$, along the diagonal boundary. Those entries are generated by $\mathcal{E}_2$ probabilistically, using (internal) sources of Bernoulli random trials (i.e., internal coin flips), with probabilities of success α, σ, β_0, and β_1, as given by (12), (8), and (11). Note that these coin flips can be driven by external sources (as is done in $\mathcal{E}_0$ and $\mathcal{E}_1$), thus contributing to the rate; however, since the boundaries occupy only $m+n-1$ bits out of the mn bits of $\Delta_{m,n}$, such a rate contribution becomes marginal when m and n are large.

The main coding task is performed by $\mathcal{E}_3$, which is fed by the outputs of $\mathcal{E}_0$ and $\mathcal{E}_1$. At each encoding step, $\mathcal{E}_3$ generates a value $X_{i,j}$ in a new location (i,j) in $\Delta^*_{m,n}$, as described in Figure 3. The value of $X_{i,j}$ depends on the values $X_{i-1,j}$, $X_{i-1,j+1}$, and $X_{i,j-1}$ (which are assumed to have already been generated), and also on at most one output bit of $\mathcal{E}_0$ or $\mathcal{E}_1$. To this end, there are two natural orders in which the values $X_{i,j}$ can be computed: they can be generated row by row, or diagonal by diagonal.

```
if X_{i-1,j} = X_{i,j-1} = 0 {
    if X_{i-1,j+1} = 0
        X_{i,j} is the output bit of E_0;
    else /* X_{i-1,j+1} = 1 */
        X_{i,j} is the output bit of E_1; }
else /* X_{i-1,j} = 1 or X_{i,j-1} = 1 */
    X_{i,j} = 0;
```

Figure 3: Encoding of $X_{i,j}$ by $\mathcal{E}_3$.

As we show in Corollary 4.3 in Section 4.1, when $\mu_{m,n}$ satisfies (4)–(11) and (12) we have $\mathsf{Prob}\{X_{i,j} = X_{i-1,j+1} = 0\} = \beta_0\sigma$ for all $(i,j) \in \Delta_{m,n} \setminus \Delta_n^{(\mathrm{h})}$. Hence, the expected number of locations $(i,j) \in \Delta^*_{m,n}$ for which $X_{i,j} = X_{i-1,j+1} = 0$, is $N = (m-1)(n-1)\beta_0\sigma$. This is the expected number of times that $\mathcal{E}_0$ or $\mathcal{E}_1$ are queried by $\mathcal{E}_3$. The expected number of times that $\mathcal{E}_0$ (respectively, $\mathcal{E}_1$) is queried is αN (respectively, $(1-\alpha)N$). Therefore, the expected rate of the overall coding scheme is

$$\begin{aligned} R_{m,n} &= \frac{(\alpha \cdot h(q_0) + (1-\alpha)\cdot h(q_1)) \cdot (1-\epsilon_{m,n}) \cdot N}{mn} \\ &= \frac{(m-1)(n-1)}{mn} \cdot \beta_0\sigma \cdot (\alpha \cdot h(q_0) + (1-\alpha)\cdot h(q_1))(1-\epsilon_{m,n}) \\ &= \frac{(m-1)(n-1)}{mn} \cdot (1-\epsilon_{m,n}) \cdot H(\boldsymbol{\mu}) \ , \end{aligned}$$

where $\lim_{m,n\to\infty} \epsilon_{m,n} = 0$. Namely, for $b \in \{0,1\}$, we bound from below the rates of $\mathcal{E}_b$ by $h(q_b)(1-\epsilon_{m,n})$; the factor $1-\epsilon_{m,n}$ also incorporates the ratio between the probability with which a typical word is generated by $\mathcal{E}_b$, compared to the probability with which such a word is generated by an ideal Bernoulli source (defined by q_b).

Simulations suggest that the rate $R_{m,n}$ is attained regardless of the boundary values set by $\mathcal{E}_2$; yet we have not proved this. On the other hand, there is clearly a fixed assignment for the boundaries that yields expected rate at least $R_{m,n}$. If we knew such an assignment, we could hard-wire it into the decoder, in which case it would be sufficient to transmit only the $(m-1)(n-1)$ non-boundary values of X, making $\mathcal{E}_2$ redundant.

Decoding is carried out as follows: Bits are read from the received $\Delta_{m,n}$-configuration in the order they were generated by the encoder, disregarding each 0 that immediately follows a 1 horizontally or vertically. The remaining bits are then divided into two bit streams according to the transformer $\mathcal{E}_b$ that generated each individual bit. The bit streams are then fed into the decoders (i.e., inverse mappings) of the respective transformers.

Our coding scheme can be simplified by combining the distribution transformers $\mathcal{E}_0$ and $\mathcal{E}_1$, in which case our encoder becomes the bit-stuffing encoder in [26] (except that the analysis here takes into account that stuffed 0's overlap, thereby improving on the lower bound of [26] on the expected rate). In such a case, we maximize $H(\boldsymbol{\mu}(q_0,q_1))$ in (13) under the restriction $q_0 = q_1$. The maximum is attained for $q_0 = q_1 \approx .644400$, and the maximum value is $H(\boldsymbol{\mu}(q_0,q_0)) \approx 0.583056$, which is within 1% of the capacity $\mathsf{cap}(\mathcal{S})$. We mention that this latter rate can be attained also by tuning the parameters in the method presented recently and independently in [15].

4 First-order Markov properties of $\boldsymbol{\mu}$

4.1 Horizontal first-order Markov process

In this section, we provide proofs for Proposition 2.1 and Proposition 2.2.

We start by verifying that the value of σ in (8) is the stationary probability of the first-order Markov process along the horizontal and diagonal boundaries. It is easy to see that

$$\sigma = \mu^{(\mathrm{h})}(0\,|\,0)\sigma + \mu^{(\mathrm{h})}(0\,|\,1)(1-\sigma) = \alpha\sigma + 1 - \sigma \ ,$$

thus implying that σ is indeed the stationary probability of $\mathsf{Prob}\{X_{0,j} = 0\}$ along the horizontal boundary. Similarly, by the choice of β_0 and β_1 in (11) we also have

$$\sigma = \mu^{(\mathrm{d})}(0\,|\,0)\sigma + \mu^{(\mathrm{d})}(0\,|\,1)(1-\sigma) = \beta_0\sigma + \beta_1(1-\sigma) \ ,$$

making σ also the stationary probability of $\mathsf{Prob}\{X_{i,-i} = 0\}$ along the diagonal boundary.

Denote by Φ_ℓ the set of all binary words of length $\le \ell$, and by Φ_ℓ^* the set of all nonempty binary words of length $\le \ell$.

For $(i,j) \in \Delta_{m,n}$ and $\boldsymbol{w} = w_1 w_2 \ldots w_\ell \in \Phi_{i+j+1}$, denote by $\mathcal{A}_{i,j}(\boldsymbol{w})$ the event

$$\mathcal{A}_{i,j}(\boldsymbol{w}) = \{X_{i,j}X_{i,j-1}X_{i,j-2}\ldots X_{i,j-\ell+1} = \boldsymbol{w}\} \ .$$

Also, for $(i,j) \in \Delta_{m,n} \setminus \Delta_n^{(\mathrm{h})}$ and $v \in \{0,1\}$, define the event

$$\mathcal{A}_{i,j}^{(v)}(\boldsymbol{w}) = \mathcal{A}_{i,j}(\boldsymbol{w}) \cap \{X_{i-1,j+1} = v\}$$

(see Figure 4). We also define the vectors

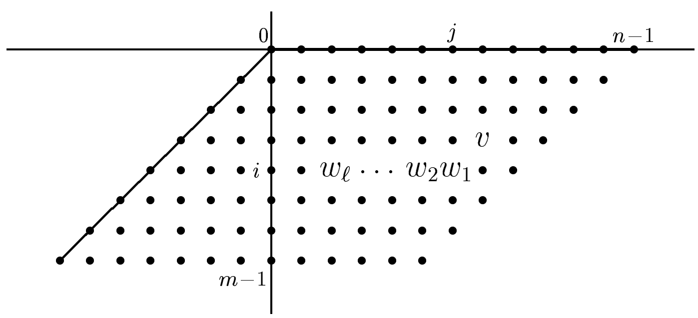

Figure 4: Event $\mathcal{A}_{i,j}^{(v)}(\boldsymbol{w})$.

$$\boldsymbol{A}_{i,j}(\boldsymbol{w}) = \begin{pmatrix} \mathsf{Prob}\{\mathcal{A}_{i,j}^{(0)}(\boldsymbol{w})\} \\ \mathsf{Prob}\{\mathcal{A}_{i,j}^{(1)}(\boldsymbol{w})\} \end{pmatrix} \ .$$

Note that by the way we set the diagonal boundary we have

$$\begin{aligned} \boldsymbol{A}_{i,-i}(0) &= \begin{pmatrix} \mathsf{Prob}\{X_{i,-i} = 0, X_{i-1,1-i} = 0\} \\ \mathsf{Prob}\{X_{i,-i} = 0, X_{i-1,1-i} = 1\} \end{pmatrix} = \begin{pmatrix} \beta_0\sigma \\ \beta_1(1-\sigma) \end{pmatrix} \\ &= \frac{1}{2-\alpha} \cdot \frac{1}{\alpha + q_1 - \alpha q_1} \cdot \begin{pmatrix} \alpha \\ (1-\alpha)q_1 \end{pmatrix} \end{aligned} \qquad (14)$$

and

$$\boldsymbol{A}_{i,-i}(1) = \begin{pmatrix} \mathsf{Prob}\{X_{i,-i}=1, X_{i-1,1-i}=0\} \\ \mathsf{Prob}\{X_{i,-i}=1, X_{i-1,1-i}=1\} \end{pmatrix} = \begin{pmatrix} (1-\beta_0)\sigma \\ (1-\beta_1)(1-\sigma) \end{pmatrix}$$
$$= \frac{1-\alpha}{2-\alpha} \cdot \frac{1}{\alpha+q_1-\alpha q_1} \cdot \begin{pmatrix} q_1 \\ \alpha(1-q_1) \end{pmatrix} . \tag{15}$$

Given $q_0, q_1, \alpha \in [0,1]$, we define the following 2×2 matrices:

$$P_{0,0} = \begin{pmatrix} \alpha q_0 & 1 \\ (1-\alpha)q_1 & 0 \end{pmatrix} , \quad P_{0,1} = \begin{pmatrix} \alpha & 1 \\ 1-\alpha & 0 \end{pmatrix} , \quad P_{1,0} = \begin{pmatrix} \alpha(1-q_0) & 0 \\ (1-\alpha)(1-q_1) & 0 \end{pmatrix} ,$$

and $P_{1,1} = 0$. The notations I and $\mathbf{1}$ will stand for the 2×2 identity matrix and the row vector $(1\ 1)$, respectively.

Recall that we say that row i in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ forms a first-order Markov process identical to the horizontal boundary if for every $1-i \le j < n-i$ and every word $\boldsymbol{c} = c_1 c_2 \ldots c_\ell \in \Phi^*_{i+j}$,

$$\mathsf{Prob}\{X_{i,j} = 0 \mid X_{i,j-1}X_{i,j-2}\ldots X_{i,j-\ell} = \boldsymbol{c}\} = \begin{cases} \alpha & \text{if } c_1 = 0 \\ 1 & \text{if } c_1 = 1 \end{cases} , \tag{16}$$

provided that the event we condition on has positive probability. The following lemma is easily verified.

Lemma 4.1 *For $m, n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11). Suppose that for some i in the range $1 \le i < m$, row $i-1$ in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ forms a first-order Markov process identical to the horizontal boundary. Then*

$$\boldsymbol{A}_{i,j}(b\boldsymbol{c}) = P_{b,c_1}\boldsymbol{A}_{i,j-1}(\boldsymbol{c}) \qquad \textit{for all } 1-i \le j < n-i \textit{ and } \boldsymbol{c} \in \Phi^*_{i+j} .$$

Proof of Proposition 2.1. We start with the "if" part and prove by induction on $i \ge 1$ that row i forms a first-order Markov process identical to the horizontal boundary. We do this by showing that (16) holds for every $1-i \le j < n-i$ and every word $\boldsymbol{c}$ of length *exactly* $i+j$ (which clearly implies that it holds for all shorter words).

First note that the sample space $\mathcal{S}(\Delta_{m,n})$ of X forces (16) to hold whenever the first bit in $\boldsymbol{c}$ is $c_1 = 1$.

We now consider words $\boldsymbol{c}$ with $c_1 = 0$. Our induction proof for row i assumes that row $i-1$ forms a first-order Markov process identical to the horizontal boundary. Clearly, this trivially holds for the induction base $i = 1$. Write $\boldsymbol{c} = 0\boldsymbol{w}$ where $\boldsymbol{w} \in \{0,1\}^{i+j-1}$, in which case (16) becomes

$$\mathsf{Prob}\{\mathcal{A}_{i,j}(00\boldsymbol{w})\} = \alpha \cdot \mathsf{Prob}\{\mathcal{A}_{i,j-1}(0\boldsymbol{w})\} .$$

This, in turn, is equivalent to

$$\mathbf{1}\boldsymbol{A}_{i,j}(00\boldsymbol{w}) = \alpha \cdot \mathbf{1}\boldsymbol{A}_{i,j-1}(0\boldsymbol{w}) . \tag{17}$$

By Lemma 4.1 and the induction hypothesis we have

$$\boldsymbol{A}_{i,j}(00\boldsymbol{w}) = P_{0,0}\boldsymbol{A}_{i,j-1}(0\boldsymbol{w}) \ .$$

Hence, (17) can be rewritten as

$$\mathbf{1}(P_{0,0} - \alpha I)\boldsymbol{A}_{i,j-1}(0\boldsymbol{w}) = 0 \ .$$

It follows that in order to show (16), it suffices to prove that for $1-i \leq j < n-i$ and $\boldsymbol{w} \in \{0,1\}^{i+j-1}$, the vector $\boldsymbol{A}_{i,j-1}(0\boldsymbol{w})$ is either the zero vector or a (right) eigenvector of $P_{0,0}$ associated with the eigenvalue α.

Now, it is easy to see that (12) implies that α is a nonnegative eigenvalue of $P_{0,0}$; indeed, α is the nonnegative root of the quadratic equation

$$\alpha^2(1-q_0) + \alpha q_1 - q_1 = 0 \tag{18}$$

obtained from the equality

$$\det(P_{0,0} - \alpha I) = \det \begin{pmatrix} \alpha q_0 - \alpha & 1 \\ (1-\alpha)q_1 & -\alpha \end{pmatrix} = 0 \ .$$

We now distinguish between two cases for the value of the word $\boldsymbol{w}$. Hereafter $\mathbf{0}_\ell$ stands for the all-zero word of length ℓ.

Case 1: $\boldsymbol{w} = \mathbf{0}_{i+j-1}$. By (14) it follows that $\boldsymbol{A}_{i,-i}(0)$ is an eigenvector of $P_{0,0}$ associated with the eigenvalue α. Hence, by Lemma 4.1 and the induction hypothesis we have

$$\boldsymbol{A}_{i,j-1}(\mathbf{0}_{i+j}) = P_{0,0}^{i+j-1}\boldsymbol{A}_{i,-i}(0) = \alpha^{i+j-1}\boldsymbol{A}_{i,-i}(0) \ , \quad 1-i \leq j < n-i \ ;$$

namely, $\boldsymbol{A}_{i,j-1}(\mathbf{0}_{i+j})$ is also an eigenvector of $P_{0,0}$ associated with the eigenvalue α.

Case 2: $\boldsymbol{w} \neq \mathbf{0}_{i+j-1}$. Write $\boldsymbol{w} = \mathbf{0}_{s-1}1\tilde{\boldsymbol{w}}$. If $s < i+j-1$ then $\tilde{\boldsymbol{w}}$ starts with a 0, or else we are in the trivial case in which the event we are conditioning on in (16) has zero probability (i.e., $\boldsymbol{A}_{i,j-s-1}(1\tilde{\boldsymbol{w}})$ is zero). By Lemma 4.1 and the induction hypothesis we obtain

$$\boldsymbol{A}_{i,j-s-1}(1\tilde{\boldsymbol{w}}) = P_{1,0}\boldsymbol{A}_{i,j-s-2}(\tilde{\boldsymbol{w}}) = A^{(0)} \cdot \begin{pmatrix} \alpha(1-q_0) \\ (1-\alpha)(1-q_1) \end{pmatrix} = \gamma \cdot \begin{pmatrix} q_1 \\ \alpha(1-q_1) \end{pmatrix} \tag{19}$$

for some real γ, where $A^{(0)}$ is the first coordinate of $\boldsymbol{A}_{i,j-s-2}(\tilde{\boldsymbol{w}})$ and where the last equality in (19) follows from (18). In fact, (19) also applies to $s = i+j-1$, in which case $\tilde{\boldsymbol{w}}$ is the empty word and $\boldsymbol{A}_{i,j-s-1}(1\tilde{\boldsymbol{w}}) = \boldsymbol{A}_{i,-i}(1)$, which, in turn, is given by (15).

Combining (19) with Lemma 4.1 we obtain

$$\boldsymbol{A}_{i,j-s}(01\tilde{\boldsymbol{w}}) = P_{0,1}\boldsymbol{A}_{i,j-s-1}(1\tilde{\boldsymbol{w}}) = \gamma \cdot \begin{pmatrix} \alpha & 1 \\ 1-\alpha & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ \alpha(1-q_1) \end{pmatrix} = \gamma \cdot \begin{pmatrix} \alpha \\ (1-\alpha)q_1 \end{pmatrix} \ .$$

That is, $\boldsymbol{A}_{i,j-s}(01\tilde{\boldsymbol{w}})$, if nonzero, is an eigenvector of $P_{0,0}$ associated with the eigenvalue α. Now, by Lemma 4.1 we have

$$\boldsymbol{A}_{i,j-1}(0\boldsymbol{w}) = P_{0,0}^{s-1}\boldsymbol{A}_{i,j-s}(01\tilde{\boldsymbol{w}}) = \alpha^{s-1}\boldsymbol{A}_{i,j-s}(01\tilde{\boldsymbol{w}}) \ .$$

We thus conclude that $\boldsymbol{A}_{i,j-1}(0\boldsymbol{w})$, if nonzero, is an eigenvector of $P_{0,0}$ associated with the eigenvalue α for every $1-i \le j < n-i$ and $\boldsymbol{w} \in \{0,1\}^{i+j-1}$. This establishes the induction step.

We now turn to the "only if" part. We show that the equality

$$\mathsf{Prob}\{X_{1,0} = 1 \mid X_{1,-1} = 0\} = 1 - \alpha \tag{20}$$

implies that α satisfies (18). Indeed,

$$\begin{aligned}
&\mathsf{Prob}\{X_{1,-1}X_{1,0} = 01\} \\
&\quad = \ \mathsf{Prob}\{X_{0,0}X_{1,-1}X_{1,0} = 001\} \\
&\quad = \ \mathsf{Prob}\{X_{0,0}X_{0,1}X_{1,-1}X_{1,0} = 0001\} + \mathsf{Prob}\{X_{0,0}X_{0,1}X_{1,-1}X_{1,0} = 0101\} \\
&\quad = \ \beta_0\sigma((1-q_0)\alpha + (1-q_1)(1-\alpha)) \ .
\end{aligned}$$

Combining this with (20), we obtain

$$\sigma(1-\alpha) = \beta_0\sigma((1-q_0)\alpha + (1-q_1)(1-\alpha)) \ ,$$

which, with (8) and (11), yields (18). □

Corollary 4.2 *For $m,n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11) and (12). Then for every $(i,j) \in \Delta^*_{m,n}$ and $\boldsymbol{w} \in \Phi_{i+j-1}$, the vector $\boldsymbol{A}_{i,j-1}(0\boldsymbol{w})$, if nonzero, is an eigenvector of $P_{0,0}$ associated with the eigenvalue α.*

Proof. By (16) we have

$$\mathbf{1}\boldsymbol{A}_{i,j}(00\boldsymbol{w}) = \alpha \cdot \mathbf{1}\boldsymbol{A}_{i,j-1}(0\boldsymbol{w})$$

for every $(i,j) \in \Delta^*_{m,n}$ and $\boldsymbol{w} \in \Phi_{i+j-1}$; so, by Lemma 4.1,

$$\mathbf{1}(P_{0,0} - \alpha I)\boldsymbol{A}_{i,j-1}(0\boldsymbol{w}) = 0 \ . \tag{21}$$

Now, the matrix $P_{0,0} - \alpha I$ is singular since α is an eigenvalue of $P_{0,0}$. On the other hand, $q_0 < 1$ implies that $\alpha \ne 1$; so, the vector $\mathbf{1}(P_{0,0} - \alpha I)$ is nonzero. Hence, that vector spans the rows of $P_{0,0} - \alpha I$, thus implying by (21) that $\boldsymbol{A}_{i,j-1}(0\boldsymbol{w})$ is an eigenvector of $P_{0,0}$ associated with the eigenvalue α. □

Corollary 4.3 *For $m,n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11) and (12). If $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ then*

$$\mathsf{Prob}\{X_{i,j} = X_{i-1,j+1} = 0\} = \beta_0\sigma$$

for every $(i,j) \in \Delta_{m,n} \setminus \Delta_n^{(\mathrm{h})}$.

Proof. Recall that for a $\Delta_{m,n}$-configuration $x \in \mathcal{S}(\Delta_{m,n})$, we denote by $\Lambda^{(\mathrm{d})}(x)$ the set of all $\Delta_{m,n+1}$-configurations in $\mathcal{S}(\Delta_{m,n+1})$ obtained from x by appending another diagonal. By the nesting property of $\mu_{m,n}$ we have

$$\mu_{m,n}(x) = \sum_{z \in \Lambda^{(\mathrm{d})}(x)} \mu_{m,n+1}(z)$$

for every $x \in \mathcal{S}(\Delta_{m,n})$. Hence, it suffices to show that when $Z \in_{\mu_{m,n+1}} \mathcal{S}(\Delta_{m,n+1})$, then

$$\mathsf{Prob}\{Z_{i,j} = Z_{i-1,j+1} = 0\} = \beta_0 \sigma$$

for every $(i,j) \in \Delta_{m,n} \setminus \Delta_n^{(\mathrm{h})}$, or—equivalently—for every $(i,j{+}1) \in \Delta^*_{m,n+1}$.

By Proposition 2.1 and Corollary 4.2, when applied to $Z \in_{\mu_{m,n+1}} \mathcal{S}(\Delta_{m,n+1})$, it follows that for every $(i,j{+}1) \in \Delta^*_{m,n+1}$, the vectors $\boldsymbol{A}_{i,j}(0)$ are eigenvectors of $P_{0,0}$ associated with the eigenvalue α; namely,

$$\boldsymbol{A}_{i,j}(0) = \begin{pmatrix} \mathsf{Prob}\{Z_{i,j} = 0, Z_{i-1,j+1} = 0\} \\ \mathsf{Prob}\{Z_{i,j} = 0, Z_{i-1,j+1} = 1\} \end{pmatrix} = \gamma_{i,j} \cdot \begin{pmatrix} \alpha \\ (1-\alpha)q_1 \end{pmatrix} \tag{22}$$

for some constants $\gamma_{i,j}$. On the other hand, we also have

$$\mathbf{1}\boldsymbol{A}_{i,j}(0) = \mathsf{Prob}\{Z_{i,j} = 0\} = \sigma \ .$$

Combining the latter equation with (22) we thus obtain

$$\boldsymbol{A}_{i,j}(0) = \frac{\sigma}{\alpha + (1-\alpha)q_1} \cdot \begin{pmatrix} \alpha \\ (1-\alpha)q_1 \end{pmatrix} = \begin{pmatrix} \beta_0 \sigma \\ \beta_1(1-\sigma) \end{pmatrix}$$

for every $(i,j{+}1) \in \Delta^*_{m,n+1}$ (compare with (14)). □

Proof of Proposition 2.2. By (4) and Corollary 4.3 we have

$$\begin{aligned} mnH(\mu_{m,n}) &= h(\sigma) + (n{-}1)\cdot\sigma\cdot h(\alpha) + (m{-}1)\cdot(\sigma\cdot h(\beta_0) + (1-\sigma)\cdot h(\beta_1)) \\ &\quad + \sum_{i=1}^{m-1} \sum_{j=-i+1}^{n-1-i} \mathsf{Prob}\{X_{i,j-1} = X_{i-1,j} = 0\} \\ &\qquad\quad \cdot \Big(\mathsf{Prob}\{X_{i-1,j+1} = 0 \mid X_{i,j-1} = X_{i-1,j} = 0\} \cdot h(q_0) \\ &\qquad\quad + \mathsf{Prob}\{X_{i-1,j+1} = 1 \mid X_{i,j-1} = X_{i-1,j} = 0\} \cdot h(q_1)\Big) \\ &= h(\sigma) + (n{-}1)\cdot\sigma\cdot h(\alpha) + (m{-}1)\cdot(\sigma\cdot h(\beta_0) + (1-\sigma)\cdot h(\beta_1)) \\ &\quad + (m{-}1)(n{-}1)\cdot\beta_0\sigma\cdot(\alpha\cdot h(q_0) + (1-\alpha)\cdot h(q_1)) \ . \end{aligned}$$

Therefore,

$$0 \le H(\mu_{m,n}) - \frac{(m{-}1)(n{-}1)}{mn}\cdot\beta_0\sigma\cdot(\alpha\cdot h(q_0) + (1-\alpha)\cdot h(q_1)) \le \frac{m{+}n{-}1}{mn} \ ,$$

as claimed. □

4.2 Diagonal first-order Markov process

In this section, we present a counterpart of Proposition 2.1 for the diagonals of $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$. We state the respective claims and point out the difference in proofs compared to those in Section 4.1.

For $(i, d{-}i) \in \Delta_{m,n}$ and $\boldsymbol{w} = w_1 w_2 \ldots w_\ell \in \Phi_i$, denote by $\mathcal{B}_{d,i}(\boldsymbol{w})$ the event

$$\mathcal{B}_{d,i}(\boldsymbol{w}) = \{X_{i,d-i} X_{i-1,d-(i-1)} X_{i-2,d-(i-2)} \ldots X_{i-\ell+1,d-(i-\ell+1)} = \boldsymbol{w}\} .$$

Also, for $(i, d{-}i) \in \Delta_{m,n} \setminus \Delta_m^{(\mathrm{d})}$ and $u \in \{0,1\}$, define the event

$$\mathcal{B}_{d,i}^{(u)}(\boldsymbol{w}) = \mathcal{B}_{d,i}(\boldsymbol{w}) \cap \{X_{i,d-1-i} = u\}$$

(see Figure 5). We also define the vectors

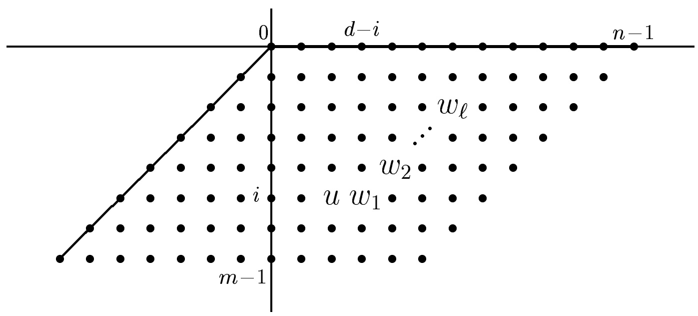

Figure 5: Event $\mathcal{B}_{d,i}^{(u)}(\boldsymbol{w})$.

$$\boldsymbol{B}_{d,i}(\boldsymbol{w}) = \begin{pmatrix} \mathsf{Prob}\{\mathcal{B}_{d,i}^{(0)}(\boldsymbol{w})\} \\ \mathsf{Prob}\{\mathcal{B}_{d,i}^{(1)}(\boldsymbol{w})\} \end{pmatrix} .$$

The counterparts of (14) and (15) take the form

$$\boldsymbol{B}_{d,0}(0) = \begin{pmatrix} \mathsf{Prob}\{X_{0,d} = 0, X_{0,d-1} = 0\} \\ \mathsf{Prob}\{X_{0,d} = 0, X_{0,d-1} = 1\} \end{pmatrix} = \begin{pmatrix} \alpha\sigma \\ 1-\sigma \end{pmatrix} = \frac{1}{2-\alpha} \cdot \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}$$

and

$$\boldsymbol{B}_{d,0}(1) = \begin{pmatrix} \mathsf{Prob}\{X_{0,d} = 1, X_{0,d-1} = 0\} \\ \mathsf{Prob}\{X_{0,d} = 1, X_{0,d-1} = 1\} \end{pmatrix} = \begin{pmatrix} (1-\alpha)\sigma \\ 0 \end{pmatrix} = \frac{1-\alpha}{2-\alpha} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} ,$$

and the counterparts of the matrices $P_{b,c}$ are

$$Q_{0,0} = \begin{pmatrix} \beta_0 q_0 & \beta_1 \\ 1-\beta_0 & 1-\beta_1 \end{pmatrix} , \qquad Q_{0,1} = \begin{pmatrix} \beta_0 q_1 & 0 \\ 1-\beta_0 & 0 \end{pmatrix} ,$$

$$Q_{1,0} = \begin{pmatrix} \beta_0(1-q_0) & 0 \\ 0 & 0 \end{pmatrix} , \qquad Q_{1,1} = \begin{pmatrix} \beta_0(1-q_1) & 0 \\ 0 & 0 \end{pmatrix} ,$$

where β_0 and β_1 are given by (11).

We say that diagonal d in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ forms a first-order Markov process identical to the diagonal boundary if for every $1 \le i < m$ and every word $\boldsymbol{c} = c_1 c_2 \ldots c_\ell \in \Phi_i^*$,

$$\mathsf{Prob}\{X_{i,d-i} = 0 \mid X_{i-1,d-(i-1)} X_{i-2,d-(i-2)} \ldots X_{i-\ell,d-(i-\ell)} = \boldsymbol{c}\} = \beta_{c_1} \, , \tag{23}$$

provided that the event we condition on has positive probability.

Lemma 4.4 *For $m, n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11). Suppose that for some d in the range $1 \le d < n$, diagonal $d-1$ in $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ forms a first-order Markov process identical to the diagonal boundary. Then*

$$\boldsymbol{B}_{d,i}(b\boldsymbol{c}) = Q_{b,c_1} \boldsymbol{B}_{d,i-1}(\boldsymbol{c}) \qquad \text{for all } 1 \le i < m \text{ and } \boldsymbol{c} \in \Phi_i^* \, .$$

Proposition 4.5 *For $m, n \ge 2$, $q_0 \in [0,1)$, and $q_1 \in (0,1]$, let $\mu_{m,n}$ be a measure defined on $\mathcal{S}(\Delta_{m,n})$ by (4)–(11). Then the entries in each diagonal $X \in_{\mu_{m,n}} \mathcal{S}(\Delta_{m,n})$ form a first-order Markov process identical to the diagonal boundary if and only (12) holds.*

Proof. The proof follows along the lines of the proof of Proposition 2.1, with a notable difference in the treatment of the case $c_1 = 1$ in (23). Specifically, in the "if" part, we also need to show that

$$\mathsf{Prob}\{\mathcal{B}_{d,i}(01\boldsymbol{w})\} = \beta_1 \cdot \mathsf{Prob}\{\mathcal{B}_{d,i-1}(1\boldsymbol{w})\} \qquad \text{for all } (i, d-i) \in \Delta_{m,n}^* \text{ and } \boldsymbol{w} \in \Phi_{i-1} \, . \tag{24}$$

The proof is carried out by induction on d, and the induction step assumes that (23) holds for diagonal $d-1$. Defining $\tau = \mathsf{Prob}\{\mathcal{B}_{d,i-1}^{(0)}(1\boldsymbol{w})\}$, we have

$$\boldsymbol{B}_{d,i-1}(1\boldsymbol{w}) = \begin{pmatrix} \tau \\ 0 \end{pmatrix} .$$

By the induction hypothesis and Lemma 4.4 we obtain

$$\boldsymbol{B}_{d,i}(01\boldsymbol{w}) = Q_{0,1} \boldsymbol{B}_{d,i-1}(1\boldsymbol{w}) = \tau \cdot \begin{pmatrix} \beta_0 q_1 \\ 1 - \beta_0 \end{pmatrix} .$$

Now, by (11) we have

$$\beta_1 = \beta_0 q_1 + 1 - \beta_0 \, .$$

Hence,

$$\mathbf{1}\boldsymbol{B}_{d,i}(01\boldsymbol{w}) = \mathbf{1}Q_{0,1}\boldsymbol{B}_{d,i-1}(1\boldsymbol{w}) = (\beta_0 q_1 + 1 - \beta_0)\tau = \beta_1 \tau = \beta_1 \cdot \mathbf{1}\boldsymbol{B}_{d,i-1}(1\boldsymbol{w}) \, ,$$

thus implying (24).

The case $c_1 = 0$ in (23) is treated by showing through the induction on d that for every $(i, d-i) \in \Delta_{m,n}^*$ and $\boldsymbol{w} \in \{0,1\}^{i-1}$, the vector $\boldsymbol{B}_{d,i-1}(0\boldsymbol{w})$, if nonzero, is an eigenvector of $Q_{0,0}$ associated with the eigenvalue β_0 given in (11). ☐

5 Fixed-rate encoding scheme

Let $B_{m,n}$ be the rectangle defined by (1). A $B_{m,n}$-configuration $x = (x_{i,j}) \in \mathcal{S}(B_{m,n})$ is called *circular* if for every $0 \le i < m$, the entries $x_{i,0}$ and $x_{i,n-1}$ are not both 1. The set of all $B_{m,n}$-configurations in $\mathcal{S}(B_{m,n})$ that are circular will be denoted by $\mathcal{S}^\circ(B_{m,n})$. Note also that if $x \in \mathcal{S}^\circ(B_{m,n})$ and if we define the $\Delta_{m,n}$-configuration $y = (y_{i,j})$ by

$$y_{i,j} = \begin{cases} x_{i,j} & \text{if } j \ge 0 \\ x_{i,n+j} & \text{otherwise} \end{cases}, \quad (i,j) \in \Delta_{m,n} ,$$

then $y \in \mathcal{S}(\Delta_{m,n})$.

In this section, we present a fixed-rate coding scheme into $\mathcal{S}^\circ(B_{m,n})$ with a rate that approaches 0.581074 for large values of n (or m). Our scheme borrows ideas from permutation codes [27], [30] combined with enumerative coding [9]. Even though the circular property is not necessary for the coding, it will make the analysis simpler. The $B_{m,n}$-configurations generated by the encoder will have the additional property that all rows in them have the same (Hamming) weight δn for a value $\delta \in [0,1]$ that will be determined in the sequel.

Let x be in $\mathcal{S}^\circ(B_{m,n})$ and assume that for some i in the range $1 \le i < m$, row $i-1$ in x has weight t. Let $j_1, j_2, \ldots, j_t$ be the indexes j for which $x_{i-1,j} = 1$. Clearly, we must have $x_{i,j_k} = 0$ for every $1 \le k \le t$. Define the words

$$x_i^{(k)} = x_{i,j_k+1} x_{i,j_k+2} \ldots x_{i,j_{k+1}-1} , \quad 1 \le k < t ,$$

and

$$x_i^{(t)} = x_{i,j_t+1} \ldots x_{i,n-1} x_{i,0} \ldots x_{i,j_1-1} .$$

The word $x_i^{(k)}$ is called the *kth phrase* of row i in x. Note that row i in x is obtained by shifting the word

$$0\, x_i^{(1)}\, 0\, x_i^{(2)}\, 0 \, \ldots \, 0\, x^{(t)}$$

cyclically j_1 entries to the right. The length of $x_i^{(k)}$ is called the *kth phrase length* in row i of x. Denoting that length by ℓ_k, the list $(\ell_1, \ell_2, \ldots, \ell_t)$ is called the *phrase profile* of row i in x. Clearly, $\ell_k = j_{k+1} - j_k - 1$, where j_{t+1} is defined to be $n + j_1$. Hence, the phrase profile of row i is completely determined by row $i-1$. Note also that $\sum_{k=1}^{t} \ell_k = n - t$. We mention that a (somewhat different) definition of phrases is used also in one-dimensional permutation codes [27], [30].

For a positive integer ℓ, let $\mathcal{S}_\ell = \mathcal{S}(B_{1,\ell})$ denote the set of all words of length ℓ that satisfy the one-dimensional $(1,\infty)$-RLL constraint. Similarly, we define $\mathcal{S}_\ell^\circ = \mathcal{S}^\circ(B_{1,\ell})$. Also, denote by $\mathcal{S}_{\ell,r}$ (respectively, $\mathcal{S}_{\ell,r}^\circ$) the set of words in $\mathcal{S}_\ell$ (respectively, $\mathcal{S}_{\ell,r}^\circ$) of weight r. It is easy to see that a $B_{m,n}$-configuration $x \in \mathcal{S}(B_{m,n})$ is in $\mathcal{S}^\circ(B_{m,n})$ if and only if every row in x is in $\mathcal{S}_n^\circ$.

Lemma 5.1 *For every two positive integers a and b such that $a+b \geq 3$, there is a mapping*

$$\phi_{a,b} \;:\; \mathcal{S}_2 \times \mathcal{S}_{a+b-2} \;\rightarrow\; \mathcal{S}_a \times \mathcal{S}_b$$

that is one-to-one and weight-preserving.

Proof. Since $a+b \geq 3$ and a and b play symmetrical roles, we can assume that $a \geq 2$. For $x_1x_2 \in \mathcal{S}_2$ and $y_1y_2\ldots y_{a+b-2} \in \mathcal{S}_{a+b-2}$, we define $\phi_{a,b}(x_1x_2, y_1y_2\ldots y_{a+b-2}) = (\boldsymbol{w}, \boldsymbol{w}')$, where $\boldsymbol{w}$ and $\boldsymbol{w}'$ are determined as follows: if $x_2 = 0$ or $y_1 = 0$, then we set

$$\boldsymbol{w} = x_1x_2y_1y_2\ldots y_{a-2} \quad \text{and} \quad \boldsymbol{w}' = y_{a-1}y_a\ldots y_{a+b-2}\;; \tag{25}$$

otherwise (if $x_1x_2 = 01$ and $y_1 = 1$),

$$\boldsymbol{w} = y_{b+1}y_{b+2}\ldots y_{a+b-2}x_1x_2 \quad \text{and} \quad \boldsymbol{w}' = y_1y_2\ldots y_b\,. \tag{26}$$

It is easy to see that the mapping $\phi_{a,b}$ is into $\mathcal{S}_a \times \mathcal{S}_b$ and is weight-preserving. To show that it is one-to-one, we need to verify that we can distinguish pairs $(\boldsymbol{w}, \boldsymbol{w}')$ generated by (25) from those that are generated by (26). Indeed, only in the latter, the last entry of $\boldsymbol{w}$ and the first entry of $\boldsymbol{w}'$ are both equal to 1. $\square$

Let x and y be two words in $\mathcal{S}_n$. We say that x is *consistent* with y if x and y form the rows of an array in $\mathcal{S}(B_{2,n})$. In other words, x and y do not have 1's in the same position.

Define

$$K(n,t) = \sum_{s=0}^{t-1} 2^s \cdot \binom{t-1}{s} \cdot |\mathcal{S}_{n-3t+2,t-s}|\,. \tag{27}$$

Lemma 5.2 *For every word $x \in \mathcal{S}^\circ_{n,t}$ there are at least $K(n,t)$ words $y \in \mathcal{S}^\circ_{n,t}$ that are consistent with x.*

Proof. The proof is based on the observation that the number of possible assignments for y depends only on the phrase profile of y, and only through the multiplicity (but not the order) with which each phrase length appears in that phrase profile. This phrase profile, in turn, is completely determined by x.

Assume first that x induces on y the phrase profile $(\ell_1, \ell_2, \ldots, \ell_t)$, where $\ell_1 = \ell_2 = \ldots = \ell_{t-1} = 2$ (and, so, $\ell_t = n-3t+2$). We refer to this profile as the *worst profile* for length n and weight t. Each of the phrases of length 2 in y can take a value from $\{00, 01, 10\}$. It follows that for $0 \leq s < t$, there are $2^s\binom{t-1}{s}$ ways to assign values to the phrases of length 2 in y so that their overall weight is s. If the overall weight of y is t, then the remaining phrase, of length $n-3t+2$, in y must have weight $t-s$. This proves the lemma assuming that x induces the worst profile on y.

It remains to establish that the worst profile is indeed the worst, in the sense that it leads to the smallest possible number of assignments for y. We show this by descending induction on the number of phrases in y whose lengths equal 2. Clearly, if at least $t-1$ phrase lengths equal 2, then, up to permutation of phrase lengths, the phrase profile is a worst profile and the claim immediately follows.

Turning to the induction step, suppose that x induces on y a phrase profile $L = (\ell_1, \ell_2, \ldots, \ell_t)$ in which $\ell_1 \neq 2$ and $\ell_2 \neq 2$. We can further assume that $\ell_1 \geq 3$; indeed, if all phrase lengths in L were less than 3, then we would have $n \leq 3t-2$, in which case $K(n,t) = 0$. We denote by Y the set of all words in $\mathcal{S}^\circ_{n,t}$ that are consistent with x.

Let x' be a word in $\mathcal{S}^\circ_{n,t}$ that induces the phrase profile $L' = (2, \ell_1+\ell_2-2, \ell_3, \ell_4, \ldots, \ell_t)$ on every word y' that is consistent with x'. The set of all such words y' in $\mathcal{S}^\circ_{n,t}$ will be denoted by Y'. Observe that L' has more phrase lengths equaling 2 than L does. So, by the induction hypothesis on Y' we have $|Y'| \geq K(n,t)$.

We next define a mapping $f : Y' \to Y$ as follows. Given $y' \in Y'$, the first two phrases of $f(y')$ are obtained by applying the mapping ϕ_{ℓ_1,ℓ_2} of Lemma 5.1 to the first two phrases of y'. The remaining $t-2$ phrases in $f(y')$ are identical to their counterparts in y'. By Lemma 5.1, the mapping f is one-to-one and weight-preserving and, so, $|Y| \geq |Y'| \geq K(n,t)$. □

Let $t_{\max} = t_{\max}(n)$ be the value of a nonnegative integer t for which $K(n,t)$ is maximized. Given n, m, and t (e.g., $t = t_{\max}(n)$), Lemma 5.2 suggests a coding scheme at a fixed rate $\lfloor m \log_2 K(n,t) \rfloor / (mn)$ into the set $\mathcal{S}^\circ(B_{m,n})$ as follows. For $i = 0, 1, \ldots, m-1$, we select row i from $\mathcal{S}^\circ_{n,t}$ so that it is consistent with row $i-1$ (for the case $i = 0$, we can assume a particular word from $\mathcal{S}^\circ_{n,t}$ to serve as a 'phantom' row -1). Lemma 5.2 guarantees that we have at least $K(n,t)$ words in $\mathcal{S}^\circ_{n,t}$ that can be selected for row i. This, in turn, implies the following result.

Proposition 5.3

$$\frac{\log_2 |\mathcal{S}^\circ(B_{m,n})|}{mn} \geq \frac{\log_2 K(n, t_{\max})}{n} .$$

The effective computation of row i in the suggested coding scheme can be done by enumerative coding, as we describe next [7],[25, p. 117],[28]. Let $(\ell_1, \ell_2, \ldots, \ell_t)$ be the phrase profile of row i as induced by row $i-1$. For this particular phrase profile, denote by $M_{k,s}$ the number of possible assignments for the first k phrases of row i so that their overall weight is s, $0 \leq s \leq t$. We have

$$M_{k,s} = \sum_{r=0}^{s} |\mathcal{S}_{\ell_k,r}| \cdot M_{k-1,s-r} , \quad 1 \leq k \leq t , \tag{28}$$

where $M_{0,0} = 1$ and $M_{0,s} = 0$ for $s > 0$. The values $|\mathcal{S}_{\ell,r}|$, in turn, can be computed by the recurrence

$$|\mathcal{S}_{\ell,r}| = |\mathcal{S}_{\ell-1,r}| + |\mathcal{S}_{\ell-2,r-1}| , \quad \ell \geq 2 , \tag{29}$$

where $|\mathcal{S}_{0,0}| = 1$, $|\mathcal{S}_{0,r}| = 0$ for $r \neq 0$, $|\mathcal{S}_{1,0}| = |\mathcal{S}_{1,1}| = 1$, and $|\mathcal{S}_{1,r}| = 0$ for $r \notin \{0,1\}$.

We can rewrite (28) and (29) in polynomial notation as follows. Let z be an indeterminate, and define the polynomials

$$T_\ell(z) = \sum_{r=0}^{\ell} T_{\ell,r} \cdot z^r = \sum_{r=0}^{\ell} |\mathcal{S}_{\ell,r}| \cdot z^r \;, \quad \ell \geq 1 \;,$$

and

$$M_k(z) = \sum_{s=0}^{t} M_{k,s} \cdot z^s \;, \quad 1 \leq k \leq t \;.$$

Then (29) becomes

$$T_\ell(z) = T_{\ell-1}(z) + z \cdot T_{\ell-2}(z) \;, \quad \ell \geq 2 \;, \tag{30}$$

where $T_0(z) = 1$ and $T_1(z) = 1 + z$. The recurrence (28), in turn, can be written as

$$M_k(z) \equiv T_{\ell_k}(z) \cdot M_{k-1}(z) \pmod{z^{t+1}} \;, \quad 1 \leq k \leq t \;,$$

where $M_0(z) = 1$. So,

$$M_k(z) \equiv \prod_{u=1}^{k} T_{\ell_u}(z) \pmod{z^{t+1}} \;, \quad 1 \leq k \leq t \;. \tag{31}$$

The latter formula can be used to accelerate the computation of the values $M_{k,s}$ through fast techniques of polynomial multiplication [1]. Note also that the polynomials $T_\ell(z)$ need to be computed for $1 \leq \ell \leq n{-}2t{+}1$ only once for the whole array.

The enumerative coding algorithm of row i is presented in Figure 6. The unconstrained input stream to be coded into row i is regarded as an integer p in the range $0 \leq p < K(n,t)$, and the phrase profile of row i is also assumed to be available. The main loop of the algorithm computes the phrases of row i, in reverse order, starting with the tth phrase. In each iteration of the main loop, the variable η determines the weight of the kth phrase, and s equals the overall weight of the first $k{-}1$ phrases. It can be easily verified by descending induction on k that each loop iteration starts with a value of p in the range $0 \leq p < M_{k,s}$, the induction base following from $0 \leq p < K(n,t) \leq M_{t,t}$. Similarly, the value of θ at the end of each loop iteration lies in the range $0 \leq \theta < T_{\ell_k,\eta} = |\mathcal{S}_{\ell_k,\eta}|$. The mapping from θ into a word in $\mathcal{S}_{\ell_k,\eta}$ assumes an ordering on the elements of each set $\mathcal{S}_{\ell,r}$. If the standard lexicographic ordering is used, then such a mapping can be efficiently implemented by (a second level of) enumerative coding, using the recurrence (29) (or (30)).

We next obtain an asymptotic estimate for $K(n,t)$ which will enable us to compute an asymptotic lower bound on $(\log_2 K(n, t_{\max}))/n$. The following lemma is a well-known asymptotic estimate for the binomial coefficients (see [17, p. 309]).

Lemma 5.4 *For $\ell \geq r \geq 0$ and $\ell > 0$,*

$$\log_2 \binom{\ell}{r} = \ell \cdot \left(h(r/\ell) - \epsilon(\ell, r) \right) \;,$$

where $\lim_{\ell\to\infty} \max_{0 \leq r \leq \ell} |\epsilon(\ell, r)| = 0$.

---

input: `integer` p `in the range` $0 \le p < K(n,t)$, `phrase profile` $(\ell_1, \ell_2, \ldots, \ell_t)$;

output: t `phrases (in descending order) of a row of an array in` $\mathcal{S}^\circ_{m,n}$;

initialize:

```
    compute T_ℓ(z) = Σ_{r=0}^{ℓ} T_{ℓ,r} · z^r by (30) for 1 ≤ ℓ ≤ max_k ℓ_k ;
    compute M_k(z) = Σ_{s=0}^{t} M_{k,s} · z^s by (31) for 1 ≤ k ≤ t;
    s ← t;
for k ← t downto 1 do    {
    η ← largest integer such that Σ_{r=0}^{η-1} T_{ℓ_k,r} · M_{k-1,s-r} ≤ p;
    p ← p − Σ_{r=0}^{η-1} T_{ℓ_k,r} · M_{k-1,s-r} ;
    s ← s − η;
    θ ← ⌊p/M_{k-1,s}⌋ ;
    set the kth phrase in the row to be the word indexed by θ in S_{ℓ_k,η} ;
    p ← p − θ · M_{k-1,s} ;
}
```

Figure 6: Enumerative coding into a row of an array in $\mathcal{S}^\circ(B_{m,n})$.

Lemma 5.5 *For $\ell \ge 2r \ge 0$ and $\ell > 0$,*

$$\log_2 |\mathcal{S}^\circ_{\ell,r}| = \log_2 \binom{\ell - r}{r} = (\ell - r) \cdot \big(h(r/(\ell - r)) - \epsilon(\ell - r, r)\big) ,$$

where $\lim_{\ell\to\infty} \max_{0\le r\le \ell} |\epsilon(\ell, r)| = 0$.

Proof. A word x is in $\mathcal{S}^\circ_{\ell,r}$ if and only if it can be written as a sequence of $\ell - r$ nonoverlapping blocks, r of which equaling 10 and the remaining equaling 0 (if the last entry in x is 1, then the last block will also include the first entry in x). Hence, $|\mathcal{S}^\circ_{\ell,r}|$ equals the number of combinations of r elements (being the indexes of the blocks 10 within x) out of $\ell - r$. □

It follows from Lemma 5.5 and the continuity of the function $t \mapsto h(t)$ that for every real $\rho \in [0, 1/2]$ we have

$$\lim_{\ell\to\infty} (1/\ell) \cdot \log_2 |\mathcal{S}^\circ_{\ell,\lceil \rho\ell \rceil}| = (1-\rho) \cdot h(\rho/(1-\rho)) ,$$

where $\lceil t \rceil$ stands for the smallest integer not greater than t. (Indeed, $(1-\rho)h(\rho/(1-\rho))$ is the entropy of a first-order Markov process defined on the $(1, \infty)$-RLL constraint, in which the probability of having 1 following 0 is $\rho/(1-\rho)$; the stationary probability of 1 is then ρ.)

Observing that

$$|\mathcal{S}^\circ_{n-3t+3,r}| \ge |\mathcal{S}_{n-3t+2,r}| \ge |\mathcal{S}_{n-3t,r}| \ge |\mathcal{S}^\circ_{n-3t,r}|$$

and that $\binom{t-1}{s} = \frac{t-s}{t} \cdot \binom{t}{s}$, we get from (27) the lower bound

$$\log_2 K(n,t) \ge \max_{0 \le s < t} \left\{ s + \log_2 \binom{t}{s} + \log_2 \tfrac{t-s}{t} + \log_2 |\mathcal{S}^{\circ}_{n-3t,t-s}| \right\} \tag{32}$$

and the upper bound

$$\log_2 K(n,t) \le \max_{0 \le s < t} \left\{ s + \log_2 \binom{t}{s} + \log_2 (t-s) + \log_2 |\mathcal{S}^{\circ}_{n-3t+3,t-s}| \right\} . \tag{33}$$

Write $\delta = t/n$, $\omega = s/n$, and $\rho = (t-s)/(n-3t)$. Assuming that $n-3t = (1-3\delta)n \ge 0$ and $\rho \le 1/2$, we can incorporate Lemmas 5.4 and 5.5 into the lower bound (32) to yield that, whenever ωn is a nonnegative integer less than δn,

$$\begin{aligned} \log_2 K(n, \delta n) &\ge \omega n + \log_2 \binom{\delta n}{\omega n} + \log_2 \tfrac{\delta - \omega}{\delta} + \log_2 |\mathcal{S}^{\circ}_{(1-3\delta)n,(1-3\delta)\rho n}| \\ &= (\omega + \delta \cdot h(\omega/\delta) + (1-3\delta) \cdot (1-\rho) \cdot h(\rho/(1-\rho)) - o(1)) \cdot n , \end{aligned}$$

where $o(1)$ stands for an expression that goes to zero as n goes to infinity. We now observe that $\omega = \delta - (1-3\delta)\rho$ and that $\omega \ge 0$ implies $\rho \le \delta/(1-3\delta)$. Hence, for every fixed rational $\delta \in [0, 1/3]$ and every n such that δn is an integer,

$$(1/n) \cdot \log_2 K(n, \delta n) \ge \sup_{\rho} F(\delta, \rho) - o(1) , \tag{34}$$

where

$$F(\delta, \rho) = \delta \cdot [1 + h((1/\delta - 3)\rho)] + (1-3\delta) \cdot [(1-\rho) \cdot h(\rho/(1-\rho)) - \rho] ,$$

and the supremum in the right-hand side of (34) is taken over all rational ρ in the range $0 \le \rho \le \min\{\delta/(1-3\delta), 1/2\}$. In fact, from the upper bound (33) it follows that the inequality in (34) can be replaced by an equality. Furthermore, since the function $F(\delta, \rho)$ is continuous we have

$$\liminf_{n \to \infty} (1/n) \cdot \log_2 K(n, \lceil \delta n \rceil) = \max_{\rho} F(\delta, \rho) ,$$

where the maximum is taken over all real $\rho \in [0, \min\{\delta/(1-3\delta), 1/2\}]$.

We now maximize the expression $F(\delta, \rho)$ over real values of $\delta \in [0, 1/3]$ and $\rho \in [0, \min\{\delta/(1-3\delta), 1/2\}]$. By taking partial derivatives of $F(\delta, \rho)$ with respect to δ and ρ, we get the equations

$$(23\delta - 4)(29\delta - 4)(8357\delta^5 - 8357\delta^4 + 3098\delta^3 - 518\delta^2 + 38\delta - 1) = 0$$

and

$$\rho = \frac{\delta \cdot (369\delta^2 - 101\delta + 4)}{1469\delta^3 - 682\delta^2 + 95\delta - 4} .$$

The maximum is attained for $(\delta_{\max}, \rho_{\max}) \approx (0.216594, 0.248986)$, in which case

$$\begin{aligned} &\liminf_{n \to \infty} (1/n) \cdot \log_2 K(n, t_{\max}(n)) \\ &\quad = \liminf_{n \to \infty} \max_{\delta} (1/n) \cdot \log_2 K(n, \lceil \delta n \rceil) \\ &\quad \ge \sup_{\delta} \liminf_{n \to \infty} (1/n) \cdot \log_2 K(n, \lceil \delta n \rceil) \\ &\quad = \max_{(\delta,\rho)} F(\delta, \rho) = F(\delta_{\max}, \rho_{\max}) \approx 0.581074 \end{aligned}$$

(in fact, one can easily show that in the third step—where we change the order between maximizing over δ and taking the limit over n—the inequality can be replaced by an equality).

Acknowledgment

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

References

[1] A.V. Aho, J.E. Hopcroft, and J.D. Ullman, *The Design and Analysis of Computer Algorithms,* Addison-Wesley, Reading, Massachusetts, 1974.

[2] J.J. Ashley, M. Blaum, B.H. Marcus, *Report on coding techniques for holographic storage,* IBM Research Report RJ 10013 (1996).

[3] R. Berger, *The Undecidability of the Domino Problem,* Mem. Amer. Math. Soc., Vol. 66, 1966.

[4] D. Brady, D. Psaltis, *Control of volume holograms, J. Opt. Soc. Am. A,* 9 (1992), 1167–1182.

[5] R. Burton, J.E. Steif, *Non-uniqueness of measures of maximal entropy for subshifts of finite type, Ergod. Th. Dynam. Sys.,* 14 (1994), 213–235.

[6] N. Calkin, H.S. Wilf, *The number of independent sets in a grid graph, SIAM J. Discrete Math.,* 11 (1997), 54–60.

[7] T.M. Cover, *Enumerative source encoding, IEEE Trans. Inform. Theory,* IT-19 (1973), 73–77.

[8] T.M. Cover, J.A. Thomas, *Elements of Information Theory,* Wiley, New York, 1991.

[9] S. Datta, S.W. McLaughlin, *An enumerative method for runlength-limited codes: permutation codes, IEEE Trans. Inform. Theory,* 45 (1999), 2199–2204.

[10] K. Engel, *On the Fibonacci number of an $m \times n$ lattice, Fibonacci Quarterly,* 28 (1990), 72–78.

[11] S. Forchhammer, J. Justesen, *Entropy bounds for constrained two-dimensional random fields, IEEE Trans. Inform. Theory,* 45 (1999), 118–127.

[12] J.F. Heanue, M.C. Bashaw, L. Hesselink, *Volume holographic storage and retrieval of digital data, Science,* 265 (1994), 749–752.

[13] J.F. Heanue, M.C. Bashaw, L. Hesselink, *Channel codes for digital holographic data storage, J. Opt. Soc. Am. A,* 12 (1995).

[14] H. Ito, A. Kato, Z. Nagy, K. Zeger, *Zero capacity region of multidimensional run length constraints, Electr. J. Combinatorics,* 6 (1999), R33.

[15] J. Justesen, Y.M. Shtarkov, *Simple models of two-dimensional information sources and codes, ISIT'98 — IEEE Int'l Symp. Inform. Theory,* Cambridge, Massachusetts, 1998.

[16] A. Kato, K. Zeger, *On the capacity of two-dimensional run length constrained channels, IEEE Trans. Inform. Theory,* 45 (1999), 1527–1540.

[17] F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes,* North Holland, Amsterdam, 1977.

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

[19] B.H. Marcus, P.H. Siegel, J.K. Wolf, *Finite-state modulation codes for data storage, IEEE J. Sel. Areas Comm.,* 10 (1992), 5–37.

[20] Z. Nagy, K. Zeger, *Capacity bounds for the* 3*-dimensional* $(0,1)$ *runlength limited channel, IEEE Trans. Inform. Theory,* 46 (2000), 1030–1033.

[21] D.K. Pickard, *A curious binary lattice process, J. Appl. Prob.,* 14 (1977), 717–731.

[22] D. Psaltis, F. Mok, *Holographic Memories, Scientific American,* 273, No. 5 (1995), 70–76.

[23] D. Psaltis, M.A. Neifeld, A. Yamamura, S. Kobayashi, *Optical memory disks in optical information processing, Applied Optics,* 29 (1990), 2038–2057.

[24] R.M. Robinson, *Undecidability and nonperiodicity for tilings of the plane, Inventiones Math.,* 12 (1971), 177–209.

[25] K.A. Schouhamer Immink, *Coding Techniques for Digital Recorders,* Prentice Hall, New York, 1991.

[26] P.H. Siegel, J.K. Wolf, *Bit-stuffing bounds on the capacity of* 2*-dimensional constrained arrays, ISIT'98 — IEEE Int'l Symp. Inform. Theory,* Cambridge, Massachusetts, 1998.

[27] D. Slepian, *Permutation codes, Proc. IEEE,* 53 (1965), 228–236.

[28] D.T. Tang, L.R. Bahl, *Block codes for a class of constrained noiseless channels, Inform. Control,* 17 (1970), 436–461.

[29] W. Weeks IV, R.E. Blahut, *The capacity and coding gain of certain checkerboard codes, IEEE Trans. Inform. Theory,* 44 (1998), 1193–1203.

[30] J.K. Wolf, *Permutation codes,* (d, k) *codes and magnetic recording, Proc. 1990 IEEE Colloq. South America, Argentina, Brazil, Chile, Uruguay,* W. Tompkins (Editor), 1990.