Universal finitary codes with exponential tails

Nate Harvey Alexander E. Holroyd* Yuval Peres†
Dan Romik

February 22, 2005 (revised April 27, 2006)

Abstract

In 1977, Keane and Smorodinsky showed that there exists a finitary homomorphism from any finite-alphabet Bernoulli process to any other finite-alphabet Bernoulli process of strictly lower entropy. In 1996, Serafin proved the existence of a finitary homomorphism with finite expected coding length. In this paper, we construct such a homomorphism in which the coding length has exponential tails. Our construction is source-universal, in the sense that it does not use any information on the source distribution other than the alphabet size and a bound on the entropy gap between the source and target distributions. We also indicate how our methods can be extended to prove a source-specific version of the result for Markov chains.

1 Introduction

Let $a, b \in \mathbb{N}$, and define the two finite alphabets $\mathbf{A} = \{i \in \mathbb{Z} : 1 \le i \le a\}$, $\mathbf{B} = \{i \in \mathbb{Z} : 1 \le i \le b\}$. Equip the sequence spaces $\mathbf{A}^{\mathbb{Z}}$ and $\mathbf{B}^{\mathbb{Z}}$ with the product σ-algebras $\mathcal{A}$ and $\mathcal{B}$ respectively. A measurable map $\varphi : \Omega \to \mathbf{B}^{\mathbb{Z}}$,

---

*Research funded in part by NSF (USA) grants DMS-0072398, CCR-0121555, NSERC (Canada) grant, CPAM, MSRI and BIRS.

†Research supported in part by NSF Grants DMS-0104073 and DMS-0244479.

Key words: Bernoulli process, entropy, finitary homomorphism, coding length, exponential tails, simulation of discrete distributions.

2000 Mathematics Subject Classifications: Primary 37A50; Secondary 28D20

where $\Omega \subseteq \mathbf{A}^{\mathbb{Z}}$ is measurable, is called **translation-equivariant** if for all $x = (x_i)_{i\in\mathbb{Z}} \in \Omega$, the left shift $\mathcal{T}(x) = (x_{i+1})_{i\in\mathbb{Z}}$ of x is also in Ω and the equality $\varphi(\mathcal{T}(x)) = \mathcal{T}(\varphi(x))$ holds. A translation-equivariant map $\varphi : \Omega \to \mathbf{B}^{\mathbb{Z}}$ is **finitary** if for all $x \in \Omega$, there exists an $N \in \mathbb{N}$ such that for all $y \in \Omega$, if $(x_i)_{|i|\le N} = (y_i)_{|i|\le N}$ then $\varphi(x)_0 = \varphi(y)_0$. In this case we let $N_\varphi(x)$ be the minimal such N, and call N_φ the **coding length** of φ.

If $p = (p(i))_{i\in\mathbf{A}}$ is a probability vector (that is to say, $p(i) \ge 0$ for all $i \in \mathbf{A}$, and $\sum_{i\in\mathbf{A}} p(i) = 1$), let $\mathbf{P}_p$ be the product measure $p^{\mathbb{Z}}$ on $\mathcal{A}$. The quadruple $B(p) = (\mathbf{A}^{\mathbb{Z}}, \mathcal{A}, \mathbf{P}_p, \mathcal{T})$ is the **Bernoulli shift** of p. Similarly if $q = (q(i))_{i\in\mathbf{B}}$ is a probability vector on $\mathbf{B}$, let $\mathbf{P}_q$ be the product measure $q^{\mathbb{Z}}$ on $\mathcal{B}$, and let $B(q) = (\mathbf{B}^{\mathbb{Z}}, \mathcal{B}, \mathbf{P}_q, \mathcal{T})$ be the Bernoulli shift of q. A **homomorphism** φ from $B(p)$ to $B(q)$ is a translation-equivariant map $\varphi : \Omega \to \mathbf{B}^{\mathbb{Z}}$ with the properties that $\Omega \in \mathcal{A}$, $\mathbf{P}_p(\Omega) = 1$, and for all $E \in \mathcal{B}$ we have $\mathbf{P}_p(\varphi^{-1}(E)) = \mathbf{P}_q(E)$.

Denote by $h(p) = -\sum_i p(i)\log p(i)$ the entropy of a probability vector. Keane and Smorodinsky [9] proved that if $h(p) > h(q)$, then there exists a finitary homomorphism φ from $B(p)$ to $B(q)$. Serafin [17] demonstrated that φ may be chosen in such a way that the expected coding length $\mathbf{E}_p(N_\varphi)$ is finite. Iwanik and Serafin [6] strengthened this result to all moments below the second.

We say that a finitary homomorphism has **exponential tails** if there exist $c > 0$ and $0 < d < 1$ such that $\mathbf{P}_p(N_\varphi \ge n) \le c \cdot d^n$ for all n. In general, we say that a non-negative sequence $(c_n)_{n\in\mathbb{N}}$ **decays exponentially** if there exist $c > 0$ and $0 < d < 1$ such that $c_n \le c \cdot d^n$ for all n. We say that a random variable W has **exponential tails** if $(\mathbf{P}(|W| \ge n))_{n\in\mathbb{N}}$ decays exponentially.

Our main result is a new construction of a finitary homomorphism from $B(p)$ to $B(q)$, when $h(p) > h(q)$. Our construction improves on the above-mentioned results in two ways. First, the coding length has exponential tails. Second, the homomorphism is source-universal, in the sense that the same function works simultaneously for all source vectors p over a given alphabet which have full support and whose entropy is greater than $h(q)$ by at least a given ε. In particular, this answers the open problem mentioned in the last two lines of [14].

The precise result is the following.

Theorem 1 *Fix a probability vector $q = (q(i))_{i\in\mathbf{B}}$, and fix $\varepsilon > 0$. There exists a measurable subset $\Omega \subseteq A^{\mathbb{Z}}$ and a finitary translation-equivariant map $\varphi : \Omega \to \mathbf{B}^{\mathbb{Z}}$, such that for any probability vector $p = (p(i))_{i\in\mathbf{A}}$ for which*

$p(i) > 0$ for all $i \in \mathbf{A}$ and $h(p) \geq h(q) + \varepsilon$, φ is a finitary homomorphism from $B(p)$ to $B(q)$ with exponential tails.

Here is a brief description of the motivation and method of the proof of Theorem 1. A homomorphism can be thought of as a translation-equivariant function that, given a sequence of independent samples with distribution p, simulates a sequence of independent samples with distribution q. Thus, it is natural to try to construct such a function using existing constructions of functions that simulate one discrete distribution using another. Such constructions have been described in [3], [4], [13], [14], [16].

Our construction combines elements from several of these constructions. First, the source sequence $(x_k)_{k\in\mathbb{Z}}$ is divided into **blocks** separated by **markers**. A marker is defined as an appearance of a certain (sufficiently rare) pattern, say a 2 followed by t 1's, where t is a large enough integer. Next, the contents of each block are fed into a specially designed function which converts these approximately independent p-distributed samples into independent unbiased bits. Next, at each block, these unbiased bits are fed into another function designed to simulate a number of independent samples of the distribution q sufficient to fill the length of that block. This function may require more bits than that block contains, but on average it requires less, because of entropy considerations. Any unused bits are then used to satisfy blocks whose simulation did not end in the first round. Continuing in this manner one obtains the required number of samples of q, which are then used to generate the value $\varphi(x)$. Everything is done simultaneously for all blocks in a translation equivariant manner, with an added bonus being the source-universality property.

The following is an extension of Theorem 1 to Markov chains.

Theorem 2 *Let $\alpha = (\alpha_{i,j})_{i,j\in\mathbf{A}}$, $\beta = (\beta_{i,j})_{i,j\in\mathbf{B}}$ be two aperiodic, irreducible Markov transition matrices over the finite alphabets $\mathbf{A}, \mathbf{B}$. Let $\mathcal{M}(\alpha) = (\mathbf{A}^{\mathbb{Z}}, \mathcal{A}, P_\alpha, \mathcal{T})$ and $\mathcal{M}(\beta) = (\mathbf{B}^{\mathbb{Z}}, \mathcal{B}, P_\beta, \mathcal{T})$ be the stationary Markov shifts of α and β respectively, and denote their entropies by $h(\alpha), h(\beta)$. If $h(\alpha) > h(\beta)$, then there exists a finitary homomorphism from $\mathcal{M}(\alpha)$ to $\mathcal{M}(\beta)$ with exponential tails.*

We indicate in Section 5 how our methods may be adapted to prove Theorem 2. For Markov chains, our construction is not source-universal, except in the weak sense that it will work, under the assumption of an entropy gap, simultaneously for all Markov transition matrices with all entries positive – see Section 5.

2 Simulations

In this section, we construct two procedures for simulating one discrete distribution from another. The constructions are variants of those used by Elias [3], Han and Hoshi [4], Knuth and Yao [13], and Romik [16]. In all of these constructions, a key point is that the loss in entropy is small.

2.1 Simulating a distribution from independent unbiased random bits

For $b \in \mathbb{N}$, let $\mathbf{B} = \{i \in \mathbb{Z} : 1 \leq i \leq b\}$ be a finite alphabet, and let $q = (q(i))_{i \in \mathbf{B}}$ be a probability vector. Let $(\{0,1\}^{\mathbb{N}}, \mathcal{F}, \mathbf{P})$ be the probability space of (one-sided) infinite binary strings, equipped with the natural product σ-algebra, and the probability measure under which the coordinate functions are independent unbiased random bits. Let $\mathbf{E}$ denote expectation with respect to the measure $\mathbf{P}$.

A **simulation of q from independent unbiased bits** is a pair of measurable functions $T : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ and $S : \{0,1\}^{\mathbb{N}} \to \mathbf{B}$, defined $\mathbf{P}$-a.s., with the following properties:

(i) If $x = (x_i)_{i \in \mathbb{N}}$ and $\tilde{x} = (\tilde{x}_i)_{i \in \mathbb{N}}$ are elements of $\{0,1\}^{\mathbb{N}}$ such that $(\tilde{x}_1, \ldots, \tilde{x}_{T(x)}) = (x_1, \ldots, x_{T(x)})$, then $T(\tilde{x}) = T(x)$ and $S(\tilde{x}) = S(x)$.

(ii) Under the measure $\mathbf{P}$, $S(x)$ has distribution q.

T is called the **stopping time** of the simulation, and S is called the **output symbol**.

The following theorem was first proved by Knuth and Yao [13]. We present an independent proof involving a more explicit construction.

Theorem 3 *There exists a simulation (T, S) of q from independent unbiased bits satisfying the additional properties:*

(i) $T(x)$ has exponential tails. More precisely, $\mathbf{P}(T(x) > k) \leq \frac{b+1}{2^k}$.

(ii) $\mathbf{E}(T(x)) \leq \frac{h(q)}{\log 2} + 6$.

Proof. Construct T and S as follows. Define a partition of $[0,1]$ by $0 = Q_0 < Q_1 < \ldots < Q_b = 1$, where $Q_j = \sum_{i=1}^{j} q(i)$. Define

$$T(x) = \min\left\{ k \in \mathbb{N} : \text{for some } 1 \le j \le b,\ \ Q_{j-1} < \sum_{i=1}^{k} \frac{x_i}{2^i} < Q_j - \frac{1}{2^k} \right\},$$

$$S(x) = \text{the unique } 1 \le j \le b \text{ for which } Q_{j-1} < \sum_{i=1}^{T(x)} \frac{x_i}{2^i} < Q_j - \frac{1}{2^{T(x)}}.$$

In words, the idea is to consider $x = (x_i)_{i\in\mathbb{N}}$ as the binary expansion of a number $u = \sum_{i=1}^{\infty} x_i 2^{-i} \in [0,1]$, and to define the output symbol $S(x)$ as that $1 \le j \le b$ for which $u \in (Q_{j-1}, Q_j)$. Determining the correct j necessitates looking at only the first $T(x)$ bits in the binary expansion of u. So $T(x), S(x)$ are defined for all x which are not the binary expansions of any of the Q_j, and property (i) in the definition of a simulation is clearly satisfied. Also, since, under the measure $\mathbf{P}$, u is uniformly distributed in $[0,1]$, we have that

$$\mathbf{P}(S(x) = j) = \text{Lebesgue measure of } (Q_{j-1}, Q_j) = q(j),$$

so property (ii) is also satisfied, and (T,S) is indeed a simulation of q from independent unbiased bits. To prove the additional properties claimed in the theorem, note that

$$\mathbf{P}(T(x) > k) = \mathbf{P} \bigcup_{j=0}^{b} \left\{ Q_j \in \left(\sum_{i=1}^{k} \frac{x_i}{2^i}, \sum_{i=1}^{k} \frac{x_i}{2^i} + \frac{1}{2^k} \right) \right\} \le \frac{b+1}{2^k},$$

establishing 3(i). For 3(ii), let for $0 \le j \le b$

$$Q_j = \sum_{i=1}^{\infty} a_{j,i} 2^{-i}, \qquad (a_{j,i} \in \{0,1\}, i \in \mathbb{N})$$

be a binary expansion of Q_j, and define for $1 \le j \le b$

$$m_j = \min\{k \in \mathbb{N} : a_{j-1,k} \ne a_{j,k}\} = \left\lceil -\frac{\log(Q_j - Q_{j-1})}{\log 2} \right\rceil.$$

Then, checking the definitions we see that, for $l \ge m_j$,

$$\begin{aligned} &\{x : S(x) = j, T(x) = l\} \\ &\qquad = \{x : (x_1, \ldots, x_l) = (a_{j-1,1}, \ldots, a_{j-1,l-1}, 1), a_{j-1,l} = 0\} \\ &\qquad\qquad \cup\ \{x : (x_1, \ldots, x_l) = (a_{j,1}, \ldots, a_{j,l-1}, 0), a_{j,l} = 1\}, \end{aligned} \tag{1}$$

while for $l < m_j$, this event is empty. Since

$$\mathbf{E}(T(x)) = \sum_{j=1}^{b} \sum_{l=1}^{\infty} l \cdot \mathbf{P}(T(x) = l, S(x) = j),$$

3(ii) will follow if we prove that for all $1 \le j \le b$,

$$\sum_{l=1}^{\infty} l \cdot \mathbf{P}(T(x) = l, S(x) = j) \le -\frac{q(j)\log q(j)}{\log 2} + 6q(j).$$

Indeed, by using (1) we see that the representation

$$q(j) = \mathbf{P}(S(x) = j) = \sum_{l=m_j}^{\infty} \mathbf{P}(S(x) = j, T(x) = l)$$

can be rewritten as a representation of $q(j)$ as a sum of negative powers of 2, namely

$$q(j) = \sum_{i=1}^{\infty} 2^{-n_{j,i}},$$

where each summand $2^{-n_{j,i}}$ is the probability of one of the two events on the right-hand side of (1). Arrange the $n_{j,i}$ such that $n_{j,1} \le n_{j,2} \le n_{j,3} \le \cdots$, and note also that $n_{j,i} < n_{j,i+2}$ for all $i \in \mathbb{N}$, since any given power of 2 appears at most twice in the sum. Then in particular we get

$$\begin{aligned} 2^{-n_{j,1}} \le q(j) &\le 2^{-n_{j,1}} + 2^{-n_{j,1}} + 2^{-n_{j,1}-1} + 2^{-n_{j,1}-1} + 2^{-n_{j,1}-2} + \ldots \\ &= 2\left(1 + \frac{1}{2} + \frac{1}{4} + \ldots\right) 2^{-n_{j,1}} = 2^{2-n_{j,1}}, \end{aligned}$$

so $n_{j,1} < -\frac{\log q(j)}{\log 2} + 2$. This then gives

$$\begin{aligned} \sum_{l=1}^{\infty} l \cdot \mathbf{P}(T(x) = l, S(x) = j) &= \sum_{i=1}^{\infty} n_{j,i} 2^{-n_{j,i}} \\ &= \sum_{i=1}^{\infty} n_{j,1} 2^{-n_{j,i}} + \sum_{i=1}^{\infty} (n_{j,i} - n_{j,1}) 2^{-n_{j,i}} \\ &\le \left(-\frac{\log q(j)}{\log 2} + 2\right) \sum_{i=1}^{\infty} 2^{-n_{j,i}} + 2^{-n_{j,1}} \cdot 2 \sum_{k=0}^{\infty} k 2^{-k} \\ &\le -\frac{q(j)\log q(j)}{\log 2} + 2q(j) + 4q(j), \end{aligned}$$

as required. $\square$

Remarks. Knuth and Yao [13] proved Theorem 3 with the better constant 2 replacing 6 in Theorem 3(ii). Their construction is described in slightly less concrete terms than the one above, but it is optimal, in the sense that the constant 2 is sharp, and in the strong sense that the stopping time is stochastically dominated by the stopping time of *any* possible simulation of q using independent unbiased bits. For more information see Section 5.12 of [1].

A generalization of the construction given above to simulation of q from a source of independent samples of an arbitrary probability vector p was given by Han and Hoshi [4] and independently by Romik [16], both of whom proved a statement analogous to Theorem 3(ii), with the base-2 entropy of q replaced by the ratio $h(q)/h(p)$, and with the constant 6 replaced by some function of the vector p.

2.2 Simulating independent unbiased random bits from a block with an excluded pattern

For $a \in \mathbb{N}$, let $\mathbf{A} = \{i \in \mathbb{Z} : 1 \leq i \leq a\}$ be a finite alphabet, and let $p = (p(i))_{i\in\mathbf{A}}$ be a probability vector. For each $n \in \mathbb{N}$, let $(\mathbf{A}^n, p^n)$ be the discrete elementary probability space of $\mathbf{A}$-valued n-tuples, with the i.i.d. product measure with marginal probabilities p. In the case $a = 2$ of a binary distribution, Elias [3] constructed a function that, given an $\mathbf{A}^n$-valued, p^n-distributed input, simulates a *random number* of independent unbiased random bits; that is, a pair (N, F), where N is an $\mathbb{N}$-valued random variable, and for each $k \in \mathbb{N}$, given $N = k$, the random vector F is distributed uniformly on the set $\{0,1\}^k$.

We shall need a generalization of Elias's construction, which takes as input n independent samples from a general discrete distribution, conditioned on the non-appearance of a certain pattern, and returns a random number of independent unbiased random bits.

For any $t \in \mathbb{N}$, let $E_{n,t}$ be the subset of $\mathbf{A}^n$ consisting of all vectors $x = (x_1, \ldots, x_n) \in \mathbf{A}^n$ for which for no $i \in \{1, \ldots, n-t\}$ is it true that

$$x_i = 2,\ x_{i+1} = x_{i+2} = \ldots = x_{i+t-1} = 1.$$

That is, $E_{n,t}$ contains all vectors which, considered as words, do not contain the pattern "2 followed by $t-1$ 1's". We sometimes call such vectors "pattern-avoiding". Let $\tilde{p}_{n,t}$ be the measure p^n conditioned on $E_{n,t}$. Let $\{0,1\}^* = \cup_{k=0}^{\infty}\{0,1\}^k$ be the set of finite strings over the alphabet $\{0,1\}$.

Theorem 4 *For any $n, t \in \mathbb{N}$ there exist functions*

$$\begin{aligned} N_{n,t} &: E_{n,t} \to \{0,1,2,\ldots\}, \\ F_{n,t} &: E_{n,t} \to \{0,1\}^*, \\ G_{n,t} &: E_{n,t} \to \{1,2,\ldots,n^{a-1}\} \end{aligned}$$

with the properties:

(i) under the measure $\tilde{p}_{n,t}$, for each $k \in \{0,1,2,\ldots\}$ for which $\tilde{p}_{n,t}(N_{n,t}(x) = k) > 0$, conditioned on $N_{n,t}(x) = k$, the random vector $F_{n,t}(x)$ is uniformly distributed on $\{0,1\}^k$;

(ii) the function $x \to (N_{n,t}(x), F_{n,t}(x), G_{n,t}(x))$ is injective;

(iii) for any $x \in E_{n,t}$ we have $N_{n,t}(x) \le (\log a / \log 2)n$.

Before going on with the proof, we explain briefly the idea behind this construction and its importance in what follows. The functions $N_{n,t}, F_{n,t}, G_{n,t}$ accept as input a $\tilde{p}_{n,t}$-distributed random variable and produce a binary string, $F_{n,t}(x)$, of length $N_{n,t}(x)$. Conditioned on the number of bits, the binary string is distributed uniformly over all binary strings of that length, in other words contains $N_{n,t}(x)$ independent unbiased bits. We would like to ensure that the construction is efficient, i.e. it extracts enough information from the input. This is guaranteed by claim (ii), which states that adding the complementary information $G_{n,t}(x)$ makes the function injective, together with the fact that the range of $G_{n,t}(x)$ is relatively small, so the amount of entropy contained in it is limited. As for claim (iii), it will be used in our proof that the homomorphism we construct has exponential tails.

Proof of Theorem 4. Throughout the proof, for convenience we shall consider n and t as fixed and in most places omit reference to the dependence of the various quantities on them. To construct the functions $N_{n,t}, F_{n,t}, G_{n,t}$, we first divide $E_{n,t}$ into classes of equiprobable elements. We do this as follows. Let

$$C = \{m = (m_1, m_2, \ldots, m_a) \in \mathbb{Z}^a : m_i \ge 0,\ 1 \le i \le a,\ m_1 + \ldots + m_a = n\}.$$

For $x = (x_1, \ldots, x_n) \in E_{n,t}$ and $1 \le i \le a$, let

$$c_i(x) = \#\{1 \le j \le n : x_j = i\},$$

and let

$$\text{count}(x) = (c_1(x), \dots, c_a(x)) \in C.$$

Then clearly $E_{n,t}$ can be written as the disjoint union

$$E_{n,t} = \bigcup_{m \in C} \{x \in E_{n,t} : \text{count}(x) = m\} =: \bigcup_{m \in C} D_m.$$

For each $m \in C$, we have for all $x \in D_m$ that

$$p^n(x) = p(1)^{m_1} p(2)^{m_2} \dots p(a)^{m_a},$$

so

$$\tilde{p}_{n,t}(x) = \frac{p(1)^{m_1} p(2)^{m_2} \dots p(a)^{m_a}}{p^n(E_{n,t})}.$$

In other words, all elements of D_m are equiprobable under $\tilde{p}_{n,t}$. Now, for each $m \in C$, let $d_m = |D_m|$, and write

$$d_m = \sum_{i=1}^{s_m} 2^{r_{m,i}}, \qquad r_{m,1} > r_{m,2} > \dots > r_{m,s_m} \geq 0,$$

for the binary expansion of d_m. The functions $N_{n,t}$, $F_{n,t}$ and $G_{n,t}$ may now be defined as follows. For each $m \in C$, arrange the elements of D_m in lexicographical order, and for each $x \in D_m$ denote by $\text{rank}(x)$ the position of x in this order. Set for each $x \in D_m$

$$\begin{aligned}
N_{n,t}(x) &= r_{m,k^*(x)}, \quad k^*(x) = \min\left\{1 \leq k \leq s_m : \sum_{i=1}^{k} 2^{r_{m,i}} \geq \text{rank}(x)\right\}, \\
F_{n,t}(x) &= \text{the length-}N_{n,t}(x) \text{ binary expansion of the number} \\
&\quad \sum_{i=1}^{k^*(x)} 2^{r_{m,i}} - \text{rank}(x), \\
G_{n,t}(x) &= \text{the position of } m \text{ in the lexicographical order on } C.
\end{aligned}$$

The functions $N_{n,t}$, $F_{n,t}$, $G_{n,t}$ are clearly defined on all $E_{n,t}$ and have the desired range. In words, we have used the lexicographical order to give an explicit partition of D_m into subsets of sizes $2^{r_{m,i}}$, $1 \leq i \leq s_m$. On each subset of size $2^{r_{m,i}}$ we define $N_{n,t} = r_{m,i}$, and for the value of $F_{n,t}$ assign to the $2^{r_{m,i}}$ possible elements the $2^{r_{m,i}}$ different binary strings of length $r_{m,i}$.

This implies claim 4(i), since the elements of D_m are equiprobable. The function $G_{n,t}$ is defined so as to encode the residual information needed to recover the value of x given $F_{n,t}(x)$. Indeed, if $x, x' \in E_{n,t}$ and $x \neq x'$, then either $x \in D_m$, $x \in D_{m'}$ for some $m \neq m'$, in which case clearly $G_{n,t}(x) \neq G_{n,t}(x')$, or x, x' are in the same D_m but $\mathrm{rank}(x) \neq \mathrm{rank}(x')$, whence $F_{n,t}(x) \neq F_{n,t}(x')$. This proves claim 4(ii). Finally, claim 4(iii) is immediate, since for all $m \in C$, $1 \leq i \leq s_m$ we have $2^{r_{m,i}} \leq d_m \leq |E_{n,t}| \leq a^n$, so $N_{n,t}(x) = r_{m,k^*(x)} \leq (\log a / \log 2)n$. $\square$

3 Construction of the homomorphism

We now construct the homomorphism φ that will be used to prove Theorem 1. We call the sequence $(x_i)_{i\in\mathbb{Z}}$ the **input sequence**, and the resulting sequence $(\varphi(x)_i)_{i\in\mathbb{Z}}$ the **output sequence**. For convenience, we fix a source probability vector $p = (p(i))_{i\in\mathbf{A}}$, which we assume has full support and satisfies $h(p) \geq h(q) + \varepsilon$, and denote $\mathbf{P} = \mathbf{P}_p$. We may also assume without loss of generality that $0 \notin (q(i))_{i\in\mathbf{B}}$. For an alphabet $\mathbf{A}$, denote by $\mathbf{A}^* = \cup_{n\geq 0}\mathbf{A}^n$ the set of **finite words over A**. For each $w \in \mathbf{A}^*$, denote by $\mathrm{length}(w)$ the length of w.

The construction is done in several stages. Here is an informal description of the steps, which are also drawn schematically in Figure 1.

Step 0: Fix a parameter $t \in \mathbb{N}$, the **marker length**. Its value will be some large integer that will be determined later, and will only depend on the target distribution q and the entropy gap bound ε.

Step 1: Divide the input sequence into **blocks**. A **marker** is an index i for which

$$x_i = 2,\ x_{i+1} = x_{i+2} = \ldots = x_{i+t-1} = 1.$$

Enumerate the markers as $\ldots, R_{-2}, R_{-1}, R_0, R_1, \ldots$, where R_1 is the first marker to the right of the origin. A **block** is the set of indices between two markers, namely $\{i : R_k < i \leq R_{k+1}\}$. The **input word associated with block *k*** is the sequence $W_k = (x_i)_{R_k+t\leq i\leq R_{k+1}-1}$, namely the sequence of input symbols in block k, not including the "211...1" patterns.

Under the measure $\mathbf{P}$, the input words $\ldots, W_{-1}, W_0, W_1, \ldots$ are independent $\mathbf{A}^*$-valued random variables. The words $(W_k)_{k\in\mathbb{Z}\setminus\{0\}}$ are identically distributed. (Note that W_0 has a different distribution owing to "size-biasing").

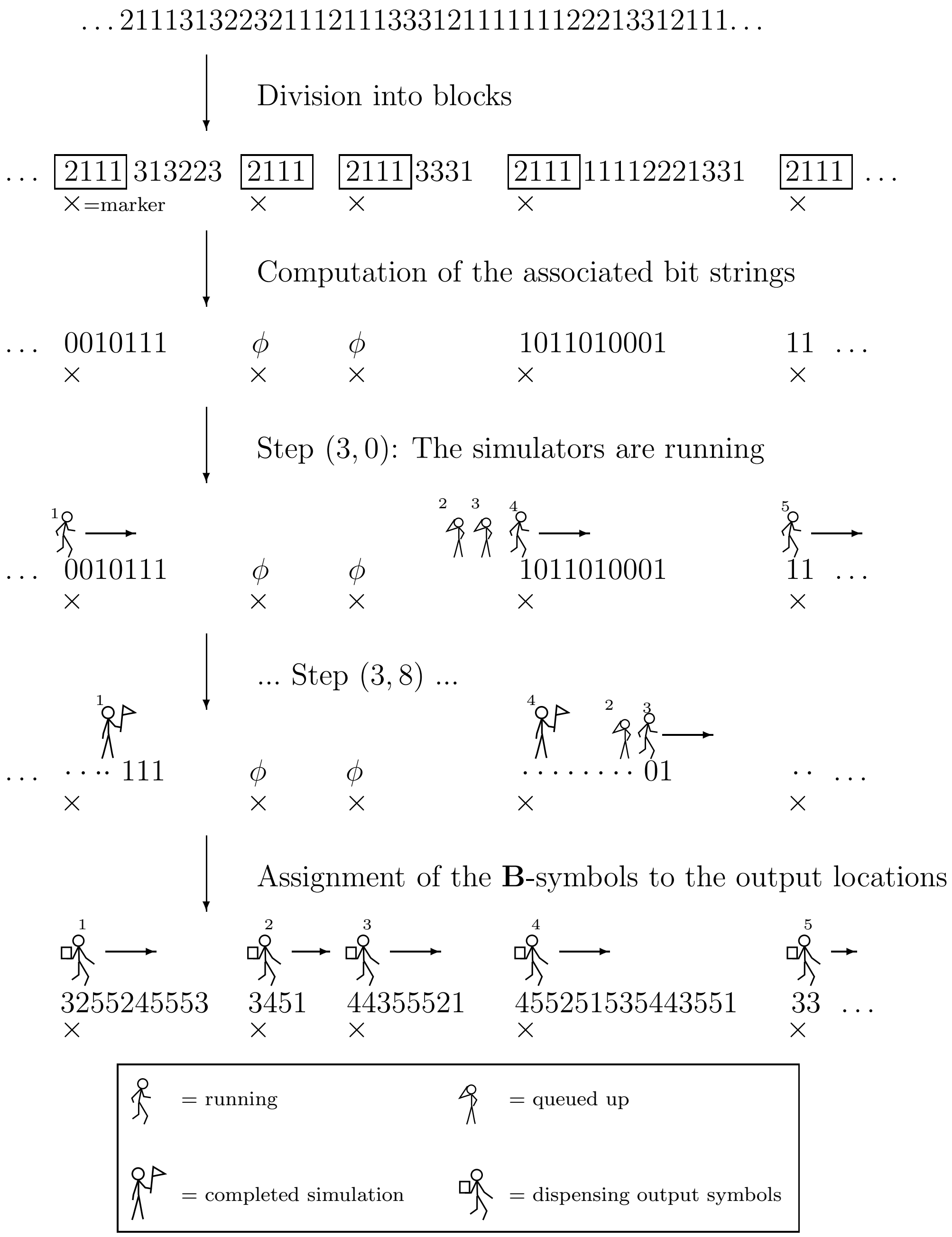

Figure 1: Illustration of φ.

All the input words have the property that, conditioned on the length of W_k being equal to n, W_k has distribution $\tilde{p}_{n,t}$.

Step 2: Apply to each input word W_k the function $F_{n,t}$ from Section 2.2, where n is the length of W_k, to obtain a string $U_k = F_{n,t}(W_k)$ of $N_{n,t}(W_k)$ independent unbiased random bits. U_k is called the **bit string associated with block $\boldsymbol{k}$**.

Step 3: For each block $k \in \mathbb{Z}$, attempt to use the random bits in U_k to simulate a $\mathbf{B}^{R_{k+1}-R_k}$-valued random variable B_k with distribution $q^{R_{k+1}-R_k}$, using the simulation (T, S) from Section 2.1. In many blocks, the stopping time will be reached. If the stopping time is reached, U_k may contain unused bits, which are still independent and unbiased. For any block k whose stopping time is not reached, look at U_{k+1} in the next block to the right to find unused bits to continue the simulation. If now the stopping time is reached, compute B_k. If not, iterate, looking one block further to the right at each step for unused bits, until the stopping time is reached. This iteration is done simultaneously for all blocks, in order to maintain translation-equivariance of the construction. A further complication arises because some bit strings may have length zero, so two or more simulators may try to read the same bits at the same time. In such situations we give priority to the simulator belonging to the rightmost block. We will refer to such a situation as a **queue-up**.

The ergodic theorem will ensure that, for a proper choice of the parameter t, a.s. enough bits are present overall to complete the process for all $k \in \mathbb{Z}$. Having computed B_k, the $\mathbf{B}$-symbols it contains are assigned to the indices of the k-th block, to produce the output sequence $\varphi(x)$.

Each block length has exponential tails. For each $k \in \mathbb{Z}$, the number of blocks which must be examined in Step 3 to simulate B_k has exponential tails. It follows that this homomorphism has exponential tails.

Formal definition of φ

Let $x = (x_i)_{i\in\mathbb{Z}}$. Let t, the **marker length**, be a positive integer to be chosen later.

We first define the marker locations, $(R_k)_{k\in\mathbb{Z}}$. Let

$$R_1 = \min\{i \geq 0 : x_i = 2, x_{i+1} = x_{i+2} = \ldots = x_{i+t-1} = 1\}.$$

Inductively, for $k \geq 2$, let

$$R_k = \min\{i > R_{k-1} : x_i = 2, x_{i+1} = x_{i+2} = \ldots = x_{i+t-1} = 1\}.$$

Let

$$R_0 = \max\{i < 0 : x_i = 2, x_{i+1} = x_{i+2} = \ldots = x_{i+t-1} = 1\}.$$

Inductively, for $k \geq 1$, let

$$R_{-k} = \max\{i < R_{-k+1} : x_i = 2, x_{i+1} = x_{i+2} = \ldots = x_{i+t-1} = 1\}.$$

It follows from well-known facts of elementary probability theory that $(R_k)_{k\in\mathbb{Z}}$ are defined for $\mathbf{P}$-almost every input sequence $(x_i)_{i\in\mathbb{Z}}$. For all $k \in \mathbb{Z}$, let $\{i : R_k < i \leq R_{k+1}\}$ be the k-th **block**. Let $W_k = (x_i)_{R_k+t\leq i\leq R_{k+1}-1}$ be the **input word** associated with block k, an $\mathbf{A}^*$-valued random variable, and let $L_k = \text{length}(W_k) = R_{k+1} - R_k - t$ be its length. Let $\lambda_k = R_{k+1} - R_k$. Note that by definition, W_k does not contain the pattern "2 followed by $t-1$ 1's".

Let $N_{n,t}, F_{n,t}, G_{n,t}$ be as in Section 2.2. For all $k \in \mathbb{Z}$, let $U_k = F_{L_k,t}(W_k)$. U_k is a $\{0,1\}^*$-valued random variable, called the **bit string** associated with block k. Denote its length by $V_k = \text{length}(U_k) = N_{L_k,t}(W_k)$. Let $U_k = (\epsilon_k(1), \epsilon_k(2), \ldots, \epsilon_k(V_k))$ be the bits comprising U_k.

For any $\ell \in \mathbb{N}$, let the pair (T_ℓ, S_ℓ) be a simulation of the distribution q^ℓ from independent unbiased bits, as in Section 2.1. (Recall that T_ℓ is the stopping time and S_ℓ is the output symbol of the simulation). For an input $x \in \{0,1\}^*$, say that **the simulation** (T_ℓ, S_ℓ) **is successful for input** x if for some $y \in \{0,1\}^{\mathbb{N}}$ (and hence also, by the definition of a simulation, for *all* $y \in \{0,1\}^{\mathbb{N}}$) we have $T_\ell(x * y) \leq \text{length}(x)$, where $x * y$ is the concatenation of x by y.

In Step 3, for each $k \in \mathbb{Z}$ we will generate a $\mathbf{B}^*$-valued random variable $B_k = (\beta_k(1), \ldots, \beta_k(\lambda_k))$ such that, conditioned on $(\lambda_k)_{k\in\mathbb{Z}}$, the B_k are independent and each B_k has distribution q^{λ_k}. To do this, in each Step (3,n) for $n \geq 0$, to each block $k \in \mathbb{Z}$ we assign the following: A pair (J_k^n, M_k^n) with $J_k^n \geq k$ and $1 \leq M_k^n \leq V_{J_k^n}$, called the **position of the k-th simulator at Step $(3, n)$** (here "position (j, k)" refers to the kth bit of block j); a word $\mathcal{Z}_k^n \in \{0,1\}^*$ called the **input read by the k-th simulator by Step $(3, n)$**; and a set $\mathcal{G}_k^n$ of pairs $(j, m) \in \mathbb{Z} \times \mathbb{N}$ such that $\mathcal{Z}_k^n$ is the concatenation of all the bits $\epsilon_j(m)$, $(j, m) \in \mathcal{G}_k^n$, arranged in lexicographical order on (j, m), called the **set of positions used by the k-th simulator by Step $(3, n)$**. If for input $\mathcal{Z}_k^n$, the simulation $(T_{\lambda_k}, S_{\lambda_k})$ is successful, then let $B_k = S_{\lambda_k}(\mathcal{Z}_k^n * y)$ for some (and hence all) $y \in \{0,1\}^{\mathbb{N}}$ and say that B_k **was computed by Step $(3, n)$**. For a pair (j, m), with $j \in \mathbb{Z}$, $V_j > 0$ and

$1 \leq m \leq V_j$, denote

$$\text{NEXT}(j,m) = \begin{cases} (j, m+1) & \text{if } m < V_j, \\ (\min\{j' > j : V_{j'} > 0\}, 1) & \text{if } m = V_j, \end{cases}$$

the **next bit position after (j, m)** (which is a random variable).

Step (3,0): For all $k \in \mathbb{Z}$, set $\mathcal{G}_k^0 = \emptyset$ (the empty set), $\mathcal{Z}_k^0 = \phi$ (the empty string). Set $J_k^0 = \min\{j \geq k : V_j > 0\}$ and $M_k^0 = 1$. (It follows easily from Lemmas 6 and 7 below that J_k^0 are a.s. defined and finite.) No B_k's are computed.

Step (3,n): For each $k \in \mathbb{Z}$, if B_k was computed by time $n-1$, set $(J_k^n, M_k^n) = (J_k^{n-1}, M_k^{n-1})$, $\mathcal{G}_k^n = \mathcal{G}_k^{n-1}$ and $\mathcal{Z}_k^n = \mathcal{Z}_k^{n-1}$. Otherwise, check if position (J_k^{n-1}, M_k^{n-1}) was used by *some* simulator by Step $(3, n-1)$, i.e. whether (J_k^{n-1}, M_k^{n-1}) is in $\cup_{k' \in \mathbb{Z}} \mathcal{G}_{k'}^{n-1}$. If yes: set $\mathcal{G}_k^n = \mathcal{G}_k^{n-1}$, $\mathcal{Z}_k^n = \mathcal{Z}_k^{n-1}$, and $(J_k^n, M_k^n) = \text{NEXT}(J_k^{n-1}, M_k^{n-1})$.

If position (J_k^{n-1}, M_k^{n-1}) was not used by any simulator by Step $(3, n-1)$: check if for some $k' > k$ we have $(J_k^{n-1}, M_k^{n-1}) = (J_{k'}^{n-1}, M_{k'}^{n-1})$, a phenomenon we refer to as a **queue-up** of the k-th simulator. If there is a queue-up, set $\mathcal{G}_k^n = \mathcal{G}_k^{n-1}$, $\mathcal{Z}_k^n = \mathcal{Z}_k^{n-1}$, and $(J_k^n, M_k^n) = \text{NEXT}(J_k^{n-1}, M_k^{n-1})$. If the k-th simulator is not queued up: Set $\mathcal{G}_k^n = \mathcal{G}_k^{n-1} \cup \{(J_k^{n-1}, M_k^{n-1})\}$ and $\mathcal{Z}_k^n = \mathcal{Z}_k^{n-1} * \epsilon_{J_k^{n-1}}(M_k^{n-1})$. If now the simulation $(T_{\lambda_k}, S_{\lambda_k})$is successful for input $\mathcal{Z}_k^n$, set $B_k = S_{\lambda_k}(\mathcal{Z}_k^n * y)$ for some (and hence all) $y \in \{0,1\}^{\mathbb{N}}$, set $(J_k^n, M_k^n) = (J_k^{n-1}, M_k^{n-1})$, and say that B_k was computed at Step $(3, n)$. Otherwise, set $(J_k^n, M_k^n) = \text{NEXT}(J_k^{n-1}, M_k^{n-1})$.

We will show later that, if the marker length t is chosen large enough, then for $\mathbf{P}$-almost every input sequence $(x_i)_{i \in \mathbb{Z}}$, for all $k \in \mathbb{Z}$ there exists an $n \geq 1$ for which B_k was computed by Step $(3, n)$. So all the B_k's are a.s. defined $\mathbf{B}^*$-valued random variables.

Note that it is immediate from the definition that $\text{length}(B_k) = \lambda_k = R_{k+1} - R_k$. Let $B_k = (\beta_k(1), \beta_k(2), \ldots, \beta_k(\lambda_k))$ be the $\mathbf{B}$-symbols comprising B_k. For each $i \in \mathbb{Z}$, we define $(\varphi(x))_i$ as follows. Let $K(i) \in \mathbb{Z}$ be the index of the block containing i, namely the unique $k \in \mathbb{Z}$ for which $R_k < i \leq R_{k+1}$, and set

$$(\varphi(x))_i = \beta_{K(i)}(i - R_{K(i)}).$$

This completes the formal definition of φ. Figure 1 shows a schematic illustration of the construction. Table 1 summarizes our main notation for convenient reference.

Table 1: Summary of notation

Symbol	Meaning
$\mathbf{A}^*$	finite words over $\mathbf{A}$
(T_ℓ, S_ℓ)	simulation of q^ℓ from independent unbiased bits
$(N_{n,t}, F_{n,t}, G_{n,t})$	simulation of independent unbiased bits from pattern-avoiding block
π	"forbidden pattern" $(2, 1, 1, \dots, 1)$ $(t-1$ ones)
$E_{n,t}$	$\mathbf{A}$-n-tuples not containing π
$x = (x_i)_{i\in\mathbb{Z}}$	input sequence
$\mathbf{A} = \{j : 1 \le j \le a\}$	source alphabet
$\mathbf{B} = \{j : 1 \le j \le b\}$	target alphabet
$p = (p(j))_{1\le i\le a}$	source distribution
$q = (q(j))_{1\le j\le b}$	target distribution
ε	lower bound on the entropy gap $h(p) - h(q)$
$\tilde{p}_{n,t}$	p^n conditioned on $E_{n,t}$
t	marker length
R_k	k-th marker
$\lambda_k = R_{k+1} - R_k$	length of k-th block
$W_k = (x_i)_{R_k+t\le i\le R_{k+1}-1}$	k-th input word
$L_k = \lambda_k - t$	length of W_k
$U_k = (\epsilon_k(1), \dots, \epsilon_k(V_k))$	bit string associated with block k
V_k	length of U_k
(J_k^n, M_k^n)	position of the k-th simulator at Step $(3, n)$
$\mathcal{Z}_k^n$	input read by the k-th simulator by Step $(3, n)$
$\mathcal{G}_k^n$	positions used by the k-th simulator by Step $(3, n)$
$\mathrm{NEXT}(j, m)$	next bit position after (j, m)
$B_k = (\beta_k(1), \dots, \beta_k(\lambda_k))$	$\mathbf{B}^*$-valued r.v. computed by the k-th simulator
$K(i)$	index of block containing i
$\varphi(x) = (\varphi(x)_i)_{i\in\mathbb{Z}}$	output sequence

4 Proof of Theorem 1

Lemma 5 *The $\mathbf{A}^*$-valued random variables $(W_k)_{k\in\mathbb{Z}}$ are independent. For each $k \in \mathbb{Z}$ and $n \in \{0, 1, 2, \ldots\}$, W_k conditioned on $\{length(W_k) = n\}$ has distribution $\tilde{p}_{n,t}$. The non-central block lengths $(R_{k+1} - R_k)_{k\in\mathbb{Z}\setminus\{0\}}$ are identically distributed with exponential tails, and the central block length $R_1 - R_0$ has exponential tails.*

Proof. Let μ_0, μ_1 be the measures on $\mathbf{A}^*$ defined as follows:

$$\begin{aligned}
&\mu_0\big(\{(a_1, a_2, \ldots, a_n)\}\big)\\
&\quad = \begin{cases} (n+t)p(2)^2 p(1)^{2(t-1)} \prod_{j=1}^n p(a_i), & (a_1, \ldots, a_n) \in E_{n,t},\\ 0 & \text{otherwise;}\end{cases}\\
&\mu_1\big(\{(a_1, a_2, \ldots, a_n)\}\big)\\
&\quad = \begin{cases} p(2)p(1)^{t-1} \prod_{j=1}^n p(a_i), & (a_1, \ldots, a_n) \in E_{n,t},\\ 0 & \text{otherwise}\end{cases}
\end{aligned}$$

We claim that for any $j \geq 0$ and $(w_k)_{-j\leq k\leq j} \subset \mathbf{A}^*$,

$$\mathbf{P}\bigg((W_k)_{-j\leq k\leq j} = (w_k)_{-j\leq k\leq j}\bigg) = \mu_0(w_0) \prod_{-j\leq k\leq j,\ k\neq 0} \mu_1(w_k). \tag{2}$$

This will prove that $(W_k)_{k\in\mathbb{Z}}$ are independent, with W_0 having distribution μ_0 and all the other W_k's having distribution μ_1. Furthermore both μ_0 and μ_1 clearly have the desired property that conditioning on the length n gives $\tilde{p}_{n,t}$. Indeed, to prove (2), let $\pi = 211\ldots1$ be the word "2 followed by $t-1$ 1's", and let w be the word obtained by the concatenation

$$w = \pi * w_{-j} * \pi * w_{-j+1} * \ldots * \pi * w_j * \pi.$$

Denote $\mathrm{len}^+ = \sum_{1\leq k\leq j} \mathrm{length}(w_k)$, $\mathrm{len}^- = \sum_{-j\leq k\leq 0} \mathrm{length}(w_k)$. Then, if all the w_k's do not contain the pattern π, we have

$$\begin{aligned}
&\Big\{(W_k)_{-j\leq k\leq j} = (w_k)_{-j\leq k\leq j}\Big\}\\
&\qquad = \bigcup_{r=0}^{\mathrm{length}(w_0)+t-1} \Big\{(x_i)_{i=r-(j+1)t-\mathrm{len}^-}^{r+(j+1)t+\mathrm{len}^+} = w\Big\},
\end{aligned} \tag{3}$$

and furthermore the above union is disjoint. (In words, this simply means that $(W_k)_{-j\le k\le j} = (w_k)_{-j\le k\le j}$ if and only if for some r, a string of x_i's centered around the origin such that the offset of the 0-th block relative to the origin is equal to r, is equal to w. This follows directly from the definitions.) Therefore

$$\mathbf{P}\Big((W_k)_{-j\le k\le j} = (w_k)_{-j\le k\le j}\Big) = \sum_{r=0}^{\mathrm{length}(w_0)+t-1} \mathbf{P}\Big((x_i)_{i=r-(j+1)t-\mathrm{len}^-}^{r+(j+1)t+\mathrm{len}^+} = w\Big)$$
$$= (\mathrm{length}(w_0)+t)\cdot p^{\mathrm{length}(w)}(w) = \mu_0(w_0) \prod_{-j\le k\le j,\ k\ne 0} \mu_1(w_k).$$

If some w_k contains π, the event on the left hand side of (3) is empty, and (2) holds trivially.

It remains to prove that the block lengths $R_{k+1}-R_k$ have exponential tails, or equivalently to prove the same for $L_k = \mathrm{length}(W_k) = R_{k+1}-R_k-t$. Denote $c = p^t(\pi) = p(2)p(1)^{t-1}$. Then for $k\ne 0$,

$$\mathbf{P}\Big(L_k = n\Big) = \sum_{w\in E_{n,t}} \mu_1(w) = c\sum_{w\in E_{n,t}} p^n(w)$$
$$= c\,\mathbf{P}\Big((x_i)_{i=0}^{n-1} \text{ does not contain } \pi\Big)$$
$$\le c\,\mathbf{P}\Big((x_i)_{i=jt}^{jt+t-1} \ne \pi,\quad j=1,2,\ldots,\left\lfloor\frac{n-t}{t}\right\rfloor\Big) = c(1-c)^{\lfloor (n-t)/t\rfloor},$$

which decays exponentially in n. Similarly, since the distribution μ_0 is at most a factor $O(n)$ times μ_1, $L_0 = \mathrm{length}(W_0)$ also has exponential tails. $\square$

Lemma 6 *The associated bit strings $(U_k)_{k\in\mathbb{Z}}$ are independent $\{0,1\}^*$-valued random variables. For each $k\in\mathbb{Z}$ and $n\in\{0,1,2,\ldots\}$, U_k conditioned on $\{V_k = n\}$ has the uniform distribution on $\{0,1\}^n$. $(U_k)_{k\in\mathbb{Z}\setminus\{0\}}$ are identically distributed. $(V_k)_{k\in\mathbb{Z}}$ have exponential tails.*

Proof. It is immediate from Lemma 5 that $(W_k)_{k\in\mathbb{Z}}$ are independent and $(W_k)_{k\in\mathbb{Z}\setminus\{0\}}$ are identically distributed, therefore the same holds for the sequence (U_k), as required. Now, for any $k\in\mathbb{Z}$, $V_k = N_{L_k,t}(W_k) \le (\log a/\log 2)L_k$ by Theorem 4(iii), so it has exponential tails by Lemma 5.

Finally, let $k \in \mathbb{Z}$ and $n \in \{0, 1, 2, \ldots\}$. Then by Theorem 4 and Lemma 5, for any $w \in \{0, 1\}^n$,

$$
\begin{aligned}
&\mathbf{P}\Big(U_k = w \;\Big|\; V_k = n\Big) \\
&= \sum_m \mathbf{P}\Big(L_k = m \;\Big|\; V_k = n\Big)\mathbf{P}\Big(U_k = w \;\Big|\; V_k = n,\ L_k = m\Big) \\
&= \sum_m \mathbf{P}\Big(L_k = m \;\Big|\; V_k = n\Big) \\
&\qquad\qquad \times \mathbf{P}\Big(F_{m,t}(W_k) = w \;\Big|\; \text{length}(W_k) = m,\ N_{m,t}(W_k) = n\Big) \\
&= \sum_m \mathbf{P}\Big(L_k = m \;\Big|\; V_k = n\Big) \cdot 2^{-n} = 2^{-n}.
\end{aligned}
$$

□

Lemma 7 *The marker length t may be chosen so that for $k \neq 0$,*

$$\mathbf{E}(V_k) > \mathbf{E}(T_{\lambda_1}). \tag{4}$$

Proof. Consider temporarily a new probability measure $\tilde{\mathbf{P}}$, with expectation operator $\tilde{\mathbf{E}}$, under which W_0 has the same distribution as W_1 and is independent of all other random variables, while all other random variables have the same distribution as before. By Lemma 5 and the computation in its proof, the process $(W_k)_{k \in \mathbb{Z}}$ is now isomorphic to the induced dynamical system $B(p)\big|_{\mathcal{M}}$, where

$$\mathcal{M} = \{x \in \mathbf{A}^{\mathbb{Z}} : (x_i)_{i=0}^{t-1} = \pi\}.$$

(Recall that $\pi = (2, 1, ..., 1)$ is the forbidden pattern.) By Abramov's formula ([15], p. 257–259), this dynamical system has entropy $h(W_1) = h(p)/p^t(\pi)$. This is also equal to $h(p)\tilde{\mathbf{E}}(\lambda_1) = h(p)\mathbf{E}(\lambda_1)$, since by Kac's formula ([15], p. 46), the expected return time (or expected block length) is the reciprocal of the probability of the inducing set.

By Theorem 4(ii), the mapping $W_1 \mapsto (U_1, L_1, G_{L_1,t}(W_1))$ is injective. Therefore, using Lemma 5 and elementary properties of the entropy function,

$$
\begin{aligned}
h(p)\mathbf{E}(\lambda_1) &= h(W_1) = h(U_1, L_1, G_{L_1,t}(W_1)) \leq h(U_1) + h(L_1, G_{L_1,t}(W_1)) \\
&= h(U_1|V_1) + h(V_1) + h(L_1, G_{L_1,t}(W_1)) \\
&= \mathbf{E}(V_1)\log 2 + h(V_1) + h(L_1, G_{L_1,t}(W_1)).
\end{aligned}
$$

By Theorem 3, we have

$$\mathbf{E}(T_{\lambda_1})\log 2 \le 6\log 2 + \mathbf{E}(\lambda_1)h(q) \le 6\log 2 + \mathbf{E}(\lambda_1)(h(p)-\varepsilon).$$

Combining the last two inequalities gives that

$$\mathbf{E}(V_1) - \mathbf{E}(T_{\lambda_1}) \ge \frac{\varepsilon}{\log 2}\mathbf{E}(\lambda_1) - 6 - \frac{h(V_1) + h(L_1, G_{L_1,t}(W_1))}{\log 2}.$$

To bound the negative terms on the right-hand side, recall the following properties of the entropy of integer-valued random variables (see [1], Lemma 12.10.2): If X is a random variable with finite expectation that takes values in $\mathbb{N}$, then $h(X) \le h(Y)$, where Y has a geometric distribution with $\mathbf{E}(Y) = \mathbf{E}(X)$. Furthermore, $h(Y) = O(\log \mathbf{E}(Y))$ when the expectation is large. This implies, using the fact that $V_1 \le (\log a/\log 2)L_1$, that for some positive constant C (that depends on a),

$$\begin{aligned}
h(L_1) &\le C\log \mathbf{E}(\lambda_1),\\
h(V_1) &\le C\log \mathbf{E}(\lambda_1),\\
h(L_1, G_{L_1,t}(W_1)) &= h(L_1) + h(G_{L_1,t}(W_1) \mid L_1)\\
&\le C\log \mathbf{E}(\lambda_1) + (a-1)\log \mathbf{E}(\lambda_1),
\end{aligned}$$

(because the range of $G_{n,t}$ is $\{1, 2, \dots, n^{a-1}\}$). Therefore

$$\mathbf{E}(V_1) - \mathbf{E}(T_{\lambda_1}) \ge f(\mathbf{E}(\lambda_1))$$

for some function

$$f(u) = f_\varepsilon(u) = \frac{\varepsilon}{\log 2}u - 6 - C'\log u,$$

where $C' > 0$ is a constant that depends only on a. Now, $f(u) \to \infty$ as $u \to \infty$. Choosing the marker length t sufficiently high will force $\mathbf{E}(\lambda_1) = 1/p^t(\pi)$ to be large, uniformly over all probability vectors p under consideration, i.e. those that satisfy $h(p) \ge h(q) + \varepsilon$ and $p(i) > 0$ for all $i \in \mathbf{A}$. (Note that $p(1)$ is bounded away from 1 because the alphabet size is fixed and $h(p)$ is bounded away from 0). So for sufficiently large t we get $\mathbf{E}(V_1) - \mathbf{E}(T_{\lambda_1}) > 0$, proving the lemma. □

From now on we consider t as having a fixed value for which (4) holds. Note that in particular it follows from Lemma 7 that almost surely, $V_k > 0$

for infinitely many positive values of k. Therefore (J_k^n, M_k^n) in Steps $(3,0)$, $(3,n)$ are a.s. defined.

Recall the notion of **stochastic domination**. Let Λ be a compact metric space on which there is defined a partial order "$\preceq$", and assume that $\preceq$ is a closed subset of $\Lambda \times \Lambda$. If X, Y are two random variables (not necessarily defined on the same probability space) taking values in Λ, denote $X \preceq_{\text{stoc}} Y$ (read: **"X is stochastically dominated by Y"**) if for any $e \in \Lambda$ we have

$$\mathbf{P}(X \succeq e) \leq \mathbf{P}(Y \succeq e).$$

It is known (see [11, Th. 2.4, p. 71]) that, under the above topological assumptions, $X \preceq_{\text{stoc}} Y$ if and only if there exist random variables X', Y', defined *on the same probability space*, such that the variable X' has the same distribution as X, the variable Y' has the same distribution as Y, and

$$\mathbf{P}(X' \preceq Y') = 1.$$

On the set $\{0,1\}^{\#} = \{0,1\}^* \cup \{0,1\}^{\mathbb{N}}$ of all finite and infinite bit strings, let $w \preceq w'$ denote the order "w is a prefix of w'". Equip $\{0,1\}^{\#}$ with the topology consisting of open sets of the form $\left\{w' \in \{0,1\}^{\#} : w \preceq w'\right\}$ for $w \in \{0,1\}^*$. It is not difficult to verify that $\{0,1\}^{\#}$ with this topology is a compact metric space, and that $\preceq$ is a closed subset of $\{0,1\}^{\#} \times \{0,1\}^{\#}$.

On the set $(\{0,1\}^{\#})^{\mathbb{Z}}$ of $\mathbb{Z}$-indexed vectors of finite and infinite bit strings, let $\preceq^{\mathbb{Z}}$ denote the (strong) product order of $\preceq$, i.e., we define

$$(w_k)_{k\in\mathbb{Z}} \preceq^{\mathbb{Z}} (w'_k)_{k\in\mathbb{Z}} \quad \Longleftrightarrow \quad w_k \preceq w'_k \text{ for all } k \in \mathbb{Z}.$$

For each $\ell \in \mathbb{N}$, let Z_ℓ be the following $\{0,1\}^*$-valued random variable: take independent unbiased bits $\nu_1, \nu_2, \nu_3, \ldots$, and set

$$Z_\ell = (\nu_1, \nu_2, \ldots, \nu_{T_\ell(\nu_1,\nu_2,\ldots)}),$$

where T_ℓ is the stopping time of the simulation of q^ℓ from independent unbiased bits as in Section 3. We call Z_ℓ (an instance of) **an acceptable input for the simulation (T_ℓ, S_ℓ)**. Let $(Z_{\ell,k})_{\ell\in\mathbb{N},k\in\mathbb{Z}}$ be an infinite array of independent random variables, where $Z_{\ell,k}$ has the same distribution as Z_ℓ. Denote $\tau_{\ell,k} = \text{length}(Z_{\ell,k})$. Clearly $\tau_{\ell,k}$ is equal in law to $T_\ell(\nu_1, \nu_2, \ldots)$.

Lemma 8 *For each $n = 0, 1, 2, \ldots$ we have* *(i) $\mathcal{Z}_k^n$ is the concatenation of the bits $\epsilon_j(m)$ for all $(j,m) \in \mathcal{G}_k^n$, arranged by lexicographical order on (j,m).*

(ii) $(\mathcal{G}_k^n)_{k\in\mathbb{Z}}$ are disjoint sets.

(iii) The conditional distribution of $(\mathcal{Z}_k^n)_{k\in\mathbb{Z}}$ given $(R_k)_{k\in\mathbb{Z}} = (r_k)_{k\in\mathbb{Z}}$ is a.s. stochastically dominated, in the order $\preceq^{\mathbb{Z}}$, by $(Z_{r_{k+1}-r_k,k})_{k\in\mathbb{Z}}$.

(iv) The conditional distribution of $(length(\mathcal{Z}_k^n))_{k\in\mathbb{Z}}$ given $(R_k)_{k\in\mathbb{Z}} = (r_k)_{k\in\mathbb{Z}}$ is a.s. stochastically dominated, in the order $\leq^{\mathbb{Z}}$ (the product order of the usual order on numbers), by $(\tau_{r_{k+1}-r_k,k})_{k\in\mathbb{Z}}$.

PROOF. Claim (i) follows trivially by induction on n, as the simulator locations (J_k^n, M_k^n) are obviously increasing in the lexicographical order.

Claim (ii) follows by induction on n, by noting that $\mathcal{G}_k^n$ is always obtained from $\mathcal{G}_k^{n-1}$ by the addition of at most one location (j,m), and, since we made allowance for the phenomenon of queue-ups, where multiple simulators are at the same location during Step $(3,n)$, any given location (j,m) is added to $\mathcal{G}_k^n$ for at most one value of $k \in \mathbb{Z}$, keeping the $\mathcal{G}_k^n$'s disjoint.

It remains to prove Claim (iii), which implies (iv) trivially. To do this, let $(\theta_{k,j})_{k\in\mathbb{Z},j\in\mathbb{N}}$ be an array of independent unbiased bits which are independent of all other random variables. For each $k \in \mathbb{Z}$ define $\mathcal{X}_k = \mathcal{Z}_k^n * (\theta_{k,j})_{j\geq 1}$. Because of Lemma 6 together with claims (i) and (ii) proven above, it follows that, conditional on $(R_k)_{k\in\mathbb{Z}} = (r_k)_{k\in\mathbb{Z}}$, we have that $(\mathcal{X}_k)_{k\in\mathbb{Z}}$ is a sequence of independent infinite sequences of independent unbiased bits. Set $\xi_k = (\mathcal{X}_{k,j})_{1\leq j\leq T_{r_{k+1}-r_k}(\mathcal{X}_k)}$. Then (still working conditionally) $(\xi_k)_{k\in\mathbb{Z}}$ are independent and for each $k \in \mathbb{Z}$, ξ_k has the distribution of $Z_{r_{k+1}-r_k}$. Also, from the construction necessarily $\mathcal{Z}_k^n \preceq \xi_k$. We have constructed a realization of $(Z_{r_{k+1}-r_k,k})_{k\in\mathbb{Z}}$ that dominates $(\mathcal{Z}_k^n)_{k\in\mathbb{Z}}$, thereby proving the stochastic domination claim. $\square$

We shall use the following **mass-transport lemma**:

Lemma 9 *Let $f : \mathbb{Z}\times\mathbb{Z} \to \mathbb{R}$ satisfy $f(x+c, y+c) = f(x,y)$ for all $x, y, c \in \mathbb{Z}$. Then for all $x \in \mathbb{Z}$,*

$$\sum_{y\in\mathbb{Z}} f(x,y) = \sum_{y\in\mathbb{Z}} f(y,x).$$

PROOF. Taking $c = x - y$ gives $f(x,y) = f(2x-y, x)$, so

$$\sum_{y\in\mathbb{Z}} f(x,y) = \sum_{y\in\mathbb{Z}} f(2x-y,x) = \sum_{u\in\mathbb{Z}} f(u,x).$$

□

Lemma 10 *Almost surely, for all $k \in \mathbb{Z}$ there exists an $n \geq 1$ for which B_k was computed by Step $(3, n)$.*

PROOF. As in the proof of Lemma 7, we introduce the probability measure $\tilde{\mathbf{P}}$ with expectation operator $\tilde{\mathbf{E}}$ under which $(W_k)_{k\in\mathbb{Z}}$ are i.i.d. Since $\mathbf{P}$ is absolutely continuous with respect to $\tilde{\mathbf{P}}$ (see the proof of Lemma 5), it is enough to prove that each B_k is eventually computed, $\tilde{\mathbf{P}}$-a.s.

For each $k \in \mathbb{Z}$, define $\mathcal{Z}_k^\infty = \lim_{n\to\infty} \mathcal{Z}_k^n$ (a possibly infinite bit sequence) and $\mathcal{G}_k^\infty = \lim_{n\to\infty} \mathcal{G}_k^n = \cup_{n=1}^\infty \mathcal{G}_k^n$. Define $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ by

$$\begin{aligned} f(k,j) &= \tilde{\mathbf{E}}\Big(\big| \big\{ (j,i) : i \in \mathbb{Z},\quad 1 \leq i \leq V_j \big\} \cap \mathcal{G}_k^\infty \big| \Big) \\ &= \tilde{\mathbf{E}}\big(\text{number of bits from } U_j \text{ read by the } k\text{-th simulator}\big). \end{aligned}$$

The function f satisfies the assumption of Lemma 9, since $(W_k)_{k\in\mathbb{Z}}$ is a stationary sequence under $\tilde{\mathbf{P}}$, and it is easy to see from the construction that $(U_k, V_k, \mathcal{Z}_k^\infty, \mathcal{G}_k^\infty)_{k\in\mathbb{Z}}$ are all generated from $(W_k)_{k\in\mathbb{Z}}$ in a shift-equivariant manner. Therefore for all $k \in \mathbb{Z}$, we have

$$\sum_{j\in\mathbb{Z}} f(k,j) = \sum_{j\in\mathbb{Z}} f(j,k). \tag{5}$$

Denote the quantity in (5) by g (by stationarity it does not depend on k). The left-hand side of (5) is equal to $\tilde{\mathbf{E}}(|\mathcal{G}_k^\infty|)$, the expected number of bits eventually read by the k-th simulator. By Lemma 8(iv), $g \leq \mathbf{E}(T_{\lambda_1})$. The right-hand side is equal to the total number of bits from the k-th associated bit string U_k eventually used by *any* simulator, and clearly cannot exceed $\tilde{\mathbf{E}}(V_k) = \mathbf{E}(V_1)$. In particular, this implies that $g < \infty$, so almost surely $\mathcal{Z}_k^\infty$ is a finite string.

Assume that with positive $\tilde{\mathbf{P}}$-probability, B_k is not computed by Step $(3, n)$ for any n. Since $\mathcal{Z}_k^\infty$ is finite, the only way for this to happen is for the k-th simulator to eventually fail to find any unused bits in the blocks to its right. Because of stationarity, by the ergodic theorem this implies that a.s., this happened to a positive proportion of the simulators. Therefore, a.s. *all* the bits are eventually used! In other words, there is the equality $g = \mathbf{E}(V_1)$. We have shown $\mathbf{E}(V_1) = g \leq \mathbf{E}(T_{\lambda_1})$, in contradiction to Lemma 7. So the assumption that B_k was not computed with positive probability is false. □

Lemma 11 *Conditioned on $(R_k)_{k\in\mathbb{Z}} = (r_k)_{k\in\mathbb{Z}}$, the random variable $(B_k)_{k\in\mathbb{Z}}$ are independent, and for each $k \in \mathbb{Z}$, B_k has distribution $q^{r_{k+1}-r_k}$.*

PROOF. By Lemma 8, conditioned on $(R_k)_{k\in\mathbb{Z}} = (r_k)_{k\in\mathbb{Z}}$, the random vector $(\mathcal{Z}_k^\infty)_{k\in\mathbb{Z}}$ is stochastically dominated in the product prefix order by $(Z_{r_{k+1}-r_k,k})_{k\in\mathbb{Z}}$. Assume that these two sequences are defined on the same space and that there is actual a.s. domination. For each $k \in \mathbb{Z}$, both $\mathcal{Z}_k^\infty$ and $Z_{r_{k+1}-r_k,k}$ are acceptable inputs for the simulation $(T_{r_{k+1}-r_k}, S_{r_{k+1}-r_k})$. So, since $\mathcal{Z}_k^\infty \preceq Z_{r_{k+1}-r_k,k}$, necessarily $\mathcal{Z}_k^\infty = Z_{r_{k+1}-r_k,k}$. We have shown that $(\mathcal{Z}_k^\infty)_{k\in\mathbb{Z}} = (Z_{r_{k+1}-r_k,k})_{k\in\mathbb{Z}}$. Therefore (still working conditionally) $(\mathcal{Z}_k^\infty)_{k\in\mathbb{Z}}$ are independent and each $\mathcal{Z}_k^\infty$ has the distribution of an acceptable input for the simulation $(T_{r_{k+1}-r_k}, S_{r_{k+1}-r_k})$. Therefore, since $B_k = S_{r_{k+1}-r_k}(\mathcal{Z}_k^\infty)$, $(B_k)_{k\in\mathbb{Z}}$ are independent, and for each $k \in \mathbb{Z}$, B_k has distribution $q^{r_{k+1}-r_k}$. □

Lemma 12 *$((\varphi(x))_i)_{i\in\mathbb{Z}}$ are i.i.d. with distribution q.*

PROOF. The mapping $i \to (K(i), i - R_{K(i)})$ is obviously injective. This means that the $(\varphi(x))_i = \beta_{K(i)}(i - R_{K(i)})$ get assigned different $\beta_k(j)$ symbols. Conditioned on $(R_k)_{k\in\mathbb{Z}}$, these symbols are all independent **B**-symbols with distribution q. This immediately implies the claim of the lemma. □

Lemma 13 *φ is translation-equivariant.*

PROOF. Let $x \in \mathbf{A}^{\mathbb{Z}}$. Denote $x' = \mathcal{T}(x)$. Our goal is to prove that for all $i \in \mathbb{Z}$, $(\varphi(x'))_i = (\varphi(x))_{i+1}$.

If X is any of the various quantities in Table 1 which are implicitly dependent on x, denote by X' the corresponding quantity taken as a function of x' rather than x. Consider separately two cases:

Case 1: $R_1 > 0$. In this case, it is easy to check directly that for all $k \in \mathbb{Z}$,

$$R'_k = R_k - 1, \quad W'_k = W_k, \quad L'_k = L_k, \quad U'_k = U_k, \quad V'_k = V_k$$

(the markers are shifted by one). By induction on n, for all $k \in \mathbb{Z}$ and all $n \geq 0$,

$$((J_k^n)', (M_k^n)') = (J_k^n, M_k^n), \quad (\mathcal{Z}_k^n)' = \mathcal{Z}_k^n, \quad (\mathcal{G}_k^n)' = \mathcal{G}_k^n.$$

Therefore, for all $k \in \mathbb{Z}$, $B'_k = B_k$. Furthermore, $K(i)' = K(i+1)$. Therefore

$$(\varphi(x'))_i = \beta'_{K(i)'}(i - R'_{K(i)'}) = \beta_{K(i+1)}(i - (R_{K(i)+1} - 1)) = (\varphi(x))_{i+1}.$$

Case 2: $R_1 = 0$. In this case, check that for all $k \in \mathbb{Z}$,

$$R'_k = R_{k+1} - 1, \quad W'_k = W_{k+1}, \quad L'_k = L_{k+1}, \quad U'_k = U_{k+1}, \quad V'_k = V_{k+1}$$

(the markers are shifted by 1, and the indexing of the blocks is shifted by 1). Therefore, by induction on n, for all $k \in \mathbb{Z}$ and for all $n \geq 0$,

$$((J_k^n)', (M_k^n)') = (J_{k+1}^n, M_{k+1}^n), \quad (\mathcal{Z}_k^n)' = \mathcal{Z}_{k+1}^n, \quad (\mathcal{G}_k^n)' = \mathcal{G}_{k+1}^n.$$

Therefore, for all $k \in \mathbb{Z}$, $B'_k = B_{k+1}$. Check as before that now $K(i)' = K(i+1) - 1$. Therefore

$$\begin{aligned}(\varphi(x'))_i &= \beta'_{K(i)'}(i - R'_{K(i)'}) \\ &= \beta_{(K(i+1)-1)+1}(i - (R_{(K(i+1)-1)+1} - 1)) = (\varphi(x))_{i+1}.\end{aligned}$$

□

The following facts concerning random variables with exponential tails will be useful.

Lemma 14

(i) If X, Y are real-valued random variables with exponential tails, then $X + Y$ has exponential tails.

(ii) If $X_1, X_2, \ldots$ are random variables with uniformly exponential tails (i.e., $(\sup_k \mathbf{P}(|X_k| \geq n))_{n \in \mathbb{N}}$ decays exponentially), and T is an $\mathbb{N}$-valued random variable with exponential tails, then the random variable X_T has exponential tails.

(iii) If $X_1, X_2, \ldots$ are i.i.d. random variables with exponential tails and mean μ, then, for any $c > 0$, the sequence

$$\left(\mathbf{P} \left(\left| \sum_{i=1}^{n} X_i - n\mu \right| \geq nc \right) \right)_{n \in \mathbb{Z}_+}$$

decays exponentially.

(iv) Suppose $X_1, X_2, \ldots$ are i.i.d. random variables with exponential tails, and T is an $\mathbb{N}$-valued random variable with exponential tails. Denote $S_m = \sum_{k=1}^{m} X_k$. Then $S_T := \sum_{k=1}^{T} X_k$ has exponential tails.

Proof. Proof of (i): we have $\mathbf{P}(|X+Y|\geq n)\leq \mathbf{P}(|X|\geq n/2)+\mathbf{P}(|Y|\geq n/2)$, which decays exponentially.

Proof of (ii): we have

$$\begin{aligned}\mathbf{P}(|X_T|\geq n) &\leq \mathbf{P}(T\geq n)+\mathbf{P}\left(\bigcup_{k=1}^{n}\{|X_k|\geq n\}\right)\\ &\leq \mathbf{P}(T\geq n)+n\sup_{1\leq k\leq n}\mathbf{P}(|X_k|\geq n),\end{aligned}$$

which decays exponentially.

Claim (iii) is a standard fact from large deviation theory; see for example ([8] Corollary 27.4).

Proof of (iv): Let $c>0$ and $0<d<1$ be such that $\mathbf{P}(T\geq n)\leq c\cdot d^n$ for all n. Denote $a=1/(2\mathbf{E}|X_1|)$. Then

$$\begin{aligned}\mathbf{P}(|S_T|\geq n) &= \sum_{m=1}^{\infty}\mathbf{P}\big(T=m,\ |S_m|\geq n\big)\leq \sum_{m=1}^{\infty}\mathbf{P}\Big(T=m,\sum_{k=1}^{m}|X_k|\geq n\Big)\\ &\leq \sum_{m=1}^{\lfloor a\cdot n\rfloor}\mathbf{P}\Big(T=m,\sum_{k=1}^{\lfloor a\cdot n\rfloor}|X_k|\geq n\Big)+\sum_{m=\lfloor a\cdot n\rfloor+1}^{\infty}\mathbf{P}\Big(T=m\Big)\\ &\leq en\cdot\mathbf{P}\Big(\Big|\sum_{k=1}^{\lfloor a\cdot n\rfloor}|X_k|-\lfloor a\cdot n\rfloor\mathbf{E}|X_1|\Big|>\frac{n}{2}\Big)+\mathbf{P}\Big(T>an\Big)\end{aligned}$$

In the last bound, the second term decays exponentially in n. The first term decays exponentially, by (iii) above. □

For each $k\in\mathbb{Z}$, let $J_k^\infty=\lim_{n\to\infty}J_k^n$ be the value of J_k^n for that n for which B_k was computed at Step $(3,n)$; that is, the index of the rightmost block used by simulator k.

Lemma 15 *J_0^∞ has exponential tails.*

Proof. Let τ be that $n>0$ for which B_0 was computed at Step $(3,n)$. If we define $I_0=0$,

$$I_1=\min\{j\geq 0: V_j>0\},$$

and inductively for $k>1$,

$$I_k=\min\{j>I_{k-1}: V_j>0\},$$

then because of Lemma 6, it is easy to prove in a manner similar to the proof of Lemma 5 that $(I_{k+1}-I_k)_{k\geq 0}$ are independent, $(I_{k+1}-I_k)_{k\geq 1}$ are identically distributed and have the geometric distribution $\mathrm{Geom}(p_0)$, where $p_0 = \mathbf{P}(V_1 > 0)$ (so in particular have exponential tails), and I_1 is stochastically dominated by $(I_2 - I_1) + 1$.

We know that

$$J_0^\infty \leq I_\tau = \sum_{j=1}^{\tau}(I_j - I_{j-1}).$$

This is because $J_0^0 = I_1$, and at any Step $(3,n)$, $1 \leq n < \tau$, we have $(J_0^n, M_0^n) = \mathrm{NEXT}(J_0^{n-1}, M_0^{n-1})$, so by induction, $J_0^n \leq I_{n+1}$. In particular $J_0^\infty = J_0^\tau = J_0^{\tau-1} \leq I_\tau$.

It follows, by Lemma 14(iv), that it is enough to prove that τ has exponential tails. Let $c>0, \delta>0$ be parameters whose value we will fix shortly. Write

$$\begin{aligned} \mathbf{P}\Big(\tau > n\Big) \quad &= \quad \mathbf{P}\Big(B_0 \text{ undefined by Step } (3,n)\Big) \\ = \quad \mathbf{P}\Big(\tau > n, \quad \mathrm{length}(\mathcal{Z}_0^n) < cn\Big) &+ \mathbf{P}\Big(\tau > n, \quad \mathrm{length}(\mathcal{Z}_0^n) \geq cn\Big). \end{aligned} \tag{6}$$

By Lemma 8(iv), the second term is at most $\mathbf{P}(T_{\lambda_1} \geq cn)$. Note that T_{λ_1} has exponential tails, since by 3(i), $\mathbf{P}(T_k \geq n) \leq \frac{b^k+1}{2^n}$, whence

$$\mathbf{P}(T_{\lambda_1} \geq n) \leq \mathbf{P}(\lambda_1 \geq dn) + \mathbf{P}\left(\bigcup_{k\leq dn}\{T_k \geq n\}\right) \leq \mathbf{P}(\lambda_1 \geq dn) + \frac{b^{dn}+1}{2^n}$$

can be seen to decay exponentially by taking $d = \log 2/(2\log b)$. It remains to deal with the first term in (6),

$$\mathbf{P}(E_n) := \mathbf{P}\Big(\tau > n, \quad \mathrm{length}(\mathcal{Z}_0^n) < cn\Big).$$

Note the event inclusion

$$E_n \subseteq \left\{\tau > n, \quad \sum_{k=1}^{J_0^n - 1} \mathrm{length}(\mathcal{Z}_k^n) \geq \sum_{k=1}^{J_0^n-1} V_k - cn\right\},$$

since on E_n, by Step $(3,n)$ simulator number 0 would have used all the bits $\epsilon_j(m)$ for $1 \leq j \leq J_0^n - 1$, $1 \leq m \leq V_j$, which were left unused by the simulators numbered $1, \ldots, J_0^n - 1$, and there cannot be more than cn such.

Next, observe that on the event $\{\tau > n\}$, because $(J_0^n, M_0^n) = \mathrm{NEXT}^n(J_0^0, M_0^0)$ (the n-th iteration of NEXT), we have the equality

$$J_0^n = \inf\left\{ j \geq 0 : \sum_{i=0}^{j} V_i > n \right\} =: \theta_n$$

(we denote the quantity on the right-hand side by θ_n). So we have shown

$$E_n \subseteq \left\{ \tau > n, \ \sum_{k=1}^{\theta_n - 1} \mathrm{length}(\mathcal{Z}_k^n) \geq \sum_{k=1}^{\theta_n - 1} V_k - cn \right\}.$$

The value of θ_n is, with probability exponentially close to 1, close to $(\mathbf{E}(V_1))^{-1}n$. More precisely, for any $\delta > 0$

$$\begin{aligned}\left\{ \theta_n > ((\mathbf{E}(V_1))^{-1} + \delta)n \right\} &\subseteq \left\{ \textstyle\sum_{0 \leq i \leq ((\mathbf{E}(V_1))^{-1}+\delta)n} V_i \leq n \right\} \\ &= \left\{ \textstyle\sum_{0 \leq i \leq ((\mathbf{E}(V_1))^{-1}+\delta)n} (V_i - \mathbf{E}(V_1)) \leq -\mathbf{E}(V_1) \cdot \delta n \right\}\end{aligned}$$

has probability which decays exponentially by Lemma 6 and Lemma 14(iii), and similarly, the probability of

$$\left\{ \theta_n < ((\mathbf{E}(V_1))^{-1} - \delta)n \right\} \subseteq \bigcup_{0 \leq j < ((\mathbf{E}(V_1))^{-1} - \delta)n} \left\{ \sum_{0=1}^{j} V_i > n \right\}$$

decays exponentially. So, summarizing the latest developments, we can now write

$$\begin{aligned} E_n \ \subseteq \ & \left\{ |\theta_n - (\mathbf{E}(V_1))^{-1} n| > \delta n \right\} \\ & \cup \bigcup_{|j - (\mathbf{E}(V_1))^{-1} n| \leq \delta n} \left\{ \sum_{k=1}^{j} \mathrm{length}(\mathcal{Z}_k^n) \geq \sum_{k=1}^{j} V_k - cn \right\}. \end{aligned}$$

The first event has probability which decays exponentially by the remarks above. The second event is a union over linearly many events, each of whose probability we can only increase by replacing length$(\mathcal{Z}_k^n)$ by T_k', an independent copy of T_{λ_1}, using Lemma 8(iv). So it is enough to prove that

$$\max_{|j-(\mathbf{E}(V_1))^{-1}n|\le\delta n}\mathbf{P}\left(\sum_{k=1}^{j}T_k'\ge\sum_{k=1}^{j}V_k-cn\right)$$

decays exponentially in n. For this, invoking Lemma 7, choose c and δ sufficiently small (the following choices will work: $\delta=(\mathbf{E}(V_1))^{-1}/2$, $c=(\mathbf{E}(V_1)-\mathbf{E}(T_{\lambda_1}))/(3(1+\mathbf{E}(V_1)))$ so that the inclusion

$$\begin{aligned}\left\{\sum_{k=1}^{j}T_k'\ge\sum_{k=1}^{j}V_k-cn\right\}\ \subseteq\ &\left\{\sum_{k=1}^{j}(T_k'-\mathbf{E}(T_{\lambda_1}))\ge cj\right\}\\ &\cup\left\{\sum_{k=1}^{j}(V_k-\mathbf{E}(V_1))\le -cj\right\},\qquad |j-(\mathbf{E}(V_1))^{-1}n|\le\delta n,\end{aligned}$$

will hold, and use Lemma 14(iii). □

Lemma 16 *φ is a finitary homomorphism from $B(p)$ to $B(q)$ with exponential tails.*

Proof. From Lemmas 12 and 13 it follows that φ is a homomorphism from $B(p)$ to $B(q)$. From the definition $(\varphi(x))_0=\beta_{K(0)}(-R_{K(0)})=\beta_0(-R_0)$ (since $K(0)=0$) we see that $(\varphi(x))_0$ is determined from the input symbols $(x_i)_{R_0\le i\le R_{J_0^\infty+1}+t}$. This proves that φ is finitary, with a coding length N_φ that satisfies

$$\begin{aligned}N_\varphi &\le \max(-R_0,R_{J_0^\infty+1}+t)\le R_{J_0^\infty+2}-R_0\\ &= \sum_{j=1}^{J_0^\infty+1}(R_{j+1}-R_j)+(R_1-R_0).\end{aligned}$$

Since by Lemma 5, $(R_{j+1}-R_j)_{j>0}$ are i.i.d. with exponential tails, and R_1-R_0 has exponential tails, it follows from Lemma 14(i),(iv) and Lemma 15 that N_φ has exponential tails. □

5 Extension to Markov chains

We indicate here the changes in the ideas presented above required to prove Theorem 2. We omit the details of the proofs, which are similar to those above.

5.1 Coding from a Markov source

If the Bernoulli source $B(p)$ is replaced by a Markov source, two changes to the construction are needed. First, the marker pattern "$211\ldots1$" must be replaced with a sequence which is assigned positive probability by the Markov chain. Here, we must give up the source-universality property, or make the rather restrictive assumption that all the entries in the matrix α are positive.

Second, it is necessary to replace the function $F_{n,t}$ that produces independent unbiased bits from a pattern-avoiding Bernoulli block with a new function, $F'_{n,t}$ designed to do the same for a pattern-avoiding Markov block conditioned to begin with a "1" (which we assume without loss of generality to be the last symbol in the left marker) and to end with a "2" (the first symbol in the right marker). Note that the construction of $F_{n,t}$ used only the symmetries of the distribution $\tilde{p}_{n,t}$, namely the fact that the space $E_{n,t}$ can be partitioned into classes of equiprobable elements, since the probability of an element only depends on the count of the different $\mathbf{A}$-symbols, and not on the order of their appearance.

For a Markov source, the probability of an element in $E_{n,t}$ will depend on the count of adjacent *pairs* of symbols. Thus, there will again be classes of equiprobable elements. The number of classes, which bounds the range of the complementary function $G'_{n,t}$ (and hence the amount of lost entropy – see the proof of Lemma 7), is at most n^{a^2}. All the proofs carry through identically to the Bernoulli case.

5.2 Coding to a Markov chain

Things get more complicated when the target process is Markov. Here, it is not enough for each block to independently generate the $\mathbf{B}$-symbols required to fill its spaces, since one must make sure that there are the correct transition probabilities on the boundaries between blocks. This problem may be solved as follows.

The first ingredient is a version of Theorem 3 for processes. That is, given a sequence of random variables $Y_1, Y_2, \ldots$, not necessarily independent, taking values in a finite alphabet $\mathbf{B}$, one may construct a sequence of simulations $(T_k, S_k)_{k\in\mathbb{N}}$, such that the stopping times T_k are increasing, and for each k, (T_k, S_k) is a simulation of the distribution of $(Y_1, Y_2, \ldots, Y_k)$ from independent unbiased bits. Each simulation is efficient, in the sense of Theorem 3(ii). The proof is obvious, by successively refining the partition $0 = Q_0 < Q_1 < \cdots < Q_b = 1$ of $(0, 1)$ used in the proof of Theorem 3.

Since the Markov matrix β is irreducible and aperiodic, there exists an $m \geq 1$ for which all the entries of β^m are positive. For each block $k \in \mathbb{Z}$, the k-th simulator will start by generating, using the independent unbiased bits U_k at her disposal (and, if necessary, bits from the blocks to her right), b separate Markov trajectories of length m, $(z_{k,i,j})_{1\leq i\leq b,\ 1\leq j\leq m}$, such that for each $1 \leq i \leq b$, $(z_{k,i,j})_{1\leq j\leq m}$ is a finite Markov chain that starts with the symbol i and has transition probabilities given by the matrix β. We call these finite chains the **preamble chains**. One of them will later be chosen to fill the first m places in the k-th block, but at the moment we don't know which.

Next, we want the k-th simulator to compute symbols to fill the remainder of her block. This can be done if in the sequences $(z_{k,i,j})$, **coupling** was achieved, i.e., if all the symbols $(z_{k,i,m})_{1\leq i\leq b}$ are identical; since in this case, no matter which of the preamble chains we later choose, we will need the same conditional distribution to compute the $(m+1)$-th symbol in the block, then the $(m+2)$-th, and so on.

Because of our choice of m, we know that coupling is achieved with a positive probability. So a positive proportion of the simulators will be able to continue and compute symbols to fill their blocks, using the nested simulations described above. But each simulator that filled her block also determined for the block to her right which of the preamble chains to use - the one that corresponds to the last symbol in the block that was filled. For each block for which the preamble was determined in this second round of computations, one may now proceed to simulate the remainder of the block. This then determines the preamble of more blocks, for which the block remainder is then computed. By iterating this process, the choice of preamble chains is propagated to the right until a single output sequence is determined.

The computation of the preamble chains uses a fixed amount of entropy per block. Since the blocks may be made arbitrarily large in expectation by choosing a long enough marker, the loss in entropy can be made negligible. The computation of the remainder of each block uses nested simulations, which are efficient. Therefore it can be shown that the total entropy loss is small, and the simulation will terminate a.s. The output sequence is clearly a stationary process, and by the construction its transition probabilities are exactly given by the transition matrix β. Therefore, it is the stationary Markov shift $\mathcal{M}(\beta)$. The construction gives a finitary homomorphism, and it can be shown to have exponential tails using large-deviations estimates similarly to the Bernoulli case.

Open Problems

(i) Do there exist source-universal finitary *isomorphisms*? More specifically, if the Bernoulli sources $B(p), B(p'), B(q)$ all have equal entropy, does there exist a finitary map (as defined in the introduction) which is simultaneously a finitary isomorphism from $B(p)$ to $B(q)$, and from $B(p')$ to $B(q)$?

Remark. If the function is not required to be a finitary map, the answer to the above question is positive, for a trivial reason. The function can simply use the law of large numbers to discern whether the input is in the almost-sure set of $B(p)$ or of $B(p')$, and apply one of two Keane-Smorodinsky [10] finitary isomorphisms accordingly.

(ii) Construct finitary isomorphisms between general Bernoulli sources with explicit bounds on the tails. (See [5],[7],[12] for such constructions in specific cases).

References

[1] T. M. Cover, J. A. Thomas (1991), *Elements of Information Theory.* John Wiley & Sons, Inc., New York.

[2] R. Durrett (1996), *Probability: Theory and Examples*, Second Edition. Duxbury Press, New York.

[3] P. Elias (1972), The efficient construction of an unbiased random sequence. *Ann. Math. Statist.* **43**, 865–870.

[4] T. S. Han, M. Hoshi (1997), Interval algorithm for random number generation. *IEEE Trans. Inform. Theory* **43**, 599–611.

[5] N. Harvey, Y. Peres (2003), An invariant of finitary codes with finite expected square root coding length. To appear in *Erg. Th. Dyn. Sys.*.

[6] A. Iwanik and J. Serafin (1999), Code length between Markov processes. *Israel J. of Math.* **11**, 29–51.

[7] S. Kalikow, B. Weiss (1992), Explicit codes for some infinite entropy Bernoulli shifts. *Ann. Probab.* **20**, 397–402.

[8] O. Kallenberg (2002), Foundations of Modern Probability, 2nd. ed. Springer, New York.

[9] M. Keane and M. Smorodinsky (1977), A class of finitary codes. *Israel J. of Math.* **26** (3-4), 352–371.

[10] M. Keane and M. Smorodinsky (1979), Bernoulli schemes of the same entropy are finitarily isomorphic. *Annals of Math.* **109**, 397–406.

[11] T. M. Liggett (1985), Interacting Particle Systems. Springer-Verlag, New York.

[12] L. D. Meshalkin (1959), A case of isomorphism of Bernoulli schemes. *Dokl. Akad. Nauk. SSSR* **128**, 41–44.

[13] D. E. Knuth and A. C. Yao (1976), The complexity of nonuniform random number generation. *Algorithms and complexity* (Proc. Sympos., Carnegie-Mellon Univ., Pittsburgh, Pa., 1976), 357–428. Academic Press, New York.

[14] Y. Peres (1992), Iterating von Neumann's procedure for extracting random bits. *Ann. Stat.* **20**, 590–597.

[15] K. Petersen (1983), *Ergodic Theory.* Cambridge University Press, Cambridge.

[16] D. Romik (1999), Sharp entropy bounds for discrete statistical simulation. *Statist. Probab. Letters* **42**, 219–227.

[17] J. Serafin (1996), The finitary coding of two Bernoulli schemes with unequal entropies has finite expectation. *Indag. Math.* (N.S.) **7**, no. 4, 503–519.

Nate Harvey
Department of Mathematics
UC Berkeley
CA 94720, USA

Alexander E. Holroyd
Department of Mathematics
University of British Columbia
Vancouver, BC V6T 1Z2, Canada
holroyd@math.ubc.ca

Yuval Peres
Departments of Statistics and Mathematics
UC Berkeley
CA 94720, USA
peres@stat.berkeley.edu

Dan Romik
The Mathematical Sciences Research Institute
17 Gauss Way
Berkeley, CA 94720-5070
USA
dromik@msri.org