On Independence and Capacity of Multidimensional Semiconstrained Systems

Ohad Elishco, *Student Member, IEEE*, Tom Meyerovitch, and Moshe Schwartz, *Senior Member, IEEE*

***Abstract*— We find a new formula for the limit of the capacity of certain sequences of multidimensional semiconstrained systems as the dimension tends to infinity. We do so by generalizing the notion of independence entropy, originally studied in the context of constrained systems, to the study of semiconstrained systems. Using the independence entropy, we obtain new lower bounds on the capacity of multidimensional semiconstrained systems in general, and d-dimensional axial-product systems in particular. In the case of the latter, we prove our bound is asymptotically tight, giving the exact limiting capacity in terms of the independence entropy. We show the new bound improves upon the best-known bound in a case study of $(0, k, p)$-RLL.**

***Index Terms*— Semiconstrained systems, capacity, independence entropy, bounds.**

I. Introduction

Error-correcting codes and constrained codes may be considered as two extreme ways of coping with a noisy channel. The former are usually data independent, and assume errors are a statistical phenomenon, reducing data-transmission rate to protect against such errors. Constrained codes, however, assume certain patterns in the data stream are responsible for the occurrence of errors. Thus, constrained codes eliminate all undesirable patterns, at the cost of reduced data-transmission rate.

Recently in [1] and [2], semiconstrained systems (SCSs) were suggested as a generalization to constrained systems (which we emphasize by calling fully constrained systems). In SCSs we do not eliminate the undesirable patterns entirely but rather we allow them to appear with a restriction on their frequency. To illustrate, consider a binary channel in which the appearance of k-consecutive 1's is forbidden. The set of allowed words is the well known inverted $(0, k)$-run-length-limited (RLL) system. However, if k-consecutive 1's are not forbidden entirely, but instead are allowed to appear in at most a fraction p of places, then the set of allowed words forms a SCS called the $(0, k, p)$-RLL system. Informally, a SCS is defined by a set Γ of probability measures over k-tuples. The allowed words in the SCS are those in which the empirical distribution of k-tuples belongs to Γ. This may be viewed as a generalization of fully constrained systems since taking Γ to be a subset with a 0-frequency restriction on some k-tuples yields a fully constrained system.

SCSs not only generalize fully constrained systems, but also subsume a range of other settings, which were mainly dealt with in an ad-hoc fashion. Among these we can find DC-free RLL coding [3], constant-weight ICI coding for flash memories [4]–[7], coding to mitigate the appearance of ghost pulses in optical communication [8], [9], and the more general, channel with cost constraints [10], [11].

In the one-dimensional case, the capacity of a SCS is given by a relatively explicit expression as the solution to a certain optimization problem on a finite dimensional space, e.g., [12]. A probabilistic encoder for SCSs was constructed in [1], and constant-bit-rate to constant-bit-rate encoders are possible by approximating a SCS with a fully constrained system, as described in [2].

A natural extension, and the goal of this work, is to study multidimensional SCSs. This is an extremely challenging problem, considering the fact that even for fully constrained systems in complete generality it is provably impossible to find an exact solution. The capacity of multidimensional fully constrained systems is known exactly only in a handful of cases [13]–[16]. In the absence of a general method for computing the capacity, various bounds and approximations were studied, e.g., [17]–[26]. It should be emphasized that apart from its independent intellectual merit, studying multidimensional systems is of practical importance since most storage media are two- or three-dimensional, including magnetic recording devices such as hard drives, optical recording devices such as CDs and DVDs, and flash memories.

The approach we take in this work is bounding the capacity by studying the independence entropy of SCSs, thus extending the works [27], [28]. The independence entropy appeared in previous works on d-dimensional shifts of finite type. Although this notion was first defined in [27], the idea stemmed from tradeoff functions studied in [29]. It was defined in a combinatorial fashion, where in this work we redefine it in a probabilistic fashion. We show that the two definitions are equal for the special case of fully constrained systems.

Manuscript received September 17, 2017; revised February 5, 2018; accepted March 24, 2018. Date of publication April 9, 2018; date of current version September 13, 2018. This work was supported in part by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) through REA under Grant 333598 and in part by the Israel Science Foundation under Grant 626/14. This paper was presented at the 2017 IEEE International Symposium on Information Theory.

O. Elishco and M. Schwartz are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva 8410501, Israel (e-mail: ohadeli@post.bgu.ac.il; schwartz@ee.bgu.ac.il).

T. Meyerovitch is with the Department of Mathematics, Ben-Gurion University of the Negev, Beer Sheva 8410501, Israel (e-mail: mtom@math.bgu.ac.il).

Communicated by N. Merhav, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2018.2824562

The motivation for the use of independence entropy is the fact that it is more easily computable, since we only need to consider independent probability measures which satisfy the constraints. We also focus on the class of d-dimensional axial-product constraints, which form a significant proportion of multidimensional fully constrained systems studied thus far. For this class, our approach has an additional major advantage in that instead of calculating the independence entropy for a d-dimensional axial product SCS, we may calculate it directly from the one-dimensional system. This dimensionality reduction offers further simplification of the calculations.

There are new features and difficulties that come up when adapting the results from fully constrained systems. We observe that fully constrained systems can be interpreted within the framework of semiconstrained systems. We elaborate on this basic yet crucial point in the following section. From this somewhat unconventional perspective, fully constrained systems are viewed as certain subsets of measures. It turns out that even in an abstract sense, the subsets of measures that describe fully constrained systems are very special among semiconstrained systems because they possess the following “extremal” property: If a measure μ is contained in (the set of measures associated to) some fully constrained system and μ is a convex combination of measures, then each of them is contained in the (set of measures associated to) the same fully constrained system. This property does not hold for general semiconstrained systems. This extremal property is manifested, for instance, in the fact that any subword of an admissible word in a fully constrained system is also admissible, leading to sub-multiplicativity of the sequence counting the number of admissible words of each length. The absence of “extremality” for more general semiconstrained systems leads to new features and difficulties. In particular, the absence of sub-multiplicativity forces us to avoid the use of Fekete’s Lemma to prove the existence of a limit for sequences related to the capacity.

The main contributions of this paper are a formulation of the independence entropy for SCSs, and its study in relation to the capacity of SCSs. As a result, we obtain a new lower bound on the capacity of multidimensional SCSs, generalizing the results of [27] and [28], and in an example test case, improving upon the best known bounds on the capacity of multidimensional $(0, 1, p)$-RLL SCSs given in [1].

In this work we also establish an equality of the limiting capacity as $d \to \infty$ and independence entropy for the d-axial-product SCSs. As the independence entropy is a lower bound on the capacity of a given SCS in every dimension, the capacity approaches the independence entropy as the dimension grows.

This paper is organized as follows. In Section II we describe the notation and give the required definitions used throughout the paper. In Section III we define the independence entropy and provide results characterizing the independence entropy. In Section IV we show that the capacity is lower bounded by the independence entropy. In Section V we show that the limiting capacity of the d-axial-product SCS as $d \to \infty$ is equal to the independence entropy. We conclude in Section VI by describing a short case study, and comparing it with previous results. The appendices provide proofs that the generalized notions we define in this paper indeed contain fully constrained systems as a special case, thus providing a generalization for them.

II. Preliminaries

Let $\mathbb{N}$ denote the set of natural numbers. We use $\mathbf{e}_i$ to denote the unit vector of direction i, $\mathbf{0}$ to denote the all-zero vector, and $\mathbf{1}$ to denote the all-one vector, where in all cases, the dimension of the vectors is implied by the context. For $n \in \mathbb{N}$ we define

$$[n] \triangleq \{0, 1, \dots, n-1\}.$$

We shall often use $[n]\mathbf{e}_i$ to denote the set $\{0 \cdot \mathbf{e}_i, 1 \cdot \mathbf{e}_i, \dots, (n-1) \cdot \mathbf{e}_i\}$. For $d, n \in \mathbb{N}$, denote by F_n^d the d-dimensional cube of length n, i.e., the set $F_n^d \triangleq [n]^d$. Obviously $|F_n^d| = n^d$. Additionally, for $(n_0, \dots, n_{d-1}) \in \mathbb{N}^d$ we conveniently denote

$$[(n_0, \dots, n_{d-1})] \triangleq [n_0] \times [n_1] \times \cdots \times [n_{d-1}].$$

Throughout the paper, Σ will be used to denote a finite alphabet. A word (or block) w of length n is a sequence of n letters from Σ, denoted $w = a_0\, a_1 \dots a_{n-1}$, with $a_i \in \Sigma$. We let $|w|$ denote the length of the word w. We can also consider infinite-sized words by a mapping from positions on the integer grid $\mathbb{Z}^d$ to letters from Σ (also known as configurations). Such a word will be denoted by $x \in \Sigma^{\mathbb{Z}^d}$, and the letter in the $\mathbf{v} \in \mathbb{Z}^d$ position will be denoted by $x_{\mathbf{v}}$ (sometimes referred to as the restriction of x to $\mathbf{v}$). More generally, given any subset of the integer grid, $S \subseteq \mathbb{Z}^d$, a word $x \in \Sigma^S$ is a mapping of positions indexed by elements of S to letters from Σ.

We require a notation for sets of probability measures and their marginals. For a set W we denote by $\mathcal{P}(W)$ the set of all probability measures over W.

Definition 1: Let $(X, \mathcal{B})$ be a measurable space. For every $\mu, \nu \in \mathcal{P}(X)$, the total variation distance *is defined as*

$$\|\mu - \nu\|_{TV} \triangleq \sup_{W \in \mathcal{B}} |\mu(W) - \nu(W)|.$$

□

Given a compact topological space X, the space $\mathcal{P}(X)$ is itself a compact topological space with respect to the weak $*$-topology. In particular, when X is a finite set with the discrete topology, the topology on $\mathcal{P}(X)$ is given by the total variation distance which also satisfies $\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{x \in X} |\mu(x) - \nu(x)|$.

Given a continuous map $f : X \to Y$ between topological spaces, and $\mu \in \mathcal{P}(X)$, let $f(\mu) \in \mathcal{P}(Y)$ be given by

$$f\ (\mu)(W) \triangleq \mu(f^{-1}(W)),\ \ W \subseteq Y.$$

Definition 2: For $d \in \mathbb{N}$, $S \subseteq \tilde{S} \subseteq \mathbb{Z}^d$, and $x \in \Sigma^{\tilde{S}}$, let x_S denote the restriction of x to the coordinates in S. Let $\pi_S^{\tilde{S}} : \Sigma^{\tilde{S}} \to \Sigma^S$ denote the restriction map given by

$$\pi_S^{\tilde{S}}(x) \triangleq x_S.$$

When $\tilde{S}$ is clear from the context, we shall write π_S instead of $\pi_S^{\tilde{S}}$. □

While having the notation $\pi_S(x)$ in addition to the equivalent notation x_S, seems superfluous, we shall require the former to simplify our presentation. As a consequence of the previous definition, for $\mu \in \mathcal{P}(\Sigma^{\tilde{S}})$ and $S \subseteq \tilde{S}$, we note that $\pi_S(\mu) \in \mathcal{P}(\Sigma^S)$ is the S-marginal of μ.

Definition 3: For $d \in \mathbb{N}$, $\mathbf{v} \in \mathbb{Z}^d$, let $\sigma_{\mathbf{v}} : \Sigma^{\mathbb{Z}^d} \to \Sigma^{\mathbb{Z}^d}$ be the shift by the vector $\mathbf{v}$*, given by*

$$(\sigma_{\mathbf{v}}(x))_{\mathbf{u}} \triangleq x_{\mathbf{u}+\mathbf{v}}, \ \mathbf{u} \in \mathbb{Z}^d, \ x \in \Sigma^{\mathbb{Z}^d}.$$

We denote by $\mathcal{P}_{\mathrm{si}}(\Sigma^{\mathbb{Z}^d})$ the space of shift-invariant probability measures on $\Sigma^{\mathbb{Z}^d}$, namely,

$$\mathcal{P}_{\mathrm{si}}(\Sigma^{\mathbb{Z}^d}) \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{\mathbb{Z}^d}) \ : \ \sigma_{\mathbf{v}}(\mu) = \mu \text{ for all } \mathbf{v} \in \mathbb{Z}^d\right\},$$

where $\sigma_{\mathbf{v}}(\mu) \triangleq \mu \circ \sigma_{\mathbf{v}}^{-1}$. For $k \in \mathbb{N}$ we say that $\mu \in \mathcal{P}(\Sigma^{F_k^d})$ is shift invariant if it is the projection of some shift-invariant measure on $\Sigma^{\mathbb{Z}^d}$, i.e., if there exists $\tilde{\mu} \in \mathcal{P}_{\mathrm{si}}(\Sigma^{\mathbb{Z}^d})$ such that $\mu = \pi_{F_k^d}\tilde{\mu}$. We denote by $\mathcal{P}_{\mathrm{si}}(\Sigma^{F_k^d})$ the space of shift-invariant probability measures on $\Sigma^{F_k^d}$, namely,

$$\mathcal{P}_{\mathrm{si}}(\Sigma^{F_k^d}) \triangleq \pi_{F_k^d}(\mathcal{P}_{\mathrm{si}}(\Sigma^{\mathbb{Z}^d})) \subseteq \mathcal{P}(\Sigma^{F_k^d}).$$

□

In the one-dimensional case, $d = 1$, it is rather easy to check whether a given probability measure $\mu \in \mathcal{P}(\Sigma^{F_k^1})$ is shift invariant. Indeed, $\mu \in \mathcal{P}_{\mathrm{si}}(\Sigma^{F_k^1})$ if and only if it satisfies the following finite system of linear equations,

$$\sum_{a\in\Sigma} \mu(a, a_1, \dots, a_{k-1}) = \sum_{a\in\Sigma} \mu(a_1, \dots, a_{k-1}, a),$$

for all $a_1, \dots, a_{k-1} \in \Sigma$.

When $d \geqslant 2$ the space of finite marginals of shift invariant measures becomes much more complicated. It is still not difficult to formulate an analogous system of linear equations that are satisfied for every $\mu \in \mathcal{P}_{\mathrm{si}}(\Sigma^{F_k^d})$. However, these linear conditions are no longer sufficient conditions for shift invariance. In fact, the problem of checking whether a given $\mu \in \mathcal{P}(\Sigma^{F_k^d})$ is shift invariant, is undecidable (assuming some computable representation of μ). See for instance [30], and references within, for a related discussion.

We are interested in defining empirical distributions of words. To that end, we give some more general definitions that we then specialize to our specific needs. Given $x \in \Sigma^{\mathbb{Z}^d}$, the delta measure at x, denoted by $\delta_x \in \mathcal{P}(\Sigma^{\mathbb{Z}^d})$, is defined by $\delta_x(\{x\}) = 1$. Additionally, given $n \in \mathbb{N}$, the *empirical measure associated with x and n*, denoted $\mathrm{fr}_{x,n} \in \mathcal{P}(\Sigma^{\mathbb{Z}^d})$, is given by

$$\mathrm{fr}_{x,n} \triangleq \frac{1}{n^d} \sum_{\mathbf{v}\in F_n^d} \delta_{\sigma_{\mathbf{v}}(x)}.$$

For $S \subseteq \mathbb{Z}^d$ we can take the S-marginal, and define $\mathrm{fr}^S_{x,n} \in \mathcal{P}(\Sigma^S)$ by

$$\mathrm{fr}^S_{x,n} \triangleq \pi_S(\mathrm{fr}_{x,n}).$$

Any word $w \in \Sigma^{F_n^d}$ may be extended periodically to the entire integer grid $\hat{w} \in \Sigma^{\mathbb{Z}^d}$ by defining

$$\hat{w}_{\mathbf{v}} \triangleq w_{\mathbf{v} \bmod n}$$

for all $\mathbf{v} \in \mathbb{Z}^d$, and where the modulo is taken entry-wise. The empirical distribution we shall be requiring may now be defined.

Definition 4: Let $d, n \in \mathbb{N}$, $w \in \Sigma^{F_n^d}$, and $S \subseteq \mathbb{Z}^d$. The empirical distribution *of w with respect to S, denoted fr^S_w, is defined by*

$$\mathrm{fr}^S_w \triangleq \mathrm{fr}^S_{\hat{w},n} .$$

□

Combinatorially speaking, the empirical distribution fr^S_w is obtained by cyclically scanning w with an S-shaped window and recording the frequency of the S-tuples in w. Thus, for instance, given a word $w = w_0 \dots w_{n-1} \in \Sigma^n$, $w_i \in \Sigma$, and $a \in \Sigma^k$ we have

$$\mathrm{fr}^{[k]}_w(a) = \frac{1}{|w|} \sum_{i=0}^{|w|-1} \mathbb{1}_a(w_i \dots w_{i+k-1})$$

where all coordinate indices are taken modulo $|w|$, and $\mathbb{1}_a : \Sigma^k \to \{0, 1\}$ is the indicator function of the singleton $\{a\}$.

Example 5: Let $\Sigma = \{0, 1\}$ and let $w = 0010111001 \in \Sigma^{F^1_{10}}$. We have that $|F^1_{10}| = 10$ and

$$\mathrm{fr}^{[3]}_w(110) = \frac{1}{10} \sum_{i=0}^{9} \mathbb{1}_{110}(w_i w_{i+1} w_{i+2}) = \frac{1}{10},$$

$$\mathrm{fr}^{[2]}_w(10) = \frac{1}{10} \sum_{i=0}^{9} \mathbb{1}_{10}(w_i w_{i+1}) = \frac{3}{10}.$$

□

Example 6: Let $\Sigma = \{0, 1\}$ and consider

$$w = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \in \Sigma^{F_4^2}, \quad a = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in \Sigma^{F_2^2}.$$

Then $\mathrm{fr}^{F_2^2}_w(a) = \frac{2}{16}$ since, of the sixteen 2×2 windows, exactly two contain a, shown in bold in the following:

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & \mathbf{0} & \mathbf{1} \\ 1 & 0 & \mathbf{1} & \mathbf{0} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{0} & 1 & 1 & \mathbf{1} \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ \mathbf{1} & 0 & 1 & \mathbf{0} \end{bmatrix}.$$

□

Lemma 7: Suppose $d, n \in \mathbb{N}$, $w \in \Sigma^{F_n^d}$, and $S \subseteq \tilde{S} \subseteq \mathbb{Z}^d$. Then

$$\pi^{\tilde{S}}_S(\mathrm{fr}^{\tilde{S}}_w) = \mathrm{fr}^S_w .$$

Proof: Let us denote $\mu \triangleq \mathrm{fr}_{\hat{w},n} \in \mathcal{P}(\Sigma^{\mathbb{Z}^d})$. By definition, for the right-hand side of the claim, for every $W \subseteq \Sigma^S$,

$$\mathrm{fr}^S_w(W) = \pi^{\mathbb{Z}^d}_S(\mu)(W) = \mu\left((\pi^{\mathbb{Z}^d}_S)^{-1}(W)\right).$$

Similarly, for the left-hand side,

$$\begin{aligned} \pi^{\tilde{S}}_S(\mathrm{fr}^{\tilde{S}}_w)(W) &= \pi^{\tilde{S}}_S\left(\pi^{\mathbb{Z}^d}_{\tilde{S}}(\mu)\right)(W) \\ &= \pi^{\mathbb{Z}^d}_{\tilde{S}}(\mu)\left((\pi^{\tilde{S}}_S)^{-1}(W)\right) \\ &= \mu\left((\pi^{\mathbb{Z}^d}_{\tilde{S}})^{-1}\left((\pi^{\tilde{S}}_S)^{-1}(W)\right)\right). \end{aligned}$$

But clearly for all $A \subseteq \Sigma^S$,

$$(\pi_S^{\mathbb{Z}^d})^{-1}(W) = (\pi_{\tilde{S}}^{\mathbb{Z}^d})^{-1}\left((\pi_S^{\tilde{S}})^{-1}(W)\right).$$

■

Lemma 7 implies that the empirical frequency of S-tuples in w can be calculated by first calculating the empirical frequency of $\tilde{S}$-tuples, and then taking the S-marginal.

Example 8: Let $\Sigma = \{0, 1\}$ *and consider*

$$w = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \in \Sigma^{F_4^2}.$$

Take $S = [1]^2 = \{(0,0)\}$ *and* $\tilde{S} = [(2,1)] = \{(0,0),(1,0)\}$. *Then*

$$\mathrm{fr}_w^{\tilde{S}}(00) = \frac{2}{16}, \quad \mathrm{fr}_w^{\tilde{S}}(01) = \frac{5}{16},$$
$$\mathrm{fr}_w^{\tilde{S}}(10) = \frac{5}{16}, \quad \mathrm{fr}_w^{\tilde{S}}(11) = \frac{4}{16}.$$

Moreover, we have that

$$\mathrm{fr}_w^{S}(0) = \frac{7}{16}, \quad \mathrm{fr}_w^{S}(1) = \frac{9}{16}.$$

We can verify now that

$$\begin{aligned} \pi_S^{\tilde{S}}(\mathrm{fr}_w^{\tilde{S}})(0) &= \mathrm{fr}_w^{\tilde{S}}\left((\pi_S^{\tilde{S}})^{-1}(0)\right) \\ &= \mathrm{fr}_w^{\tilde{S}}(\{00, 01\}) \\ &= \frac{7}{16} \\ &= \mathrm{fr}_w^{S}(0). \end{aligned}$$

□

We are now ready to define multidimensional semiconstrained systems.

Definition 9: For $d \in \mathbb{N}$, *a* $\mathbb{Z}^d$-semiconstrained system *(SCS) is a set* $\Gamma \subseteq \mathcal{P}(\Sigma^S)$ *for some finite set* $S \subseteq \mathbb{Z}^d$. *For* $n \in \mathbb{N}$, *the* admissible n-blocks *of* Γ *are*

$$\mathcal{B}_n(\Gamma) \triangleq \left\{ w \in \Sigma^{F_n^d} \ : \ \mathrm{fr}_w^S \in \Gamma \right\}.$$

□

Since all SCSs we study in this paper are $\mathbb{Z}^d$-SCSs, we shall abbreviate and call them just SCSs, where the dimension, d, will be clear from the context.

Note that SCSs generalize a subclass of d-dimensional fully constrained systems that correspond to subshifts of finite type in symbolic dynamics. Recall that those fully constrained systems are defined by a set of "forbidden patterns", $A \subseteq \Sigma^{F_k^d}$, such that a word $w \in \Sigma^{\mathbb{Z}^d}$ is admissible if and only if none of the elements of A appear as an F_k^d-tuple of w. In our notation, we therefore have the following.

Definition 10: For $d, k \in \mathbb{N}$, *we say that* $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$ *is* fully constrained *if there exists some* $L \subseteq \Sigma^{F_k^d}$ *such that*

$$\Gamma = \{\mu \in \mathcal{P}(\Sigma^{F_k^d}) \ : \ \mu(L) = 1\}.$$

□

Example 11: Let $\Sigma = \{0, 1\}$, *take*

$$L = \Sigma^{F_2^2} \setminus \left\{ \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\},$$

and consider the fully constrained system, Γ, *defined by*

$$\Gamma = \left\{ \mu \in \mathcal{P}(\Sigma^{F_2^2}) \ : \ \mu(L) = 1 \right\}.$$

Note that $\mathcal{B}_n(\Gamma)$ *is the set of all* $n \times n$ *two-dimensional binary arrays such that none of the six patterns above appears within a* 2×2 *window in them. It is simple to verify that in fact, no two horizontally adjacent* 1*'s may appear, and no two vertically adjacent* 1*'s may appear, in any admissible word. Thus, the* $n \times n$ *arrays in* $\mathcal{B}_n(\Gamma)$ *are the admissible words of the (cyclical)* $(1, \infty)$*-RLL fully constrained system.* □

An important figure of merit we associate with any set of words, and in particular, with SCSs, is the capacity, which we now define.

Definition 12: Let $d \in \mathbb{N}$, *and let* $S \subseteq \mathbb{Z}^d$ *be a finite subset. For any SCS,* $\Gamma \subseteq \mathcal{P}(\Sigma^S)$, *and for* $\epsilon > 0$, *let*

$$\mathbb{B}_\epsilon(\Gamma) \triangleq \left\{ \mu \in \mathcal{P}(\Sigma^S) \ : \ \inf_{\nu \in \Gamma} \|\mu - \nu\|_{TV} \leqslant \epsilon \right\}.$$

The capacity *of* Γ *is defined as,*

$$\mathsf{cap}(\Gamma) \triangleq \lim_{\epsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n^d} \log_2 (|\mathcal{B}_n(\mathbb{B}_\epsilon(\Gamma))|).$$

□

First, we mention that $\lim_{\epsilon \to 0^+}$ in the definition of the capacity exists due to monotonicity, since $|\mathcal{B}_n(\mathbb{B}_\epsilon(\Gamma))|$ is non-increasing in ϵ.

To avoid certain pathological scenarios, [1], [2] defined sets of weakly-admissible words and their capacity. We contend that the capacity definition provided here is the proper multidimensional generalization of these definitions. Intuitively, the capacity measures the exponential growth rate of the number of words that "almost" satisfy the semiconstraints given by Γ. Additionally, it has the nice property that the capacity of a set Γ is equal to the capacity of the closure of Γ.

At first glance this definition of capacity may seem odd. A naive definition, which we call the *internal capacity*, might be as follows.

Definition 13: Let $d \in \mathbb{N}$, $S \subseteq \mathbb{Z}^d$ *finite, and* $\Gamma \subseteq \mathcal{P}(\Sigma^S)$ *be a SCS. The* internal capacity *of* Γ *is defined as*

$$\widehat{\mathsf{cap}}(\Gamma) \triangleq \limsup_{n \to \infty} \frac{1}{n^d} \log_2 (|\mathcal{B}_n(\Gamma)|).$$

□

By definition we have

$$\mathsf{cap}(\Gamma) = \lim_{\epsilon \to 0^+} \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma))$$

which means that

$$\widehat{\mathsf{cap}}(\Gamma) \leqslant \mathsf{cap}(\Gamma). \tag{1}$$

We also observe that for some "nice" SCSs Γ, $\widehat{\mathsf{cap}}(\Gamma) = \mathsf{cap}(\Gamma)$. For instance, we have the following result for one-dimensional SCSs.

Theorem 14: [2, Section 2] Let $k \in \mathbb{N}$, *and* $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ *be convex and equal to the closure of its relative interior in* $\mathcal{P}_{\mathrm{si}}(\Sigma^k)$. *Then*

$$\mathsf{cap}(\Gamma) = \widehat{\mathsf{cap}}(\Gamma) = \log_2 |\Sigma| - \inf_{\eta \in \Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k)} H(\eta|\mu),$$

where $H(\eta|\mu)$ *is the relative entropy function*[1] *and* μ *is defined by* $\mu(\phi a) \triangleq \frac{1}{|\Sigma|}\sum_{a' \in \Sigma} \eta(\phi a')$ *for all* $\phi \in \Sigma^{k-1}$ *and* $a \in \Sigma$.

Remark 15: Consider the (compact) space $M = \mathcal{P}(\Sigma^S)$ *and let* $C(M)$ *be the set of all closed (hence, compact) subsets of* M. *Thus,* $C(M)$ *is a compact topological space (under the Hausdorff metric). Since* $\widehat{\mathsf{cap}}(\Gamma)$ *is monotone, the set of* Γ*s for which* $\widehat{\mathsf{cap}}(\Gamma) \neq \mathsf{cap}(\Gamma)$ *is meager. In other words, if we consider* $\widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))$ *as a function of* ϵ, $f(\epsilon) \triangleq \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma))$, *then* $\mathsf{cap}(\mathbb{B}_\delta(\Gamma)) = \widehat{\mathsf{cap}}(\mathbb{B}_\delta(\Gamma))$ *whenever* f *is continuous in* δ. *Since* f *is a monotone function, it is discontinuous on a countable number of places. In practice, it means that if for a specific* Γ, $\mathsf{cap}(\Gamma) \neq \widehat{\mathsf{cap}}(\Gamma)$ *an arbitrary small change in* Γ *will give an equality.* □

Remark 16: For a fully constrained system, $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$, *non-emptyness of* $\mathcal{B}_n(\Gamma)$ *for all* $n > 0$ *is equivalent to the fact that the subshift of finite type*

$$\left\{ w \in \Sigma^{\mathbb{Z}^d} \;:\; \forall \mathbf{v} \in \mathbb{Z}^d, (\sigma_{\mathbf{v}}(w))_{F_k^d} \in L \right\},$$

is not empty. Berger's Theorem [31] implies that it is undecidable whether a subshift of finite type is empty given L, for $d > 1$. *Because (under reasonable assumptions on the representation) it is undecidable if a given multidimensional SCS is non-empty, it is difficult to understand what a SCS really looks like.* □

At this point we pause to ponder the following: Note that the definition of empirical frequency is cyclic (in the sense that coordinates are taken modulo n) while in traditional fully constrained systems it is not. This seems at odds with our claim of SCSs generalizing fully constrained systems. The necessity of the modulo in the definition of SCSs stems from working with the space of shift-invariant measures and their associated admissible words. Shift-invariant measures are defined over $\mathbb{Z}^d$, hence, it is necessary to complete a word $w \in \Sigma^{F_n^d}$ to a word from $\Sigma^{\mathbb{Z}^d}$. We choose to do this completion periodically using the modulo notion, extending w to $\hat{w}$. This choice simplifies the analysis which follows. We contend that with respect to this issue, the capacity is more natural than the internal capacity, since it is equal to the non-cyclic capacity of fully constrained systems. To avoid a lengthy detour, the full details are provided in Appendix A.

Finally, we raise the question: what multidimensional SCSs are of interest? If we examine the extensive literature for fully constrained systems, a significant proportion of multidimensional fully constrained systems are defined as an axial product of one-dimensional fully constrained systems. Intuitively speaking, if we have a set of "forbidden patterns" defining a one-dimensional fully constrained system, we can define its d-dimensional axial product by forbidding these patterns along each dimension. We now formally define this for the case of d-dimensional SCSs with slightly more generality. This definition generalizes the d-dimensional axial product defined in [27].

Definition 17: Consider $S_0, \dots, S_{d-1} \subseteq \mathbb{N}$, *with* $0 \in S_i$ *for all* $i \in [d]$, *and SCSs* $\Gamma_i \subseteq \mathcal{P}(\Sigma^{S_i})$. *Denote* $S \triangleq \bigcup_{i \in [d]} S_i \mathbf{e}_i \subseteq \mathbb{Z}^d$. *The* d-axial-product SCS, *denoted* $\otimes_{i \in [d]} \Gamma_i$, *is defined by*

$$\otimes_{i \in [d]} \Gamma_i \triangleq \left\{ \mu \in \mathcal{P}(\Sigma^S) \;:\; \forall i \in [d],\ \pi_{S_i \mathbf{e}_i}(\mu) \in \Gamma_i \right\}.$$

□

It follows from the above definition, that for every $n \in \mathbb{N}$ we have

$$\mathcal{B}_n\left(\otimes_{i \in [d]} \Gamma_i\right) = \left\{ w \in F_n^d \;:\; \forall i \in [d],\ \mathrm{fr}_w^{S_i \mathbf{e}_i} \in \Gamma_i \right\},$$

with coordinates taken modulo n. Intuitively, the arrays of a d-axial-product SCS satisfy that along the ith direction, the empirical distribution of S_i-tuples is in Γ_i. Note that $\otimes_{i \in [d]} \Gamma_i$ induces a set of measures over $\Sigma^{F_k^d}$ where $k = \max_i \{k_i \;:\; k_i \in S_i\}$. Hence, we sometimes consider a d-axial-product SCS $\otimes_{i \in [d]} \Gamma_i$ as a subset of $\mathcal{P}(\Sigma^{F_k^d})$.

Example 18: Let $\Sigma = \{0, 1\}$ *Consider two real constants* $0 \leqslant p_0, p_1 \leqslant 1$, *and the one-dimensional SCSs,* Γ_0 *and* Γ_1, *given by*

$$\Gamma_0 = \left\{ \mu \in \mathcal{P}(\Sigma^2) \;:\; \mu(11) \leqslant p_0 \right\},$$
$$\Gamma_1 = \left\{ \mu \in \mathcal{P}(\Sigma^2) \;:\; \mu(11) \leqslant p_1 \right\}.$$

Here we are taking $S_0 = S_1 = \{0, 1\}$ *The admissible words in the 2-axial-product SCS,* $\Gamma_0 \otimes \Gamma_1$, *are all two-dimensional words in which the empirical frequency of two horizontally adjacent* 1*s is at most* p_0, *and the empirical frequency of two vertically adjacent* 1*s is at most* p_1, *i.e., all the words* $w \in \Sigma^{F_n^2}$ *such that*

$$\mathrm{fr}_w^{\{(0,0),(1,0)\}}(11) \leqslant p_0,$$
$$\mathrm{fr}_w^{\{(0,0),(0,1)\}}(11) \leqslant p_1.$$

We may also consider $\Gamma_0 \otimes \Gamma_1$ *as a subset of* $\mathcal{P}(\Sigma^{F_2^2})$

$$\Gamma_0 \otimes \Gamma_1 = \left\{ \mu \in \mathcal{P}(\Sigma^{F_2^2}) \;:\; \pi_{\{(0,0),(0,1)\}}(\mu)(11) \leqslant p_0, \right.$$
$$\left. \pi_{\{(0,0),(1,0)\}}(\mu)(11) \leqslant p_1 \right\}.$$

Note that

$$\pi_{\{(0,0),(1,0)\}}(\mu)(11) = \mu\left(\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\right) + \mu\left(\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}\right) + \mu\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\right) + \mu\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right),$$

$$\pi_{\{(0,0),(0,1)\}}(\mu)(11) = \mu\left(\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\right) + \mu\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\right) + \mu\left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\right) + \mu\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right).$$

□

[1] The relative entropy function is also referred to as the KL-divergence function, which is defined as $D(\eta\|\mu) \triangleq \int \log\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}\right)\mathrm{d}\eta$.

In this paper we are interested in the capacity and the internal capacity of multidimensional SCSs. Although the capacity is easier to work with, as we will see later on, the task of computing it is still daunting. Thus, there is a necessity for more easily computable bounds on the capacity. To this end, we define the independence entropy of a d-dimensional SCS, which is the basis of the main results of this paper.

III. Independence Entropy

In this section we define the independence entropy of multidimensional SCSs and present some of its properties. It will be used to bound the capacity. The independence entropy is not a new notion, and has appeared previously in [27] in relation to the capacity of fully constrained systems. However, the formulation of the independence entropy was combinatorial and therefore less suitable for our purposes. Thus, we modify the definition of independence entropy and formulate it as a statistical notion.

The admissible words of SCSs (see Definition 9) have their empirical S-tuple distribution from Γ. Finding such words inexorably involves intricate dependencies between coordinates. This affects not only the task of generating such words, but also the very basic problem of calculating or bounding the capacity of the SCS – the problem that is the focus of this paper.

In an attempt to simplify this problem, we study the independence-entropy approach. We eliminate all dependencies by considering only product measures, i.e., where the symbol in each coordinate of the word is chosen independently of other coordinates. Accordingly, we only require the *average* of S-marginals to be in Γ. We then ask what is the entropy of such a system. Intuitively, we are seeking the maximum rate of transmission in a system where word coordinates are transmitted independently and in parallel, designed such that the average S-marginals are in Γ. The following model provides a rough interpretation of the independence entropy: Suppose each bit of the output is transmitted by a different agent, and the number of agents is very large. The agents are allowed to coordinate a protocol in advance, but are unable to communicate once they receive the messages to be transmitted. In addition, the statistics of the output should roughly satisfy the constraints given by Γ, with high probability (as a function of the number of agents). In this case under suitable assumptions, the maximal transmission rate would coincide with the independence entropy. We proceed with formal definitions, starting with a product measure.

Definition 19: Let $d \in \mathbb{N}$, and let $S \subseteq \mathbb{Z}^d$ be a finite set. We say that $\mu \in \mathcal{P}(\Sigma^S)$ is an independent probability measure *or* a product measure *if $\mu(w) = \prod_{\mathbf{v}\in S} \pi_{\{\mathbf{v}\}}(\mu)(w)$. For $S \subseteq \mathbb{Z}^d$ that is possibly infinite, $\mu \in \mathcal{P}(\Sigma^S)$ is a product measure whenever $\pi_{S'}(\mu)$ is a product measure for every finite $S' \subseteq S$.* □

In other words, we say that μ is independent if there exists $\{p_{\mathbf{v}} \in \mathcal{P}(\Sigma) \ : \ \mathbf{v} \in S\}$ such that $\mu = \prod_{\mathbf{v}\in S} p_{\mathbf{v}}$. We naturally identify the set of product measures in $\mathcal{P}(\Sigma^S)$ with $(\mathcal{P}(\Sigma))^S$.

Next, we define the average of a marginal.

Definition 20: Given $d, n \in \mathbb{N}$, $\mu \in \mathcal{P}(\Sigma^{F_n^d})$, and $S \subseteq F_n^d$, let $\overline{\pi}_S(\mu) \in \mathcal{P}(\Sigma^S)$ be the average of the S-marginals over translates of μ:

$$\overline{\pi}_S(\mu) \triangleq \frac{1}{|F_n^d|} \sum_{\mathbf{v}\in F_n^d} \pi_{S+\mathbf{v}}(\mu),$$

where the coordinates $S + \mathbf{v}$ are taken modulo n. □

Let $S \subseteq F_k^d$ and let $\Gamma \subseteq \mathcal{P}(\Sigma^S)$ be a SCS. For $n \geqslant k$ we define

$$\overline{\mathcal{P}}_n(\Gamma) \triangleq \left\{\mu \in (\mathcal{P}(\Sigma))^{F_n^d} \ : \ \overline{\pi}_S(\mu) \in \Gamma\right\}.$$

Thus, $\overline{\mathcal{P}}_n(\Gamma)$ consists of product measures on $\Sigma^{F_n^d}$ such that the average of the S-marginals is in Γ. We can now define the independence entropy of a SCS.

Definition 21: Let $d, k \in \mathbb{N}$, $S \subseteq F_k^d$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^S)$ be a d-dimensional SCS. The internal independence entropy *of Γ is defined by*

$$\widehat{h_{\mathrm{ind}}}(\Gamma) \triangleq \limsup_{n\to\infty} \sup_{\mu\in\overline{\mathcal{P}}_n(\Gamma)} \frac{1}{n^d} H(\mu),$$

where $H(\mu) \triangleq -\sum_{w\in\Sigma^{F_n^d}} \mu(w)\log_2 \mu(w)$ is the entropy of μ. The independence entropy *of Γ is defined by*

$$h_{\mathrm{ind}}(\Gamma) \triangleq \lim_{\epsilon\to 0^+} \widehat{h_{\mathrm{ind}}}(\mathbb{B}_\epsilon(\Gamma)).$$

□

Again, it is clear by definition that

$$\widehat{h_{\mathrm{ind}}}(\Gamma) \leqslant h_{\mathrm{ind}}(\Gamma). \tag{2}$$

The notion of independence entropy which appears here is a generalization of the combinatorial notion for fully constrained systems that appears in [27].

Theorem 22: Let $d, k \in \mathbb{N}$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$ be a fully constrained system. Then

$$h_{\mathrm{ind}}(\Gamma) = h_{\mathrm{ind}}^{\mathrm{com}}(\Gamma)$$

where $h_{\mathrm{ind}}^{\mathrm{com}}$ is the combinatorial independence entropy from [27].

To avoid a significant diversion from the main discussion, the proof of Theorem 22, together with the required definitions from [27], are given in Appendix B.

We now show properties of $\widehat{h_{\mathrm{ind}}}$ and h_{ind} which make them easier to analyze by reducing the multidimensional case to the one-dimensional case. We start with an inequality given in the following lemma. The proof follows the same argument that was used in [27] to show the inequality for fully constrained systems. However, the equality for fully constrained systems holds in an easier and stronger sense.

Lemma 23: Let $k \in \mathbb{N}$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a one-dimensional SCS. Then for all $d \in \mathbb{N}$,

$$\widehat{h_{\mathrm{ind}}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}(\Gamma^{\otimes d}).$$

Proof: Take $\hat{\mu} \in \overline{\mathcal{P}}_n(\Gamma)$. Since $\hat{\mu}$ is a product measure, it can be written as $\hat{\mu} = \prod_{i=0}^{n-1} \pi_{\{i\}}(\hat{\mu})$. We now construct a measure $\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})$ using $\hat{\mu}$. For every $\mathbf{v} \in F_n^d$ set

$$\pi_{\{\mathbf{v}\}}(\mu) \triangleq \pi_{\{\ell(\mathbf{v})\}}(\hat{\mu}),$$

where $\ell(\mathbf{v}) \triangleq \left(\sum_{i=0}^{d-1} v_i\right) \bmod n$ is the modulo n of the sum of the coordinates of $\mathbf{v}$.

Observe that μ is such that in every row in every direction, i.e., a set of coordinates of the form $\mathbf{v}+[n]\mathbf{e}_i$, we obtain some cyclic rotation of $\hat{\mu}$ by t positions, denoted $\sigma_t(\hat{\mu})$. However, $\hat{\mu} \in \overline{\mathcal{P}}_n(\Gamma)$ implies $\sigma_t(\hat{\mu}) \in \overline{\mathcal{P}}_n(\Gamma)$. Thus, we obtain that $\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})$ and

$$\frac{1}{n}H(\hat{\mu}) = \frac{1}{n^d}H(\mu).$$

Since we are taking the supremum over all measures $\hat{\mu}$, we have $\widehat{h_{\mathrm{ind}}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}(\Gamma^{\otimes d})$. ■

Theorem 24: Let $k \in \mathbb{N}$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a one-dimensional SCS. Then for all $d \in \mathbb{N}$,

$$h_{\mathrm{ind}}(\Gamma^{\otimes d}) = h_{\mathrm{ind}}(\Gamma).$$

Proof: We first show that $h_{\mathrm{ind}}(\Gamma^{\otimes d}) \leqslant h_{\mathrm{ind}}(\Gamma)$. Fix $\delta > 0$ and take $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma^{\otimes d}))$. Recall that

$$\overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma^{\otimes d})) \subseteq \mathcal{P}(\Sigma^{F_n^d}).$$

Let $(\mathbf{v}_i)_{i\in[n^{d-1}]}$ be an enumeration of $\{0\}\times F_n^{d-1}$, i.e.,

$$\{\mathbf{v}_0, \dots, \mathbf{v}_{n^{d-1}-1}\} = \{0\}\times F_n^{d-1}.$$

For $i \in [n^{d-1}]$, define $\mu_i \in \mathcal{P}(\Sigma^n)$ by $\mu_i \triangleq \pi_{[n]\mathbf{e}_0+\mathbf{v}_i}(\mu)$. Now let $\hat{\mu} \in \mathcal{P}(\Sigma^{n^d})$ be the product measure that is the product of all the μ_i's. This means that for a word $a = a_0 \dots a_{n^d-1} \in \Sigma^{n^d}$,

$$\hat{\mu}(a) \triangleq \mu_0(a_0 \dots a_{n-1}) \cdot \mu_1(a_n \dots a_{2n-1}) \cdots \cdot \mu_{n^d-1}(a_{n(n^{d-1}-1)} \dots a_{n^d-1}),$$

Since each of the μ_i's is already a product measure, $\hat{\mu} \in \mathcal{P}(\Sigma^{n^d})$ is also a product measure. We have

$$\begin{aligned}
&\overline{\pi}_{[k]}(\hat{\mu})\\
&= \frac{1}{n^d}\sum_{j=0}^{n^d-1}\pi_{j+[k]}(\hat{\mu})\\
&= \frac{1}{n^d}\left(\sum_{i=0}^{n^{d-1}-1}\sum_{j=in}^{(i+1)n-1}\pi_{j+[k]}(\hat{\mu})\right)\\
&= \frac{1}{n^d}\Bigg(\sum_{i=0}^{n^{d-1}-1}\Bigg(\sum_{j=in}^{(i+1)n-k}\pi_{j+[k]}(\hat{\mu}) + \sum_{j=(i+1)n-k+1}^{(i+1)n-1}\pi_{j+[k]}(\hat{\mu})\Bigg)\Bigg)\\
&\overset{(a)}{=} \frac{1}{n^d}\Bigg(\sum_{i=0}^{n^{d-1}-1}\sum_{j=in}^{(i+1)n-k}\pi_{(j-in)+[k]}(\mu_i) + \sum_{i=0}^{n^{d-1}-1}\sum_{j=(i+1)n-k+1}^{(i+1)n-1}\pi_{j+[k]}(\hat{\mu})\Bigg)\\
&= \frac{1}{n^d}\Bigg(\sum_{i=0}^{n^{d-1}-1}\sum_{j=in}^{(i+1)n-1}\pi_{(j-in)+[k]}(\mu_i) - \sum_{i=0}^{n^{d-1}-1}\sum_{j=(i+1)n-k+1}^{(i+1)n-1}\pi_{(j-in)+[k]}(\mu_i) + \sum_{i=0}^{n^{d-1}-1}\sum_{j=(i+1)n-k+1}^{(i+1)n-1}\pi_{j+[k]}(\hat{\mu})\Bigg)\\
&= \frac{1}{n^{d-1}}\Bigg(\sum_{i=0}^{n^{d-1}-1}\overline{\pi}_{[k]}(\mu_i) - \frac{1}{n}\sum_{i=0}^{n^{d-1}-1}\sum_{j=(i+1)n-k+1}^{(i+1)n-1}\left(\pi_{(j-in)+[k]}(\mu_i) - \pi_{j+[k]}(\hat{\mu})\right)\Bigg)\\
&= \overline{\pi}_{[k]\mathbf{e}_0}(\mu) - \frac{1}{n^d}\sum_{i=0}^{n^{d-1}-1}\sum_{j=(i+1)n-k+1}^{(i+1)n-1}\left(\pi_{(j-in)+[k]}(\mu_i) - \pi_{j+[k]}(\hat{\mu})\right),
\end{aligned}$$

where (a) follows from the definition of $\hat{\mu}$ and since the coordinates are taken modulo n when calculating $\pi_{[k]}(\mu_i)$. Each $\left(\pi_{(j-in)+[k]}(\mu_i) - \pi_{j+[k]}(\hat{\mu})\right)$ is a signed measure of total variation norm at most 1. Therefore,

$$\left\|\overline{\pi}_{[k]}(\hat{\mu}) - \overline{\pi}_{[k]\mathbf{e}_0}(\mu)\right\|_{TV} \leqslant \frac{k}{n}.$$

This means that $\overline{\pi}_{[k]}(\hat{\mu}) \in \mathbb{B}_{\frac{k}{n}+\delta}(\Gamma)$. We obtained that for every $\epsilon > \delta > 0$, and every $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma^{\otimes d}))$, we can find $n_0 \in \mathbb{N}$ such that for every $n > n_0$, $\hat{\mu} \in \overline{\mathcal{P}}_{n^d}(\mathbb{B}_\epsilon(\Gamma))$. Since μ and $\hat{\mu}$ are both product measures we have

$$\begin{aligned}
H(\mu) &= \sum_{\mathbf{v}\in F_n^d} H(\pi_{\{\mathbf{v}\}}(\mu))\\
&= \sum_{i\in[n^d]} H(\pi_{\{i\}}(\hat{\mu}))\\
&= H(\hat{\mu}).
\end{aligned}$$

This implies that for every $\epsilon > \delta > 0$,

$$\limsup_{n\to\infty}\sup_{\mu\in\overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma^{\otimes d}))}\frac{1}{n^d}H(\mu) \leqslant \limsup_{n\to\infty}\sup_{\mu\in\overline{\mathcal{P}}_{n^d}(\mathbb{B}_\epsilon(\Gamma))}\frac{1}{n^d}H(\mu).$$

We therefore obtain $h_{\mathrm{ind}}(\Gamma^{\otimes d}) \leqslant \widehat{h_{\mathrm{ind}}}(\mathbb{B}_\epsilon(\Gamma))$ for every $\epsilon > 0$. Taking the limit as $\epsilon \to 0^+$, by the definition of $h_{\mathrm{ind}}(\Gamma)$ we have

$$h_{\mathrm{ind}}(\Gamma^{\otimes d}) \leqslant h_{\mathrm{ind}}(\Gamma).$$

We now show the other direction. By Lemma 23, For every $\delta > 0$ we have

$$\widehat{h_{\mathrm{ind}}}(\mathbb{B}_\delta(\Gamma)) \leqslant \widehat{h_{\mathrm{ind}}}(\mathbb{B}_\delta(\Gamma)^{\otimes d}).$$

By monotonicity of $\widehat{h_{\mathrm{ind}}}$ it thus follows that for every $\delta > 0$,

$$h_{\mathrm{ind}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}(\mathbb{B}_\delta(\Gamma)^{\otimes d}).$$

Now observe that for every $\epsilon > 0$ there exists $\delta > 0$ so that

$$\mathbb{B}_\delta(\Gamma)^{\otimes d} \subseteq \mathbb{B}_\epsilon\left(\Gamma^{\otimes d}\right).$$

It follows that for every $\epsilon > 0$

$$h_{\mathrm{ind}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}\left(\mathbb{B}_\epsilon\left(\Gamma^{\otimes d}\right)\right).$$

Thus, by taking the limit $\epsilon \to 0^+$,

$$h_{\mathrm{ind}}(\Gamma) \leqslant h_{\mathrm{ind}}\left(\Gamma^{\otimes d}\right).$$

■

We conclude this section by noting that Lemma 23 and Theorem 24 show that for $\Gamma \subseteq \mathcal{P}(\Sigma^k)$,

$$\widehat{h_{\text{ind}}}(\Gamma) \leqslant \widehat{h_{\text{ind}}}(\Gamma^{\otimes d}) \leqslant h_{\text{ind}}(\Gamma^{\otimes d}) = h_{\text{ind}}(\Gamma). \tag{3}$$

IV. Independence Entropy Lower Bounds the Capacity

This section and the next explore the relationship between the independence entropy and the capacity. In this section we show that the capacity of any d-dimensional SCS (not necessarily an axial product) is lower bounded by the independence entropy.

Before proceeding we require a simple lemma.

Lemma 25: Let $d, n \in \mathbb{N}$, and $S \subseteq F_n^d$, then π_S and $\overline{\pi}_S$ are contractions with respect to the total-variation distance, i.e., for all $\mu, \nu \in \mathcal{P}(\Sigma^{F_n^d})$,

$$\|\pi_S(\mu) - \pi_S(\nu)\|_{TV} \leqslant \|\mu - \nu\|_{TV},$$
$$\|\overline{\pi}_S(\mu) - \overline{\pi}_S(\nu)\|_{TV} \leqslant \|\mu - \nu\|_{TV}.$$

Proof: For every $W \subseteq \Sigma^S$ we have

$$\begin{aligned}|\pi_S(\mu)(W) - \pi_S(\nu)(W)| &= \left|\mu(\pi_S^{-1}(W)) - \nu(\pi_S^{-1}(W))\right| \\ &\leqslant \sup_{A' \subset \Sigma^S} \left|\mu(W') - \nu(W')\right| \\ &= \|\mu - \nu\|_{TV}.\end{aligned}$$

Hence the function $\pi_{S+\mathbf{v}}$ is a contraction for every $\mathbf{v} \in F_n^d$. Then $\overline{\pi}_S$, being an average of contractions, is itself a contraction. ∎

We are now ready to state and prove the main result of this section – a lower bound on the capacity. The corresponding result for fully constrained systems was obtained in [27].

Theorem 26: Let $d \in \mathbb{N}$, $S \subseteq \mathbb{Z}^d$ be a finite set, and let $\Gamma \subseteq \mathcal{P}(\Sigma^S)$ be a SCS. Then $h_{\text{ind}}(\Gamma) \leqslant \mathsf{cap}(\Gamma)$.

Proof: For ease of reading, throughout the proof, we omit the superscript F_n^d unless the shape is different or essential. Hence, instead of $\text{fr}^{F_n^d}$ we write fr, instead of $\hat{\text{fr}}^{F_n^d}$ we write $\hat{\text{fr}}$, and instead of $\pi^{F_n^d}$ we write π. Fix $\delta > 0$, $n \in \mathbb{N}$ such that $S \subseteq F_n^d$, and let $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma))$. For $m \in \mathbb{N}$, we have a natural identification isomorphism $\Sigma^{F_{nm}^d} \cong (\Sigma^{F_n^d})^{F_m^d}$ that identifies $\mathbf{v} \in F_{nm}^d$ with the unique pair $\mathbf{r} \in F_n^d$ and $\mathbf{q} \in F_m^d$ such that $\mathbf{v} = n\mathbf{q} + \mathbf{r}$. Consider the product measure $\mu^m \in \mathcal{P}(\Sigma^{F_n^d})^{F_m^d} \subseteq \mathcal{P}(\Sigma^{F_{nm}^d})$ satisfying

$$\mu^m(\{x\}) = \prod_{\mathbf{v} \in F_m^d} \mu(\pi_{F_n^d}(\sigma_{n\mathbf{v}}(x))).$$

Note that since μ is a product measure, μ^m is also a product measure.

For a word $w \in \Sigma^{F_{nm}^d}$, denote by $\hat{\text{fr}}_w$ the empirical distribution of *non-overlapping* F_n^d-tuples, i.e.,

$$\hat{\text{fr}}_w \triangleq \frac{1}{|F_m^d|} \sum_{\mathbf{u} \in F_m^d} \delta_{\pi_{F_n^d}(\sigma_{n\mathbf{u}}(\hat{w}))}.$$

Additionally, observe that

$$\frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \hat{\text{fr}}_{\sigma_{\mathbf{v}}(w)} = \text{fr}_w.$$

Also, because pi_S is an affine map, it follows that

$$\frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \pi_S\left(\hat{\text{fr}}_{\sigma_{\mathbf{v}}(w)}\right) = \pi_S(\text{fr}_w).$$

By Lemma 7, $\pi_S(\text{fr}_w) = \text{fr}_w^S$.

Note that by the construction of μ^m we have $\overline{\pi}_S(\mu) = \overline{\pi}_S(\mu^m)$, and we obtain,

$$\begin{aligned}
&\left\|\text{fr}_w^S - \overline{\pi}_S(\mu^m)\right\|_{TV} \\
&= \left\|\text{fr}_w^S - \overline{\pi}_S(\mu)\right\|_{TV} \\
&= \left\|\frac{1}{|F_{nm}^d|} \sum_{\mathbf{u} \in F_{nm}^d} \pi_S(\delta_{\sigma_{\mathbf{u}}(w)}) - \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \pi_{S+\mathbf{v}}(\mu)\right\|_{TV} \\
&= \left\|\frac{1}{|F_n^d||F_m^d|} \sum_{\mathbf{v} \in F_n^d} \sum_{\mathbf{u} \in F_m^d} \pi_S(\delta_{\sigma_{n\mathbf{u}+\mathbf{v}}(w)}) - \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \pi_{S+\mathbf{v}}(\mu)\right\|_{TV} \\
&\overset{(a)}{=} \left\|\frac{1}{|F_n^d||F_m^d|} \sum_{\mathbf{v} \in F_n^d} \sum_{\mathbf{u} \in F_m^d} \pi_{S+\mathbf{v}}(\delta_{\sigma_{n\mathbf{u}}(w)}) - \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \pi_{S+\mathbf{v}}(\mu)\right\|_{TV} \\
&\overset{(b)}{\leqslant} \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \left\|\frac{1}{|F_m^d|} \sum_{\mathbf{u} \in F_m^d} \pi_{S+\mathbf{v}}(\delta_{\sigma_{n\mathbf{u}}(w)}) - \pi_{S+\mathbf{v}}(\mu)\right\|_{TV} \\
&\overset{(c)}{=} \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \left\|\pi_{S+\mathbf{v}}\left(\frac{1}{|F_m^d|} \sum_{\mathbf{u} \in F_m^d} \delta_{\sigma_{n\mathbf{u}}(w)}\right) - \pi_{S+\mathbf{v}}(\mu)\right\|_{TV} \\
&= \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \left\|\pi_{S+\mathbf{v}}(\hat{\text{fr}}_w) - \pi_{S+\mathbf{v}}(\mu)\right\|_{TV} \\
&\overset{(d)}{\leqslant} \frac{1}{|F_n^d|} \sum_{\mathbf{v} \in F_n^d} \left\|\hat{\text{fr}}_w - \mu\right\|_{TV} \\
&= \left\|\hat{\text{fr}}_w - \mu\right\|_{TV}
\end{aligned}$$

where:

- (a) follows since $\pi_S(\delta_{\sigma_{\mathbf{v}}(w)}) = \pi_{S+\mathbf{v}}(\delta_w)$.
- (b) follows by the triangle inequality.
- (c) follows since π_S is an affine map.
- (d) follows by Lemma 25.

Thus, for $\epsilon > \delta$, if $\|\hat{\text{fr}}_w - \mu\|_{TV} < \epsilon - \delta$ then $\|\text{fr}_w^S - \overline{\pi}_S(\mu^m)\|_{TV} < \epsilon - \delta$. Therefore,

$$\begin{aligned}&\left\{w \in \Sigma^{F_{nm}^d} : \left\|\text{fr}_w^S - \overline{\pi}_S(\mu)\right\|_{TV} \geqslant \epsilon - \delta\right\} \\ &\qquad \subseteq \left\{w \in \Sigma^{F_{nm}^d} : \left\|\hat{\text{fr}}_w - \mu\right\|_{TV} \geqslant \epsilon - \delta\right\}.\end{aligned}$$

Using the fact that $\overline{\pi}_S(\mu) \in \mathbb{B}_\delta(\Gamma)$, it follows that

$$\left\{w \in \Sigma^{F^d_{nm}} \ : \ \mathrm{fr}^S_w \notin \mathrm{int}\,(\mathbb{B}_\epsilon(\Gamma))\right\} \subseteq \left\{w \in \Sigma^{F^d_{nm}} \ : \ \left\|\hat{\mathrm{fr}}_w - \mu\right\|_{TV} \geqslant \epsilon - \delta\right\}, \tag{4}$$

where $\mathrm{int}(\cdot)$ denotes the interior of a set, i.e., $\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)) = \left\{\nu \in \mathcal{P}(\Sigma^S) \ : \ \inf_{\mu\in\Gamma} \|\nu-\mu\|_{TV} < \epsilon\right\}$.

If $w \in \Sigma^{F^d_{nm}}$ was randomly drawn according to μ^m, the non-overlapping F^d_n-tuples are distributed i.i.d. according to μ. Apply Cramer's Theorem (as in [32, Th. 2.2.3, Remark c]) to deduce that for $\epsilon > \delta$ and for every m,

$$\mu^m\left(\left\{w \in \Sigma^{F^d_{nm}} \ : \ \left\|\hat{\mathrm{fr}}_w - \mu\right\|_{TV} \geqslant \epsilon - \delta\right\}\right) \leqslant 2\exp\left(-m \inf_{\nu\in\mathcal{P}(\Sigma^{F^d_n}):\ \|\nu-\mu\|_{TV}\geqslant\epsilon-\delta} H(\nu|\mu)\right).$$

Note that the function $\nu \times \mu \mapsto H(\nu|\mu)$ is continuous and strictly positive off the diagonal. Thus, for every $\epsilon > \delta$ we have

$$c_\mu(\epsilon) \triangleq \inf_{\nu\in\mathcal{P}(\Sigma^{F^d_n}):\ \|\nu-\mu\|_{TV}\geqslant\epsilon-\delta} H(\nu|\mu) > 0.$$

Hence

$$\mu^m\left(\left\{w \in \Sigma^{F^d_{nm}} \ : \ \left\|\hat{\mathrm{fr}}_w - \mu\right\|_{TV} \geqslant \epsilon - \delta\right\}\right) \leqslant 2\exp\left(-mc_\mu(\epsilon)\right). \tag{5}$$

Recall that

$$\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma))) = \left\{w \in \Sigma^{F^d_{nm}} \ : \ \mathrm{fr}^S_w \in \mathrm{int}(\mathbb{B}_\epsilon(\Gamma))\right\}.$$

By (4), we have

$$\begin{aligned}\mu^m&\left(\Sigma^{F^d_{nm}} \setminus \mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))\right)\\ &= \mu^m\left(\left\{w \in \Sigma^{F^d_{nm}} \ : \ \mathrm{fr}^S_w \notin \mathrm{int}(\mathbb{B}_\epsilon(\Gamma))\right\}\right)\\ &\leqslant \mu^m\left(\left\{w \in \Sigma^{F^d_{nm}} \ : \ \left\|\hat{\mathrm{fr}}_w - \mu\right\|_{TV} \geqslant \epsilon - \delta\right\}\right).\end{aligned} \tag{6}$$

Combining (5) and (6) we have,

$$\xi \triangleq \mu^m\left(\Sigma^{F^d_{nm}} \setminus \mathcal{B}_{nm}\ (\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))\right) \leqslant 2\exp\left(-mc_\mu(\epsilon)\right).$$

It now follows that,

$$\begin{aligned}&\frac{1}{n^d}H(\mu)\\ &\quad= \frac{1}{(nm)^d}H(\mu^m)\\ &\quad= -\frac{1}{(nm)^d}\sum_{w\in\Sigma^{F^d_{nm}}} \mu^m(w)\log_2\mu^m(w)\\ &\quad= -\frac{1}{(nm)^d}\sum_{w\in\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))} \mu^m(w)\log_2\mu^m(w)\\ &\qquad\quad -\frac{1}{(nm)^d}\sum_{w\notin\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))} \mu^m(w)\log_2\mu^m(w)\\ &\quad\overset{(a)}{\leqslant} (1-\xi)\cdot\frac{\log_2|\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))|}{(nm)^d}\\ &\qquad\quad + \xi\cdot\frac{\log_2\left|\Sigma^{F^d_{nm}} \setminus \mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))\right|}{(nm)^d} + H_2(\xi)\\ &\quad\leqslant \frac{\log_2|\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))|}{(nm)^d}\\ &\qquad\quad + 2e^{-mc_\mu(\epsilon)}\frac{\log_2|\Sigma|^{(nm)^d}}{(nm)^d} + H_2(\xi)\\ &\quad\leqslant \frac{1}{(nm)^d}\log_2|\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))|\\ &\qquad\quad + 2e^{-mc_\mu(\epsilon)}\frac{1}{(nm)^d}\log_2|\Sigma|^{(nm)^d} + H_2(\xi).\end{aligned}$$

where (a) follows from standard maximization of entropy arguments, and where $H_2(\xi) \triangleq -\xi\log_2\xi - (1-\xi)\log_2(1-\xi)$ is the binary entropy function. This implies

$$\begin{aligned}\frac{1}{n^d}H(\mu) &= \limsup_{m\to\infty}\frac{1}{n^d}H(\mu)\\ &\leqslant \limsup_{m\to\infty}\frac{1}{(nm)^d}\log_2|\mathcal{B}_{nm}(\mathrm{int}(\mathbb{B}_\epsilon(\Gamma)))|\\ &\leqslant \limsup_{m\to\infty}\frac{1}{(nm)^d}\log_2|\mathcal{B}_{nm}(\mathbb{B}_\epsilon(\Gamma))|\\ &\leqslant \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma)),\end{aligned}$$

This is true for every $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma))$ and hence

$$\sup_{\mu\in\overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma))}\frac{1}{n^d}H(\mu) \leqslant \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma)).$$

Since this holds for every n we have that for every $\epsilon > \delta > 0$,

$$\widehat{h_{\mathrm{ind}}}(\mathbb{B}_\delta(\Gamma)) \leqslant \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma)).$$

Taking the limit as $\delta \to 0$, this implies that for every $\epsilon > 0$,

$$h_{\mathrm{ind}}(\Gamma) \leqslant \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma)).$$

Finally, taking the limit as $\epsilon \to 0$, it follows that

$$h_{\mathrm{ind}}(\Gamma) \leqslant \mathsf{cap}(\Gamma).$$

■

We summarize our results thus far by noting that for a SCS $\Gamma \subseteq \mathcal{P}(\Sigma^k)$, since $\widehat{h_{\mathrm{ind}}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}(\Gamma^{\otimes d})$, Theorem 26 together with (3) show that

$$\widehat{h_{\mathrm{ind}}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}(\Gamma^{\otimes d}) \leqslant h_{\mathrm{ind}}(\Gamma^{\otimes d}) \leqslant \mathsf{cap}(\Gamma^{\otimes d}), \tag{7}$$

$$\widehat{h_{\mathrm{ind}}}(\Gamma) \leqslant \widehat{h_{\mathrm{ind}}}(\Gamma^{\otimes d}) \leqslant h_{\mathrm{ind}}(\Gamma^{\otimes d}) = h_{\mathrm{ind}}(\Gamma) \leqslant \mathsf{cap}(\Gamma). \tag{8}$$

V. Upper Bound on Limiting Capacity

In this section we prove that if $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ is a convex one-dimensional SCS and $\Gamma^{\otimes d}$ its d-axial product, then

$$\limsup_{d\to\infty}\mathsf{cap}(\Gamma^{\otimes d}) \leqslant h_{\mathrm{ind}}(\Gamma^{\otimes d}).$$

The main idea is to show that for any $\epsilon > 0$ we are able to find d large enough for which the independence entropy is ϵ-close to $\mathsf{cap}(\Gamma^{\otimes d})$. This is the main result of [28] and the proof here is an adaptation of it.

Before going into details we introduce a different form of d-axial product which we call the *weak* d-axial product. For a one dimensional SCS, $\Gamma \subseteq \mathcal{P}(\Sigma^k)$, define

$$\Gamma^{\boxtimes d} \triangleq \left\{ \mu \in \mathcal{P}(\Sigma^{F_k^d}) \ : \ \frac{1}{d}\sum_{i\in[d]} \pi_{[k]\mathbf{e}_i}(\mu) \in \Gamma \right\},$$

and thus

$$\mathcal{B}_n\left(\Gamma^{\boxtimes d}\right) = \left\{ w \in F_n^d \ : \ \frac{1}{d}\sum_{i\in[d]} \mathrm{fr}_w^{[k]\mathbf{e}_i} \in \Gamma \right\}.$$

For the weak d-axial product we define,

$$\overline{\mathcal{P}}_n(\Gamma^{\boxtimes d}) \triangleq \left\{ \mu \in (\mathcal{P}(\Sigma))^{F_n^d} \ : \ \frac{1}{d}\sum_{i\in[d]} \overline{\pi}_{[k]\mathbf{e}_i}(\mu) \in \Gamma \right\}.$$

This last definition is a relaxed version of $\Gamma^{\otimes d}$, since $\overline{\mathcal{P}}_n(\Gamma^{\otimes d})$ is the set of all independent measures for which the average of the k-marginals in each direction (separately) belongs to Γ, whereas $\overline{\mathcal{P}}_n(\Gamma^{\boxtimes d})$ is the set of all independent measures for which the average of k-marginals (over all directions) belongs to Γ.

Correspondingly, we have,

$$h_{\mathrm{ind}}(\Gamma^{\boxtimes d}) \triangleq \lim_{\epsilon\to 0^+} \limsup_{n\to\infty} \sup_{\mu\in\overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma)^{\boxtimes d})} \frac{1}{n^d} H(\mu),$$

where $H(\mu) \triangleq -\sum_{w\in\Sigma^{F_n^d}} \mu(w)\log_2\mu(w)$ is the entropy of μ.

As will become clearer later on, it will be somewhat easier to use $h_{\mathrm{ind}}(\Gamma^{\boxtimes d})$ than $h_{\mathrm{ind}}(\Gamma^{\otimes d})$ in this section. First, the following lemma shows that the relaxation leading to $h_{\mathrm{ind}}(\Gamma^{\boxtimes d})$ does not affect the independence entropy.

Lemma 27: Let $k \in \mathbb{N}$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a convex one-dimensional SCS, then

$$h_{\mathrm{ind}}(\Gamma) = h_{\mathrm{ind}}(\Gamma^{\otimes d}) = h_{\mathrm{ind}}(\Gamma^{\boxtimes d}).$$

Proof: By Theorem 24 we already know that $h_{\mathrm{ind}}(\Gamma^{\otimes d}) = h_{\mathrm{ind}}(\Gamma)$. Thus, we are left with proving the last equality. Since Γ is convex, for every $\delta > 0$,

$$\overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma^{\otimes d})) \subseteq \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma)^{\otimes d}) \subseteq \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma)^{\boxtimes d}).$$

Hence,

$$h_{\mathrm{ind}}(\Gamma^{\otimes d}) \leqslant h_{\mathrm{ind}}(\Gamma^{\boxtimes d}).$$

The other direction follows essentially by using the same method as in the proof of Theorem 24, as we now describe. Let $(\mathbf{v}_i^j)_{i\in[n^{d-1}]}$ be an enumeration of $F_n^{j-1}\times\{0\}\times F_n^{d-j}$, i.e.,

$$\left\{\mathbf{v}_0^j,\ldots,\mathbf{v}_{n^{d-1}-1}^j\right\} = F_n^{j-1}\times\{0\}\times F_n^{d-j}.$$

Fix $\delta > 0$ and $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma)^{\boxtimes d})$. For $i \in [n^{d-1}]$ and $j \in [d]$, define $\mu_i^j \in \mathcal{P}(\Sigma^n)$ by $\mu_i^j \triangleq \pi_{[n]\mathbf{e}_j+\mathbf{v}_i^j}(\mu)$. Now let $\hat{\mu} \in \mathcal{P}(\Sigma^{dn^d})$ be the product measure satisfying

$$\hat{\mu}(\{a\}) = \prod_{j\in[d]}\prod_{i\in[n^{d-1}]} \mu_i^j(a_{in+jn^d}\ldots a_{(i+1)n+jn^d-1})$$

for every word $a = a_0\ldots a_{dn^d-1} \in \Sigma^{dn^d}$. It is clear that $\hat{\mu}$ is indeed a product measure, because every μ_i^j is also a product measure. Now,

$$\begin{aligned}
&\overline{\pi}_{[k]}(\hat{\mu}) \\
&= \frac{1}{dn^d}\sum_{i\in[dn^d]} \pi_{i+[k]}(\hat{\mu}) \\
&= \frac{1}{dn^d}\sum_{j\in[d]}\sum_{i\in[n^{d-1}]}\sum_{\ell\in[n]} \pi_{[k]+in+\ell+jn^d}(\hat{\mu}) \\
&= \frac{1}{dn^d}\sum_{j\in[d]}\sum_{i\in[n^{d-1}]}\left(\sum_{\ell\in[n-k]} \pi_{[k]+in+\ell+jn^d}(\hat{\mu}) + \sum_{\ell=n-k}^{n-1} \pi_{[k]+in+\ell+jn^d}(\hat{\mu})\right) \\
&= \frac{1}{dn^d}\sum_{j\in[d]}\sum_{i\in[n^{d-1}]}\left(\sum_{\ell\in[n-k]} \pi_{[k]+\ell}(\mu_i^j) + \sum_{\ell=n-k}^{n-1} \pi_{[k]+in+\ell+jn^d}(\hat{\mu})\right) \\
&= \frac{1}{dn^d}\sum_{j\in[d]}\sum_{i\in[n^{d-1}]}\left(\sum_{\ell\in[n]} \pi_{[k]+\ell}(\mu_i^j) - \sum_{\ell=n-k}^{n-1} \pi_{[k]+\ell}(\mu_i^j) + \sum_{\ell=n-k}^{n-1} \pi_{[k]+in+\ell+jn^d}(\hat{\mu})\right) \\
&\overset{(a)}{=} \frac{1}{d}\sum_{j\in[d]} \overline{\pi}_{[k]\mathbf{e}_j}(\mu) \\
&\quad - \frac{1}{dn^d}\sum_{j\in[d]}\sum_{i\in[n^{d-1}]}\sum_{\ell=n-k}^{n-1}\left(\pi_{[k]+\ell}(\mu_i^j) - \pi_{[k]+in+\ell+jn^d}(\hat{\mu})\right).
\end{aligned}$$

Recall that from the definition of $\overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma)^{\boxtimes d})$, we have

$$\frac{1}{d}\sum_{j\in[d]} \overline{\pi}_{[k]\mathbf{e}_j}(\mu) \in \mathbb{B}_\delta(\Gamma).$$

Since $\left(\pi_{[k]+\ell}(\mu_i^j) - \pi_{[k]+in+\ell+jn^d}(\hat{\mu})\right)$ is a signed measure of total variation norm at most 2, it follows that $\overline{\pi}_{[k]}(\hat{\mu}) \in \mathbb{B}_{\frac{2k}{n}+\delta}(\Gamma)$, so $\hat{\mu} \in \overline{\mathcal{P}}_{dn^d}\left(\mathbb{B}_{\frac{2k}{n}+\delta}(\Gamma)\right)$. Hence, for every $\epsilon > \delta > 0$, and every $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma)^{\boxtimes d})$, we can find $n_0 \in \mathbb{N}$ such that for every $n > n_0$, $\hat{\mu} \in \overline{\mathcal{P}}_{dn^d}(\mathbb{B}_\epsilon(\Gamma))$, and therefore,

$$\limsup_{n\to\infty} \sup_{\mu\in\overline{\mathcal{P}}_n(\mathbb{B}_\delta(\Gamma)^{\boxtimes d})} \frac{1}{n^d}H(\mu) \leqslant \limsup_{n\to\infty} \sup_{\mu\in\overline{\mathcal{P}}_{dn^d}(\mathbb{B}_\epsilon(\Gamma))} \frac{1}{dn^d}H(\mu).$$

Thus, we obtain $h_{\mathrm{ind}}(\Gamma^{\boxtimes d}) \leqslant \widehat{h_{\mathrm{ind}}}(\mathbb{B}_\epsilon(\Gamma))$ for every $\epsilon > 0$, and by definition it follows that

$$h_{\mathrm{ind}}(\Gamma^{\boxtimes d}) \leqslant h_{\mathrm{ind}}(\Gamma).$$

■

Given a probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$, denote by $L^2(\mathcal{X}, \mathcal{F}, \mathbb{P}, \mathbb{C}^n)$ the Hilbert space of $\mathcal{F}$-measurable functions $f : \mathcal{X} \to \mathbb{C}^n$ satisfying

$$\|f\|_{L^2}^2 \triangleq \int \langle f, f\rangle \mathrm{d}\mathbb{P} < \infty,$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{C}^n$.

The following lemma is based on Dirichlet's "pigeon hole principle" and different versions of it are used in many de-Finetti type proofs (see, for example, [33] [34, Lemma 4.1]).

Lemma 28: Let $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_m \subseteq \mathcal{F}$ be a sequence of sub-σ-algebras. Let $f \in L^2(\mathcal{X}, \mathcal{F}, \mathbb{P}, \mathbb{C}^n)$, and denote $f_j \triangleq E[f|\mathcal{F}_j]$, the conditional expectation of f with respect to the sub-σ-algebra $\mathcal{F}_j$. Then, there exists $t \in [m]$ such that

$$\|f_{t+1} - f_t\|_{L^2}^2 \leqslant \frac{1}{m}\|f\|_{L^2}^2$$

Proof: For every ℓ, let $V_\ell \triangleq L^2(\mathcal{X}, \mathcal{F}_\ell, \mathbb{P}, \mathbb{C}^n)$ denote the corresponding sub-space of the Hilbert space $V \triangleq L^2(\mathcal{X}, \mathcal{F}, \mathbb{P}, \mathbb{C}^n)$. Then f_ℓ is an orthogonal projection of f onto V_ℓ. Thus, $\langle f - f_\ell, g\rangle_{L^2} = 0$ for every $g \in V$. Therefore,

$$\|f_m\|_{L^2}^2 = \sum_{\ell \in [m]} \|f_{\ell+1} - f_\ell\|_{L^2}^2 + \|f_0\|_{L^2}^2.$$

Additionally, $0 \leqslant \|f_m\|_{L^2}^2 \leqslant \|f\|_{L^2}^2$. The result follows by noticing that if m non-negative real numbers sum to at most $\|f\|_{L^2}^2$ then the value of at least one element is at most $\frac{1}{m}\|f\|_{L^2}^2$. ■

Before stating the lemmas, we need the following notation. Recall that for $k \in \mathbb{N}$, we defined $[k] \triangleq \{0, \ldots, k-1\}$. We now define $[-k] \triangleq \{-1, \ldots, -k\}$.

Lemma 29: For every $\epsilon > 0$, and any $m \in \mathbb{N}$, there exists $d_0 \in \mathbb{N}$ such that for every $d \geqslant d_0$, and every $n, j \in \mathbb{N}$, $n \geqslant j+2$, there exists a sequence of $m+1$ random subsets $X_0, X_1, \ldots, X_m \subseteq F_n^d$, and random variables $I_{t,\mathbf{v}} \in [d]$, for all $t \in [m]$, $\mathbf{v} \in F_n^d$, all defined on an appropriate probability space $(\mathcal{X}, 2^{\mathcal{X}}, \mathbb{P})$, such that all the following hold:

1) $\mathbb{P}(X_i \subseteq X_{i+1}) = 1$ *for all* $i \in [m]$.
2) $\mathbb{P}(|X_m| \leqslant \epsilon |F_n^d|) \geqslant 1 - \epsilon$.
3) *For all $\mathbf{v} \in F_n^d$ and $t \in [m]$, $I_{t,\mathbf{v}}$ is distributed uniformly on $[d]$ and is independent of X_t. Furthermore, for every value of X_t,*

$$\mathbb{P}\left(X_t \cup ([-(j+1)]\mathbf{e}_{I_{t,\mathbf{v}}} + \mathbf{v}) \subseteq X_{t+1} \;\middle|\; X_t\right) \geqslant 1 - \epsilon.$$

Proof: Choose $0 < p < 1$ small enough so that $1 - (1 - p)^{m+1} \leqslant \frac{\epsilon}{2}$, and conveniently denote $p_i \triangleq 1 - (1 - p)^{i+1}$. For all $i \in [m+1]$, consider random subsets $A_i \subseteq F_n^d$ whose coordinates are chosen i.i.d. Bernoulli(p), i.e., $\mathbb{P}(\mathbf{v} \in A_i) = p$ for all $\mathbf{v} \in F_n^d$, independently of $F_n^d \setminus \{\mathbf{v}\}$. Define $X_{-1} \triangleq \emptyset$, and for all $i \in [m+1]$, define

$$X_i \triangleq X_{i-i} \cup A_i.$$

Thus, $\mathbb{P}(\mathbf{v} \in X_i) = p_i$ for all $\mathbf{v} \in F_n^d$, independently of $F_n^d \setminus \{\mathbf{v}\}$. We contend that for large enough d, the claims hold.

First, it is clear that $\mathbb{P}(X_i \subseteq X_{i+1}) = 1$ for $i \in [m+1]$ by construction. Second, we have

$$\mathbb{P}\left(|X_m| \leqslant \epsilon \;\middle|\; F_n^d\right) \geqslant \mathbb{P}\left(|X_m| < 2p_m \;\middle|\; F_n^d\right) \geqslant 1 - e^{-2p_m^2 n^d},$$

where the last inequality follows from Hoeffding's inequality. Since the right-hand side approaches 1 when $n \geqslant 2$ and $d \to \infty$, claim 2 holds for large enough d.

We now address claim 3. Fix $t \in [m]$ and consider A_{t+1}. For a coordinate $\mathbf{v} \in F_n^d$, denote by $D(t, \mathbf{v})$ the set

$$D(t, \mathbf{v}) \triangleq \{i \in [d] \;:\; \mathbf{v} + [-(j+1)]\mathbf{e}_i \subseteq A_{t+1}\}.$$

If $D(t, \mathbf{v}) \neq \emptyset$ then draw $I_{t,\mathbf{v}}$ uniformly from $D(t, \mathbf{v})$. Otherwise, draw $I_{t,\mathbf{v}}$ uniformly from $[d]$. Note that $I_{t,\mathbf{v}}$ is distributed uniformly on $[d]$ since the distribution of A_{t+1} is invariant under coordinate permutation. Since the coordinates in A_{t+1} are chosen independently of $A_t, A_{t-1}, \ldots, A_0$ we obtain that $I_{t,\mathbf{v}}$ is independent of X_t. Finally, we have

$$\begin{aligned}
&\mathbb{P}\left(X_t \cup ([-(j+1)]\mathbf{e}_{I_{t,\mathbf{v}}} + \mathbf{v}) \subseteq X_{t+1} \;\middle|\; X_t\right) \\
&\quad \geqslant \mathbb{P}(D(t, \mathbf{v}) \neq \emptyset) \\
&\quad = 1 - (1 - p^{j+1})^d.
\end{aligned}$$

Since the right-hand side approaches 1 as $d \to \infty$, claim 3 holds for large enough d. ■

If X is a random variable over some probability space, we use $\mathbb{P}_X$ to denote its distribution. Let $X_0, \ldots, X_{k-1}$ be random variables over the same probability space $(\mathcal{X}, 2^{\mathcal{X}}, \mathbb{P})$. We denote by $(X_0, \ldots, X_{k-1})$ the vector distributed according to their joint probability, $\mathbb{P}_{X_0,\ldots,X_{k-1}}$, and denote by $(X_0 \times \cdots \times X_{k-1})$ the vector distributed according to their product probability, i.e., $\mathbb{P}_{X_0 \times \cdots \times X_{k-1}} \triangleq \prod_{i \in [k]} \mathbb{P}_{X_i}$.

Lemma 30: Let $\mathcal{X}$ be a finite set, and $X_0, \ldots, X_{k-1}$ be k random variables defined over the same probability space $(\mathcal{X}, 2^{\mathcal{X}}, \mathbb{P})$. Then

$$\begin{aligned}
&\left\|\mathbb{P}_{X_0,\ldots,X_{k-1}} - \mathbb{P}_{X_0 \times \cdots \times X_{k-1}}\right\|_{TV} \\
&\qquad \leqslant \sum_{i=0}^{k-2} E_{X_0,\ldots,X_i}\left[\left\|\mathbb{P}_{X_{i+1}|X_0,\ldots,X_i} - \mathbb{P}_{X_{i+1}}\right\|_{TV}\right].
\end{aligned}$$

Proof: We prove this by induction on k. The case of $k = 1$ is trivially true. In the base case of $k = 2$ we have,

$$\begin{aligned}
&\left\|\mathbb{P}_{X_0,X_1} - \mathbb{P}_{X_0 \times X_1}\right\|_{TV} \\
&\quad = \frac{1}{2}\sum_{x_0,x_1} \left|\mathbb{P}_{X_0,X_1}(x_0, x_1) - \mathbb{P}_{X_0}(x_0)\mathbb{P}_{X_1}(x_1)\right| \\
&\quad = \frac{1}{2}\sum_{x_0,x_1} \left|\mathbb{P}_{X_0}(x_0)\mathbb{P}_{X_1|X_0}(x_1|x_0) - \mathbb{P}_{X_0}(x_0)\mathbb{P}_{X_1}(x_1)\right| \qquad (9)
\end{aligned}$$

where the sum of x_0 and x_1 is over the support of X_0 and X_1, respectively. Since $\mathbb{P}_{X_0}(x_0) \geqslant 0$ we have

$$\begin{aligned}
&\frac{1}{2}\sum_{x_0,x_1 \in \mathcal{X}} \left|\mathbb{P}_{X_0}(x_0)\mathbb{P}_{X_1|X_0}(x_1|x_0) - \mathbb{P}_{X_0}(x_0)\mathbb{P}_{X_1}(x_1)\right| \\
&= \sum_{x_0 \in \mathcal{X}} \mathbb{P}_{X_0}(x_0)\left(\frac{1}{2}\sum_{x_1 \in \mathcal{X}} \left|\mathbb{P}_{X_1|X_0}(x_1|x_0) - \mathbb{P}_{X_1}(x_1)\right|\right). \qquad (10)
\end{aligned}$$

Combining (9) and (10) and using the total variation distance definition we obtain

$$\left\|\mathbb{P}_{X_0,X_1} - \mathbb{P}_{X_0 \times X_1}\right\|_{TV} = E_{X_0}\left[\left\|\mathbb{P}_{X_1|X_0} - \mathbb{P}_{X_1}\right\|_{TV}\right].$$

Now assume the statement is correct for $k-1$ random variables and we show it is correct for k random variables. We write

$$\begin{aligned}&\left\|\mathbb{P}_{X_0,\dots,X_{k-1}}-\mathbb{P}_{X_0\times\cdots\times X_{k-1}}\right\|_{TV}\\&=\left\|\mathbb{P}_{X_0,\dots,X_{k-1}}-\mathbb{P}_{(X_0,\dots,X_{k-2})\times X_{k-1}}\right.\\&\quad\left.+\mathbb{P}_{(X_0,\dots,X_{k-2})\times X_{k-1}}-\mathbb{P}_{X_0\times\cdots\times X_{k-1}}\right\|_{TV}.\end{aligned}$$

By applying the triangle inequality we obtain

$$\begin{aligned}&\left\|\mathbb{P}_{X_0,\dots,X_{k-1}}-\mathbb{P}_{X_0\times\cdots\times X_{k-1}}\right\|_{TV}\\&\leqslant\left\|\mathbb{P}_{X_0,\dots,X_{k-1}}-\mathbb{P}_{(X_0,\dots,X_{k-2})\times X_{k-1}}\right\|_{TV}\\&\quad+\left\|\mathbb{P}_{(X_0,\dots,X_{k-2})\times X_{k-1}}-\mathbb{P}_{X_0\times\cdots\times X_{k-1}}\right\|_{TV}.\end{aligned}\tag{11}$$

Considering $Y=(X_0,\dots,X_{k-2})$ as a tuple-valued radom variable, and applying the case $k=2$ on the pair of random variables (Y,X_{k-1}) we have:

$$\begin{aligned}&\left\|\mathbb{P}_{X_0,\dots,X_{k-1}}-\mathbb{P}_{(X_0,\dots,X_{k-2})\times X_{k-1}}\right\|_{TV}\\&\leqslant E_{X_0,\dots,X_{k-2}}\left[\left\|\mathbb{P}_{X_{k-1|(X_0,\dots,X_{k-2})}}-\mathbb{P}_{X_{k-1}}\right\|_{TV}\right]\end{aligned}\tag{12}$$

It is easy to check that

$$\begin{aligned}&\left\|\mathbb{P}_{(X_0,\dots,X_{k-2})\times X_{k-1}}-\mathbb{P}_{X_0\times\cdots\times X_{k-1}}\right\|_{TV}\\&=\left\|\mathbb{P}_{X_0,\dots,X_{k-2}}-\mathbb{P}_{X_0\times\cdots\times X_{k-2}}\right\|_{TV}\end{aligned}\tag{13}$$

By the induction hypothesis we have

$$\begin{aligned}&\left\|\mathbb{P}_{X_0,\dots,X_{k-2}}-\mathbb{P}_{X_0\times\cdots\times X_{k-2}}\right\|_{TV}\\&\leqslant\sum_{i=0}^{k-3}E_{X_0,\dots,X_i}\left[\left\|\mathbb{P}_{X_{i+1}|X_0,\dots,X_i}-\mathbb{P}_{X_{i+1}}\right\|_{TV}\right].\end{aligned}$$

Combining this with (11), (12) and (13) completes the proof. ∎

For $A\subseteq F_n^d$, let $\mathcal{F}_A\subseteq 2^{\Sigma^{F_n^d}}$ denote the σ-algebra generated by the coordinates in A, namely,

$$\mathcal{F}_A\triangleq\left\{\left\{x\in\Sigma^{F_n^d}:\pi_A(x)\in W\right\}\;:\;W\subseteq\Sigma^A\right\}.$$

Definition 31: Let $d,k,n\in\mathbb{N}$, $A\subseteq F_n^d$, and let $y\in\Sigma^{F_n^d}$. For a one-dimensional SCS, $\Gamma\subseteq\mathcal{P}(\Sigma^k)$, and its d-axial-product SCS, $\Gamma^{\otimes d}$, we define the following:

$$\begin{aligned}&\mu^{n,d}\ \text{ is the uniform measure over }\mathcal{B}_n(\Gamma^{\otimes d}),\\&\mu_{y,A}\triangleq\mu^{n,d}\left(\cdot\mid\mathcal{F}_A\right)(y),\\&\eta_{y,A}\triangleq\prod_{\mathbf{v}\in F_n^d}\pi_{\{\mathbf{v}\}}\left(\mu_{y,A}\right).\end{aligned}$$

□

In other words, $\mu_{y,A}$ is the uniform distribution on $\mathcal{B}_n(\Gamma^{\otimes d})$ given whose positions in A agree with y_A. Moreover, $\eta_{y,A}$ is the independent version of $\mu_{y,A}$. The following statement is a particular application of Lemma 30 above.

Lemma 32: For every $d,n\in\mathbb{N}$, $i\in[d]$, and $A\subseteq F_n^d$, we have

$$\begin{aligned}&\sum_{\mathbf{v}\in F_n^d}E\left[\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})-\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})\right\|_{TV}\right]\\&\leqslant\sum_{\mathbf{v}\in F_n^d}\sum_{j\in[k]}E\left[\left\|\pi_{\{\mathbf{v}\}}(\eta_{y,A})-\pi_{\{\mathbf{v}\}}(\mu_{y,(A\cup([-j]\mathbf{e}_i+\mathbf{v}))})\right\|_{TV}\right].\end{aligned}$$

Proof: First note that if $k=1$ the result is immediate since all the summands on the left-hand side are 0. We now examine the case of $k\geqslant 2$. For the time being, let us fix $\mathbf{v}\in F_n^d$ and $y\in\Sigma^{F_n^d}$. We define the random variables X_j, $j\in[k]$, where $X_0,\dots,X_{k-1}$ is distributed according to $\mathbb{P}^y_{X_0,\dots,X_{k-1}}\triangleq\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})$. In particular, each X_j is distributed according to $\mathbb{P}^y_{X_j}\triangleq\pi_{j\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})=\pi_{j\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})$. Additionally, $\mathbb{P}^y_{X_0\times\cdots\times X_{k-1}}=\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})$. We use the superscript y to emphasize that these distributions depend y. Also for $z\in\Sigma^{F_n^d}$ such that $z_A=y_A$, the conditional probability $\mathbb{P}^y_{X_{j+1}|X_0,\dots,X_j}$ evaluated at z is equal to the measure $\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\mu_{z,A\cup([j+1]\mathbf{e}_i+\mathbf{v})})$. By Lemma 30, we have

$$\begin{aligned}&\left\|\mathbb{P}^y_{X_0,\dots,X_{k-1}}-\mathbb{P}^y_{X_0\times\cdots\times X_{k-1}}\right\|_{TV}\\&\leqslant\sum_{j=0}^{k-2}E\left[\left\|\mathbb{P}^y_{X_{j+1}|X_0,\dots,X_j}-\mathbb{P}^y_{X_{i+1}}\right\|_{TV}\right].\end{aligned}\tag{14}$$

The expectations in the right-hand side are with respect to the conditioning on the random variables $X_0,\dots,X_j$. We can rewrite the above equation as (15), as shown at the bottom of this page. Integrating the inequality (15) over y with respect to $\mu^{n,d}$ we have (16), as shown at the bottom of this page. By definition of $\mu_{y,A}$ as the conditional measure, for every $f:\Sigma^{F_n^d}\to\mathbb{R}$ we have

$$\iint f(z)d\mu_{y,A}(z)\mathrm{d}\mu^{n,d}(y)=\int f(y)\mathrm{d}\mu^{n,d}(y).$$

Writing the integeral with repect to $\mu^{n,d}$ as $E[\cdot]$, we thus have

$$\begin{aligned}&E\left[\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})-\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right]\\&\leqslant\sum_{j=0}^{k-2}E\left[\left\|\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\mu_{y,A\cup([j+1]\mathbf{e}_i+\mathbf{v})})\right.\right.\\&\qquad\left.\left.-\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right].\end{aligned}$$

$$\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})-\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\leqslant\sum_{j=0}^{k-2}\int\left\|\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\mu_{z,A\cup([j+1]\mathbf{e}_i+\mathbf{v})})-\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\eta_{z,A})\right\|_{TV}\mathrm{d}\mu_{y,A}(z)\tag{15}$$

$$\begin{aligned}&\int\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})-\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\mathrm{d}\mu^{n,d}(y)\\&\leqslant\sum_{j=0}^{k-2}\iint\left\|\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\mu_{z,A\cup([j+1]\mathbf{e}_i+\mathbf{v})})-\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\eta_{z,A})\right\|_{TV}\mathrm{d}\mu_{y,A}(z)\mathrm{d}\mu^{n,d}(y).\end{aligned}\tag{16}$$

Summing over all $\mathbf{v} \in F_n^d$ we obtain

$$\sum_{\mathbf{v}\in F_n^d} E\left[\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A}) - \pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right] \leqslant \sum_{\mathbf{v}\in F_n^d}\sum_{j=0}^{k-2} E\left[\left\|\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\mu_{y,A\cup([j+1]\mathbf{e}_i+\mathbf{v})}) - \pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right]. \tag{17}$$

Recall that $[-j] \triangleq \{-1,\dots,-j\}$, hence

$$[j+1]\mathbf{e}_i = (j+1)\mathbf{e}_i + [-(j+1)]\mathbf{e}_i.$$

Thus, (17) can be written as

$$\sum_{\mathbf{v}\in F_n^d} E\left[\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A}) - \pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right] \leqslant \sum_{\mathbf{v}\in F_n^d}\sum_{j=0}^{k-2} E\left[\left\|\pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\mu_{y,A\cup((j+1)\mathbf{e}_i+\mathbf{v}+[-(j+1)]\mathbf{e}_i)}) - \pi_{(j+1)\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right]. \tag{18}$$

Since we are summing over all $\mathbf{v} \in F_n^d$, and since coordinates are taken modulo n, we may write (18) as follows,

$$\sum_{\mathbf{v}\in F_n^d} E\left[\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A}) - \pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A})\right\|_{TV}\right] \leqslant \sum_{\mathbf{v}\in F_n^d}\sum_{j=0}^{k-2} E\left[\left\|\pi_{\{\mathbf{v}\}}(\mu_{y,A\cup(\mathbf{v}+[-(j+1)]\mathbf{e}_i)}) - \pi_{\{\mathbf{v}\}}(\eta_{y,A})\right\|_{TV}\right]. \tag{19}$$

Since the total variation distance is non-negative, (19) implies the lemma. ■

The following proposition, which is used to prove the main result of this section, considers the following scenario. Assume $y \in \Sigma^{F_n^d}$ is randomly drawn using the measure $\mu^{n,d}$, i.e., it is drawn uniformly at random from the set of admissible words $\mathcal{B}_n(\Gamma^{\otimes d})$. We then study the random variable $\eta_{y,A}$ (a measure in itself), and ask what is the probability that it resides within the set of measures $\overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\Gamma))^{\boxtimes d}\right)$. For convex SCSs, we prove this probability is ϵ-close to 1, assuming d is sufficiently large.

Proposition 33: Let $k \in \mathbb{N}$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a convex SCS. For any $\epsilon > 0$, there exists $d_0 \in \mathbb{N}$, such that for all $d \in \mathbb{N}$, $d \geqslant d_0$, $n \in \mathbb{N}$, $n \geqslant k+2$, there exists $A \subseteq F_n^d$, $|A| \leqslant \epsilon n^d$, such that for $y \in \Sigma^{F_n^d}$ drawn randomly using the measure $\mu^{n,d}$,

$$\mu^{n,d}\left(\eta_{y,A} \in \overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\Gamma))^{\boxtimes d}\right)\right) \geqslant 1-\epsilon.$$

Proof: Recall that by Definition 31, $\eta_{y,A}$ is a product measure, while $\mu_{y,A}$ is not necessarily so. Additionally, we contend that $\overline{\pi}_{[k]\mathbf{e}_i}(\mu_{y,A}) \in \Gamma$ for all $y \in \mathcal{B}_n(\Gamma^{\otimes d})$, $A \subseteq F_n^d$ and $i \in [d]$. Indeed,

$$\begin{aligned}\overline{\pi}_{[k]\mathbf{e}_i}(\mu_{y,A}) &= \frac{1}{|F_n^d|}\sum_{\mathbf{v}\in F_n^d}\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})\\ &= \frac{1}{|F_n^d|}\sum_{\mathbf{v}\in F_n^d}\frac{1}{|\pi_A^{-1}(y_A)|}\sum_{x\in\pi_A^{-1}(y_A)}\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\delta_{\hat{x}})\\ &= \frac{1}{|\pi_A^{-1}(y_A)|}\sum_{x\in\pi_A^{-1}(y_A)}\frac{1}{|F_n^d|}\sum_{\mathbf{v}\in F_n^d}\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\delta_{\hat{x}})\\ &= \frac{1}{|\pi_A^{-1}(y_A)|}\sum_{x\in\pi_A^{-1}(y_A)}\pi_{[k]\mathbf{e}_i}\left(\frac{1}{|F_n^d|}\sum_{\mathbf{v}\in F_n^d}\delta_{\sigma_{\mathbf{v}}(\hat{x})}\right)\\ &= \frac{1}{|\pi_A^{-1}(y_A)|}\sum_{x\in\pi_A^{-1}(y_A)}\mathrm{fr}_x^{[k]\mathbf{e}_i},\end{aligned}$$

where we recall that

$$\pi_A^{-1}(y_A) = \left\{x \in \mathcal{B}_n(\Gamma^{\otimes d}) \;:\; x_A = y_A\right\}.$$

Since $\mathrm{fr}_x^{[k]\mathbf{e}_i} \in \Gamma$ for every $x \in \pi_A^{-1}(y_A)$ and since Γ is convex the contention is proved. Additionally, by the convexity of Γ, $\overline{\pi}_{[k]\mathbf{e}_i}(\mu_{y,A}) \in \Gamma$ implies

$$\frac{1}{d}\sum_{i\in[d]}\overline{\pi}_{[k]\mathbf{e}_i}(\mu_{y,A}) \in \Gamma.$$

Draw $y \in \Sigma^{F_n^d}$ randomly using the measure $\mu^{n,d}$. For any $A \subseteq F_n^d$, let us denote

$$D_{A,y} \triangleq \left\|\frac{1}{d}\sum_{i\in[d]}\overline{\pi}_{[k]\mathbf{e}_i}(\eta_{y,A}) - \frac{1}{d}\sum_{i\in[d]}\overline{\pi}_{[k]\mathbf{e}_i}(\mu_{y,A})\right\|_{TV}.$$

We will use $E_y[\cdot]$ to denote expectation with respect to the random variable y which is randomly drawn using the measure $\mu^{n,d}$. Denote

$$D_A \triangleq E_y[D_{A,y}].$$

To prove the theorem, it suffices to show that for any $\epsilon > 0$, if d is large enough there exists $A \subseteq F_n^d$, $|A| \leqslant \epsilon n^d$, and with probability at least $1-\epsilon$ (with respect to $\mu^{n,d}$) we have $D_{A,y} \leqslant \epsilon$. By a standard application of the Markov inequality, it is sufficient to show that (under the above conditions) $D_A \leqslant \epsilon^2$.

By definition, for any $A \subseteq F_n^d$ and any $y \in \Sigma^{F_n^d}$ we have

$$D_{A,y} = \left\|\frac{1}{d|F_n^d|}\sum_{i\in[d]}\sum_{\mathbf{v}\in F_n^d}(\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A}) - \pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A}))\right\|_{TV}.$$

Applying the triangle inequality we obtain

$$D_{A,y} \leqslant \frac{1}{d|F_n^d|}\sum_{i\in[d]}\sum_{\mathbf{v}\in F_n^d}\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A}) - \pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})\right\|_{TV}.$$

Taking the expectation, E_y, on both sides and using its linearity we get

$$D_A \leqslant \frac{1}{d|F_n^d|}\sum_{i\in[d]}\sum_{\mathbf{v}\in F_n^d}E_y\left[\left\|\pi_{[k]\mathbf{e}_i+\mathbf{v}}(\eta_{y,A}) - \pi_{[k]\mathbf{e}_i+\mathbf{v}}(\mu_{y,A})\right\|_{TV}\right].$$

$$\sum_{(\mathbf{v},a)\in F_n^d\times\Sigma} E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]\cdot 1$$

$$\overset{\text{C.S}}{\leqslant}\sqrt{\left(\sum_{(\mathbf{v},a)\in F_n^d\times\Sigma}\left(E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]\right)^2\right)\left(\sum_{(\mathbf{v},a)\in F_n^d\times\Sigma}1^2\right)}$$

$$=\sqrt{\left(\sum_{(\mathbf{v},a)\in F_n^d\times\Sigma}\left(E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]\right)^2\right)\cdot\left|F_n^d\right|\cdot|\Sigma|}. \tag{21}$$

$$D_A\leqslant\frac{\sqrt{|\Sigma|}}{2d\sqrt{\left|F_n^d\right|}}\sum_{j\in[k]}\sum_{i\in[d]}\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}\left(E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]\right)^2}. \tag{22}$$

$$D_A\leqslant\frac{\sqrt{|\Sigma|}}{2d\sqrt{\left|F_n^d\right|}}\sum_{j\in[k]}\sum_{i\in[d]}\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}\left(E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right)^2\right]\right)}. \tag{23}$$

By Lemma 32 and the linearity of the expectation we obtain

$$D_A\leqslant\sum_{j\in[k]}\frac{1}{d|F_n^d|}\cdot\sum_{i\in[d]}\sum_{\mathbf{v}\in F_n^d}E_y\left[\left\|\pi_{\{\mathbf{v}\}}(\eta_{y,A})-\pi_{\{\mathbf{v}\}}(\mu_{y,A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})})\right\|_{TV}\right].$$

Consider another random variable $x\in\Sigma^{F_n^d}$, also randomly drawn using the measure $\mu^{n,d}$. Now define $f:\Sigma^{F_n^d}\to\{0,1\}^{F_n^d\times\Sigma}$ by

$$f(x)_{(\mathbf{v},a)}\triangleq\begin{cases}1 & x_{\mathbf{v}}=a,\\ 0 & \text{otherwise},\end{cases}$$

for all $\mathbf{v}\in F_n^d$ and $a\in\Sigma$. Thus, by definition we have that

$$\pi_{\{\mathbf{v}\}}(\eta_{y,A})(a)=\pi_{\{\mathbf{v}\}}(\mu_{y,A})(a)=E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y).$$

Since Σ is finite we can write the total variation distance as a sum, and then apply the triangle inequality, which results in

$$D_A\leqslant\frac{1}{2}\sum_{j\in[k]}\frac{1}{d|F_n^d|}\sum_{i\in[d]}\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]. \tag{20}$$

For any $j\in[k]$, viewing the expression

$$\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]$$

as an inner product of a vector in $\mathbb{R}^{F_n^d\times\Sigma}$ whose $(\mathbf{v},a)$'th coordinate is equal to

$$E_y\left[\left|E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_A\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{A\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right|\right]$$

and $\mathbf{1}$, we may apply Cauchy-Schwarz (C.S) inequality and obtain (21), as shown at the top of this page. Thus, combining (20) and (21) we have (22), as shown at the top of this page. Using the fact that $(E[|X|])^2\leqslant E[X^2]$ (again, by C.S), we have (23), as shown at the top of this page.

Choose m large enough such that $\frac{1}{\sqrt{m}}\leqslant\frac{\epsilon^2}{k\sqrt{|\Sigma|}}$ and denote $\epsilon_0=\frac{\epsilon^2}{k|\Sigma|}$. Now let $\mathbb{P},I_{t,\mathbf{v}},X_0,X_1,\ldots,X_m$ be as given by Lemma 29 with $n>k+2$ and with ϵ_0 and obtain d_0. From here on, assume $d\geqslant d_0$. Let $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}$.

First, from (23) we may bound D_{X_t}, for any $t\in[m+1]$, by

$$D_{X_t}\leqslant\frac{\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\sum_{j\in[k]}\frac{1}{d}\sum_{i\in[d]}\left(\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right)^2\right]\right)^{\frac{1}{2}}. \tag{24}$$

By the properties of X_t and X_{t+1} given in Lemma 29, for every $\mathbf{v}\in F_n^d$ there is a random variable $I_{t,\mathbf{v}}$ independent of X_t and distributed uniformly on $[d]$ so that $\mathbb{P}(X_t\cup([-(j+1)]\mathbf{e}_{I_{t,\mathbf{v}}}+\mathbf{v})\subseteq X_{t+1}\mid X_t)\geqslant 1-\epsilon_0$. Denote

$$X_{t,\mathbf{v}}\triangleq X_t\cup([-(j+1)]\mathbf{e}_{I_{t,\mathbf{v}}}+\mathbf{v}).$$

Since $I_{t,\mathbf{v}}$ is independent of X_t we have (25), as shown at the top of the next page.

From (24) and (25) we obtain (26), as shown at the top of the next page.

Since we may view $E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right]$ as the orthogonal projection of $f(x)_{(\mathbf{v},a)}$ on $L^2(F_n^d,\mathcal{F}_{X_t},\mu^{n,d},\mathbb{R})$, if $X_{t,\mathbf{v}}\subseteq X_{t+1}$ we have

$$E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t,\mathbf{v}}}\right](y)\right)^2\right]\leqslant E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t+1}}\right](y)\right)^2\right].$$

$$\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t,\mathbf{v}}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]$$
$$=\frac{1}{d}\sum_{i\in[d]}\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t\cup([-(j+1)]\mathbf{e}_i+\mathbf{v})}\right](y)\right)^2\right]}. \tag{25}$$

$$D_{X_t}\leqslant\frac{\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\sum_{j\in[k]}\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t,\mathbf{v}}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]. \tag{26}$$

$$\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t,\mathbf{v}}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]$$
$$\leqslant(1-\epsilon_0)\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t+1}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]+\epsilon_0\sqrt{|\Sigma|\left|F_n^d\right|}$$
$$\leqslant\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t+1}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]+\epsilon_0\sqrt{|\Sigma|\left|F_n^d\right|}. \tag{27}$$

$$D_{X_t}\leqslant\frac{\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\sum_{j\in[k]}\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t+1}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]$$
$$+\frac{\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\sum_{j\in[k]}\epsilon_0\sqrt{|\Sigma|\left|F_n^d\right|}$$
$$=\frac{k\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\mathbb{E}\left[\sqrt{\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t+1}}\right](y)\right)^2\right]}\;\middle|\;X_t\right]+\frac{k|\Sigma|\epsilon_0}{2}. \tag{28}$$

Otherwise, if $X_{t,\mathbf{v}}\nsubseteq X_{t+1}$, then

$$E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t,\mathbf{v}}}\right](y)\right)^2\right]\leqslant 1.$$

By the properties of X_t given in Lemma 29, we have that

$$\mathbb{P}(X_{t,\mathbf{v}}\subseteq X_{t+1}\mid X_t)\geqslant 1-\epsilon_0.$$

Thus, we have (27), as shown at the top of this page.

From (26) and (27) we obtain (28), as shown at the top of this page.

Observe that viewing f as a random variable with respect to $\mu^{n,d}$ we have $\|f\|_2=\sqrt{\left|F_n^d\right|}$. Note also that

$$\left\|E\left[f\mid\mathcal{F}_{X_t}\right]-E\left[f\mid\mathcal{F}_{X_{t+1}}\right]\right\|_2$$
$$=\left(\sum_{\mathbf{v}\in F_n^d}\sum_{a\in\Sigma}E_y\left[\left(E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_t}\right](y)-E_x\left[f(x)_{(\mathbf{v},a)}\mid\mathcal{F}_{X_{t+1}}\right](y)\right)^2\right]\right)^{\frac{1}{2}}. \tag{29}$$

From (28) and (29) we obtain that for every $t\in[m]$,

$$D_{X_t}\leqslant\frac{k\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\mathbb{E}\left[\left\|E[f\mid\mathcal{F}_{X_t}]-E[f\mid\mathcal{F}_{X_{t+1}}]\right\|_2\;\middle|\;X_t\right]+\frac{k|\Sigma|\epsilon_0}{2}. \tag{30}$$

Note that the probability that a random variable is greater than or equal to its expectation is always strictly positive. Because X_t takes only finitely many values, this means that for every $t\in[m]$, for every realization of X_t, denoted as χ_t, there exists a realization of X_{t+1}, denoted as $\chi_{t+1}=\chi_{t+1}(\chi_t)$ such that $\mathbb{P}\left(X_{t+1}=\chi_{t+1}\mid X_t\right)>0$ and

$$\mathbb{E}\left[\left\|E\left[f\mid\mathcal{F}_{X_t}\right]-E\left[f\mid\mathcal{F}_{X_{t+1}}\right]\right\|_2\mid X_t\right]\leqslant\left\|E\left[f\mid\mathcal{F}_{\chi_t}\right]-E\left[f\mid\mathcal{F}_{\chi_{t+1}}\right]\right\|_2.$$

Together with (30) we obtain

$$D_{\chi_t}\leqslant\frac{k\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\left\|E\left[f\mid\mathcal{F}_{\chi_t}\right]-E\left[f\mid\mathcal{F}_{\chi_{t+1}}\right]\right\|_2+\frac{k|\Sigma|\epsilon_0}{2}. \tag{31}$$

Since (31) holds for every t, we obtain that there exists a sequence $(\chi_t)_{t\in[m+1]}$ of realizations of $(X_t)_{t\in[m+1]}$ with positive probabilities, such that for every $t \in [m]$,

$$D_{\chi_t} \leqslant \frac{k\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}} \left\| E\left[f \mid \mathcal{F}_{\chi_t}\right] - E\left[f \mid \mathcal{F}_{\chi_{t+1}}\right]\right\|_2 + \frac{k\,|\Sigma|\,\epsilon_0}{2}. \tag{32}$$

From Lemma 28, there exists $t \in [m]$ such that

$$\left\| E\left[f \mid \mathcal{F}_{\chi_t}\right] - E\left[f \mid \mathcal{F}_{\chi_{t+1}}\right]\right\|_2^2 \leqslant \frac{1}{m}\,\|f\|_2^2. \tag{33}$$

Combining (33) with (32) we obtain that there exists $t \in [m]$ such that

$$\begin{aligned} D_{\chi_t} &\leqslant \frac{k\sqrt{|\Sigma|}}{2\sqrt{\left|F_n^d\right|}}\frac{1}{\sqrt{m}}\|f\|_2 + \frac{k\,|\Sigma|\,\epsilon_0}{2} \\ &= \frac{k\sqrt{|\Sigma|}}{2\sqrt{m}} + \frac{k\,|\Sigma|\,\epsilon_0}{2}. \end{aligned}$$

Taking $A = \chi_t$, and recalling our choice of $\epsilon_0 = \frac{\epsilon^2}{k|\Sigma|}$ and $\frac{1}{\sqrt{m}} \leqslant \frac{\epsilon^2}{k\sqrt{|\Sigma|}}$, we obtain that

$$D_A \leqslant \frac{k\sqrt{|\Sigma|}}{2\sqrt{m}} + \frac{k\,|\Sigma|\,\epsilon_0}{2} \leqslant \epsilon^2,$$

which completes the proof. ∎

We have reached the main result of this section. We show that capacity of a convex d-axial product is arbitrarily close to the independence entropy, as the dimension grows.

Theorem 34: Let $k \in \mathbb{N}$, and let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a convex one-dimensional SCS. Then

$$\limsup_{d\to\infty} \mathsf{cap}(\Gamma^{\otimes d}) = h_{\mathrm{ind}}(\Gamma).$$

Proof: First note that $\limsup_{d\to\infty} \mathsf{cap}(\Gamma^{\otimes d}) \geqslant h_{\mathrm{ind}}(\Gamma)$ by applying Theorem 26 to $\Gamma^{\otimes d}$ for every d and taking $d \to \infty$ on both sides. For the other direction, fix $\epsilon_0 > 0$ and choose

$$0 < \epsilon < \min\left\{\frac{\epsilon_0}{2\log_2|\Sigma|}, 1\right\}, \quad 0 < \delta < \frac{\epsilon}{2}.$$

Replace Γ by $\mathbb{B}_\delta(\Gamma)$ in Definition 31 and denote the resulting measures by $\mu_\delta^{n,d}$, $\mu_{y,A}^\delta$, and $\eta_{y,A}^\delta$.

Recall that for a measure μ and a σ-algebra $\mathcal{F}$,

$$H(\mu \mid \mathcal{F}) \triangleq E\left[H(\mu(\cdot \mid \mathcal{F}))\right] = \int H(\mu(\cdot \mid \mathcal{F}))(x)\mathrm{d}\mu(x). \tag{34}$$

In other words, $H(\mu \mid \mathcal{F})$ is the expected entropy of the conditional measure $\mu(\cdot \mid \mathcal{F})$. Also recall that for $A \subseteq F_n^d$, $\pi_A(\mu_\delta^{n,d})$ denotes the A-marginal of $\mu_\delta^{n,d}$, and that $\mathcal{F}_A$ denotes the σ-algebra generated by the coordinates in A. We have that

$$H(\mu_\delta^{n,d}) = H(\pi_A(\mu_\delta^{n,d})) + H\left(\mu_\delta^{n,d} \,\middle|\, \mathcal{F}_A\right).$$

By Proposition 33, for any $n \in \mathbb{N}$, $n \geqslant k+2$, there exists $d_0 \in \mathbb{N}$, such that for every $d \geqslant d_0$, there exists $A \subseteq \Sigma^{F_n^d}$, $|A| \leqslant \epsilon n^d$, such that,

$$\mu_\delta^{n,d}\left(\eta_{y,A}^\delta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)\right) \geqslant 1-\epsilon > 0.$$

In particular, there exists a word $y \in \Sigma^{F_n^d}$ such that $\eta_{y,A} \in \overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon\ (\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)$. Since clearly

$$H(\pi_A(\mu_\delta^{n,d})) \leqslant \log_2\left|\Sigma^A\right|,$$

by combining the above we have

$$H(\mu_\delta^{n,d}) \leqslant H\left(\mu_\delta^{n,d} \,\middle|\, \mathcal{F}_A\right) + \epsilon n^d \log_2|\Sigma|. \tag{35}$$

Because the joint entropy of a finite set of random variables is bounded from above by the sum of their entropies (and the same statement holds for conditional entropy), we have:

$$H\left(\mu_\delta^{n,d} \,\middle|\, \mathcal{F}_A\right) \leqslant \sum_{\mathbf{v}\in F_n^d} H\left(\pi_{\{\mathbf{v}\}}(\mu_\delta^{n,d}) \,\middle|\, \mathcal{F}_A\right).$$

By definition of the random measure $\eta_{y,A}^\delta$ and from (34), we have

$$H\left(\pi_{\{\mathbf{v}\}}(\mu_\delta^{n,d}) \,\middle|\, \mathcal{F}_A\right) = \sum_{y\in\Sigma^{F_n^d}} H\left(\pi_{\{\mathbf{v}\}}(\eta_{y,A}^\delta)\right)\mu_\delta^{n,d}(y).$$

Thus,

$$H\left(\mu_\delta^{n,d} \,\middle|\, \mathcal{F}_A\right) \leqslant \sum_{\mathbf{v}\in F_n^d}\sum_{y\in\Sigma^{F_n^d}} H\left(\pi_{\{\mathbf{v}\}}(\eta_{y,A}^\delta)\right)\mu_\delta^{n,d}(y).$$

Now, since $\eta_{y,A}$ is a product measure, we have

$$H(\eta_{y,A}^\delta) = \sum_{\mathbf{v}\in F_n^d} H\left(\pi_{\{\mathbf{v}\}}(\eta_{y,A}^\delta)\right).$$

It follows that,

$$H\left(\mu_\delta^{n,d} \,\middle|\, \mathcal{F}_A\right) \leqslant \sum_{y\in\Sigma^{F_n^d}} H\left(\eta_{y,A}^\delta\right)\mu_\delta^{n,d}(y). \tag{36}$$

Let us conveniently use p to denote the value

$$p \triangleq \mu_\delta^{n,d}\left(\eta_{y,A}^\delta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)\right),$$

and recall that $p \geqslant 1-\epsilon > 0$. Then

$$\begin{aligned} &\sum_{y\in\Sigma^{F_n^d}} H(\eta_{y,A}^\delta)\mu_\delta^{n,d}(y) \\ &\quad\leqslant p \cdot \sup_{\eta\in\overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)} H(\eta) + (1-p)\cdot\log_2\left|\Sigma^{F_n^d}\right|. \end{aligned} \tag{37}$$

Using the fact that $p \geqslant 1-\epsilon > 0$ combined with (36) and (37), it follows that

$$\begin{aligned} &H\left(\mu_\delta^{n,d} \,\middle|\, \mathcal{F}_A\right) \\ &\quad\leqslant p \cdot \sup_{\eta\in\overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)} H(\eta) + (1-p)n^d\cdot\log_2|\Sigma| \\ &\quad\leqslant \sup_{\eta\in\overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)} H(\eta) + \epsilon n^d\log_2|\Sigma|. \end{aligned} \tag{38}$$

Combining (38) with (35) we obtain

$$\frac{1}{n^d}H\left(\mu_\delta^{n,d}\right) \leqslant \frac{1}{n^d}\sup_{\eta\in\overline{\mathcal{P}}_n\left((\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\boxtimes d}\right)} H(\eta) + 2\epsilon\log_2|\Sigma|.$$

By our choice of ϵ, we have $\epsilon + \delta \leqslant \epsilon_0$, hence $(\mathbb{B}_\epsilon(\mathbb{B}_\delta(\Gamma)))^{\otimes d} \subseteq (\mathbb{B}_{\epsilon_0}(\Gamma))^{\otimes d}$, as well as

$$\frac{1}{n^d} H\left(\mu_\delta^{n,d}\right) \leqslant \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))^{\boxtimes d}\right)} H(\eta) + \epsilon_0. \quad (39)$$

Since $\mu_\delta^{n,d}$ is the uniform measure on $\mathcal{B}_n(\mathbb{B}_\delta(\Gamma)^{\otimes d})$,

$$H(\mu_\delta^{n,d}) = \log_2 \left|\mathcal{B}_n(\mathbb{B}_\delta(\Gamma)^{\otimes d})\right|.$$

Thus,

$$\frac{1}{n^d} \log_2 \left|\mathcal{B}_n\left((\mathbb{B}_\delta(\Gamma))^{\otimes d}\right)\right| \leqslant \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))^{\boxtimes d}\right)} H(\eta) + \epsilon_0.$$

Taking $\limsup_{n\to\infty}$ we obtain

$$\widehat{\mathsf{cap}}\left((\mathbb{B}_\delta(\Gamma))^{\otimes d}\right) \leqslant \limsup_{n\to\infty} \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))^{\boxtimes d}\right)} H(\eta) + \epsilon_0.$$

Since

$$\mathbb{B}_\delta\left(\Gamma^{\otimes d}\right) \subseteq (\mathbb{B}_\delta(\Gamma))^{\otimes d},$$

we have

$$\widehat{\mathsf{cap}}\left(\mathbb{B}_\delta\left((\Gamma)^{\otimes d}\right)\right) \leqslant \widehat{\mathsf{cap}}\left((\mathbb{B}_\delta(\Gamma))^{\otimes d}\right) \leqslant \limsup_{n\to\infty} \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))^{\boxtimes d}\right)} H(\eta) + \epsilon_0.$$

Taking $\lim_{\delta\to 0^+}$, we get

$$\mathsf{cap}\left(\Gamma^{\otimes d}\right) \leqslant \limsup_{n\to\infty} \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))^{\boxtimes d}\right)} H(\eta) + \epsilon_0. \quad (40)$$

At this point we take a slight detour. For $\xi > 0$, $\mathbb{B}_{\epsilon_0}(\Gamma) \subseteq \mathbb{B}_\xi\left(\mathbb{B}_{\epsilon_0}(\Gamma)\right)$ and hence we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))^{\boxtimes d}\right)} H(\eta) + \epsilon_0 \\
&\quad \leqslant \limsup_{\xi\to 0^+} \limsup_{n\to\infty} \frac{1}{n^d} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_\xi(\mathbb{B}_{\epsilon_0}(\Gamma)))^{\boxtimes d}\right)} H(\eta) + \epsilon_0 \\
&\quad \overset{(a)}{=} h_{\text{ind}}\left(\left(\mathbb{B}_{\epsilon_0}(\Gamma)\right)^{\boxtimes d}\right) + \epsilon_0 \\
&\quad \overset{(b)}{=} h_{\text{ind}}\left(\left(\mathbb{B}_{\epsilon_0}(\Gamma)\right)\right) + \epsilon_0,
\end{aligned}$$

where (a) follows by definition, and (b) follows by Lemma 27. Substituting this in (40) and taking $d \to \infty$ we obtain

$$\limsup_{d\to\infty} \mathsf{cap}\left(\Gamma^{\otimes d}\right) \leqslant h_{\text{ind}}\left(\left(\mathbb{B}_{\epsilon_0}(\Gamma)\right)\right) + \epsilon_0. \quad (41)$$

Note that since Γ is convex we have that for $\epsilon_1 > 0$, $\mathbb{B}_{\epsilon_1}\left(\mathbb{B}_{\epsilon_0}(\Gamma)\right) = \mathbb{B}_{\epsilon_1+\epsilon_0}(\Gamma)$. Therefore, by the definition of limit we have

$$\begin{aligned}
&\limsup_{\epsilon_0\to 0^+} \limsup_{\epsilon_1\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_1}(\mathbb{B}_{\epsilon_0}(\Gamma)))\right)} H(\eta) \\
&\quad = \limsup_{\epsilon_0\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \sup_{\eta \in \overline{\mathcal{P}}_n\left((\mathbb{B}_{\epsilon_0}(\Gamma))\right)} H(\eta).
\end{aligned}$$

Therefore, taking the limit as $\epsilon_0 \to 0$ in (41) we obtain

$$\limsup_{d\to\infty} \mathsf{cap}\left(\Gamma^{\otimes d}\right) \leqslant h_{\text{ind}}(\Gamma).$$

■

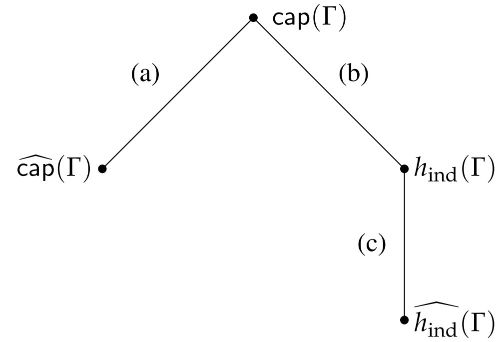

Fig. 1. The Hasse diagram for a general d-dimensional SCS $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$, where (a) follows from (1), (b) follows from Theorem 26, and (c) follows from (2).

VI. Discussion

Our initial motivation behind this work is to approximate the capacity of multidimensional SCSs using "meaningful" expressions. The main challenges were defining exactly what is the capacity of multidimensional SCSs, and obtaining the connections between the capacity and the independence entropy. The type of semiconstrained systems and constrained systems considered in this paper correspond to a class of subshifts called "subshifts of finite type". More general systems, such as sofic shifts also appear in the context of constrained coding. We point out that the main results of [27]–[29] apply to more general subshifts. Thus, Theorem 34 only generalizes the main result of [28] for subshifts of finite type.

At the core of our results, for $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ and its axial product $\Gamma^{\otimes d}$, by Theorem 24 and Theorem 26 it follows that

$$h_{\text{ind}}(\Gamma) \leqslant \mathsf{cap}(\Gamma^{\otimes d}).$$

Thus, the problem of bounding the capacity of a d-dimensional axial-product SCS is simplified by having to consider only product measures, which are much easier to handle. Moreover, any number of dimensions d, may be reduced via this bound to the one-dimensional case. This bound is asymptotically tight, as together with Theorem 34, for convex Γ,

$$\limsup_{d\to\infty} \mathsf{cap}(\Gamma^{\otimes d}) = h_{\text{ind}}(\Gamma).$$

It also appears that the capacity cap, and independence entropy h_{ind}, are robust generalizations of their one-dimensional combinatorial counterparts.

The paper contains many connections between the various capacities and entropies. Figure 1 shows the Hasse diagram[2] for the bounds pertaining to general d-dimensional $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$. In the case of a convex one-dimensional $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ and its d-axial-product SCS $\Gamma^{\otimes d}$, a more elaborate Hasse diagram emerges, which is shown in Figure 2.

We note here that following the same arguments used in the proof of Theorem 24 would show that $\mathsf{cap}(\Gamma^{\otimes d}) \geqslant \mathsf{cap}(\Gamma^{\otimes d+1})$ which means that in $\limsup_{d\to\infty} \mathsf{cap}(\Gamma^{\otimes d})$ the limit actually exists and equals to $\inf_d \mathsf{cap}(\Gamma^{\otimes d})$.

[2] In a Hasse diagram expressions are represented by nodes. An edge between two nodes represents an inequality where the higher node has a higher value.

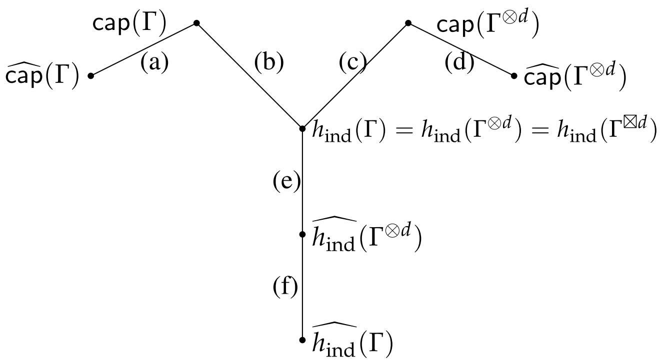

Fig. 2. The Hasse diagram for a convex one-dimensional $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ and its d-axial-product SCS $\Gamma^{\otimes d}$, where (a) and (d) follow from (1), (b) and (c) follow from Theorem 26, (e) follows from (2), and (f) follows from Lemma 23.

We would also like to compare our results, as they apply to a specific case study described in [1]. Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a convex one-dimensional SCS, and recall that the axial product $\Gamma^{\otimes d}$ is defined as

$$\Gamma^{\otimes d} \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{F_k^d}) \;:\; \forall i \in [d],\; \pi_{[k]\mathbf{e}_i}(\mu) \in \Gamma\right\},$$

and thus

$$\mathcal{B}_n\left(\Gamma^{\otimes d}\right) = \left\{w \in F_n^d \;:\; \forall i \in [d],\; \mathrm{fr}_w^{[k]\mathbf{e}_i} \in \Gamma\right\}.$$

The SCSs studied in [1] were an averaged version of the axial product, namely,

$$\Gamma^{\boxtimes d} \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{F_k^d}) \;:\; \frac{1}{d}\sum_{i\in[d]} \pi_{[k]\mathbf{e}_i}(\mu) \in \Gamma\right\},$$

and thus

$$\mathcal{B}_n\left(\Gamma^{\boxtimes d}\right) = \left\{w \in F_n^d \;:\; \frac{1}{d}\sum_{i\in[d]} \mathrm{fr}_w^{[k]\mathbf{e}_i} \in \Gamma\right\}.$$

By convexity, it easily follows that

$$\mathcal{B}_n\left(\Gamma^{\otimes d}\right) \subseteq \mathcal{B}_n\left(\Gamma^{\boxtimes d}\right),$$

and thus

$$\mathsf{cap}(\Gamma^{\otimes d}) \leqslant \mathsf{cap}(\Gamma^{\boxtimes d}).$$

We now focus on the simple example known as the $(0, k, p)$-RLL SCS over the binary alphabet $\Sigma = \{0, 1\}$, which was the case study of [1]. The one-dimensional $(0, k, p)$-RLL SCS, $0 \leqslant p \leqslant 1$, is defined by

$$\Gamma_{k,p} \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{k+1}) \;:\; \mu(1^{k+1}) \leqslant p\right\}, \tag{42}$$

where 1^{k+1} denotes the all-ones string of length $k+1$. This example is a generalization of the well known inverted $(0, k)$-RLL fully constrained system, since if we take $p = 0$ we obtain the inverted $(0, k)$-RLL. In [1], the authors found lower and upper bounds on the internal capacity of $\Gamma_{k,p}^{\boxtimes d}$. We recall the relevant lower bound here.

Theorem 35: [1, Th. 20] Let $\Gamma_{k,p}$ denote the one-dimensional $(0, k, p)$-RLL SCS given in (42). Then, for all $0 \leqslant p \leqslant \frac{1}{2^{k+1}}$,

$$\widehat{\mathsf{cap}}(\Gamma_{k,p}^{\boxtimes d}) \geqslant 1 + d\left(\widehat{\mathsf{cap}}(\Gamma) - 1\right),$$

whereas for all $\frac{1}{2^{k+1}} \leqslant p \leqslant 1$, $\widehat{\mathsf{cap}}(\Gamma_{k,p}^{\boxtimes d}) = 1$.

We first note that this theorem implies a lower bound on $\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d})$,

$$\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d}) \geqslant \widehat{\mathsf{cap}}(\Gamma_{k,p}^{\boxtimes d}) \geqslant 1 + d\left(\widehat{\mathsf{cap}}(\Gamma) - 1\right).$$

The lower bound of [1] eventually becomes negative, as the dimension d grows, and therefore, degenerate. However, using the results of this paper,

$$\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d}) \geqslant \widehat{h_{\text{ind}}}(\Gamma_{k,p}),$$

and this bound does not depend on the dimension, and therefore, does not degenerate. We provide an explicit numerical example:

Example 36: Let us take $k = 2$, and $p = 0.05$, meaning that we restrict the frequency of the pattern 111 to be at most 0.05. Fix $d = 3$. The lower bound on $\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d})$ from [1] uses $\widehat{\mathsf{cap}}(\Gamma_{k,p})$. The latter can be calculated by solving an optimization problem using a computer. We obtain that $\widehat{\mathsf{cap}}(\Gamma_{k,p}) \approx 0.976$ which means that

$$\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d}) \geqslant 1 + 3 \cdot (0.976 - 1) \approx 0.928.$$

Using the results of this paper, we use $\widehat{h_{\text{ind}}}(\Gamma_{k,p})$ as a lower bound to $\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d})$. Finding the supremum involved in the definition of $\widehat{h_{\text{ind}}}(\Gamma_{k,p})$ is also not easy, and we lower bound it by guessing a specific measure. We take each coordinate to be i.i.d. Bernoulli $\sqrt[3]{0.05}$, and we get

$$\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d}) \geqslant \widehat{h_{\text{ind}}}(\Gamma_{k,p}) \geqslant H_2(\sqrt[3]{0.05}) \approx 0.949,$$

which is a better lower bound than that of [1]. Note that the upper bound gives $\widehat{\mathsf{cap}}(\Gamma') \leqslant 0.983$. We further mention that the lower bound of [1] gets increasingly worse as the dimension grows. For example, when $d = 10$ we obtain by Theorem 35 that $\mathsf{cap}(\Gamma_{k,p}^{\boxtimes d}) \geqslant 0.76$ whereas using the independence entropy, the bound stays the same, i.e., $\mathsf{cap}(\Gamma^{\boxtimes d}) \geqslant 0.949$. Finally, for all $d \geqslant 42$, the lower bound of [1] becomes degenerate. □

We present another example for $(0, 1, p)$ with a more elaborate lower bound.

Example 37: Take $k = 1$ and consider $\Gamma_{k,p}$. From the results of this paper,

$$\begin{aligned}\limsup_{d\to\infty} \mathsf{cap}(\Gamma_{1,p}^{\boxtimes d}) &\geqslant \limsup_{d\to\infty} \mathsf{cap}(\Gamma_{1,p}^{\otimes d}) \\ &= h_{\text{ind}}(\Gamma_{1,p}) \\ &\geqslant \widehat{h_{\text{ind}}}(\Gamma_{1,p}).\end{aligned}$$

We lower bound $\widehat{h_{\text{ind}}}(\Gamma_{1,p})$ by devising a product measure $\mu_{2n} \in (\mathcal{P}(\Sigma))^{2n}$, for all $n \in \mathbb{N}$. The measures use two parameters $0 \leqslant x, y \leqslant 1$, using a Bernoulli$(x)$ *distribution*

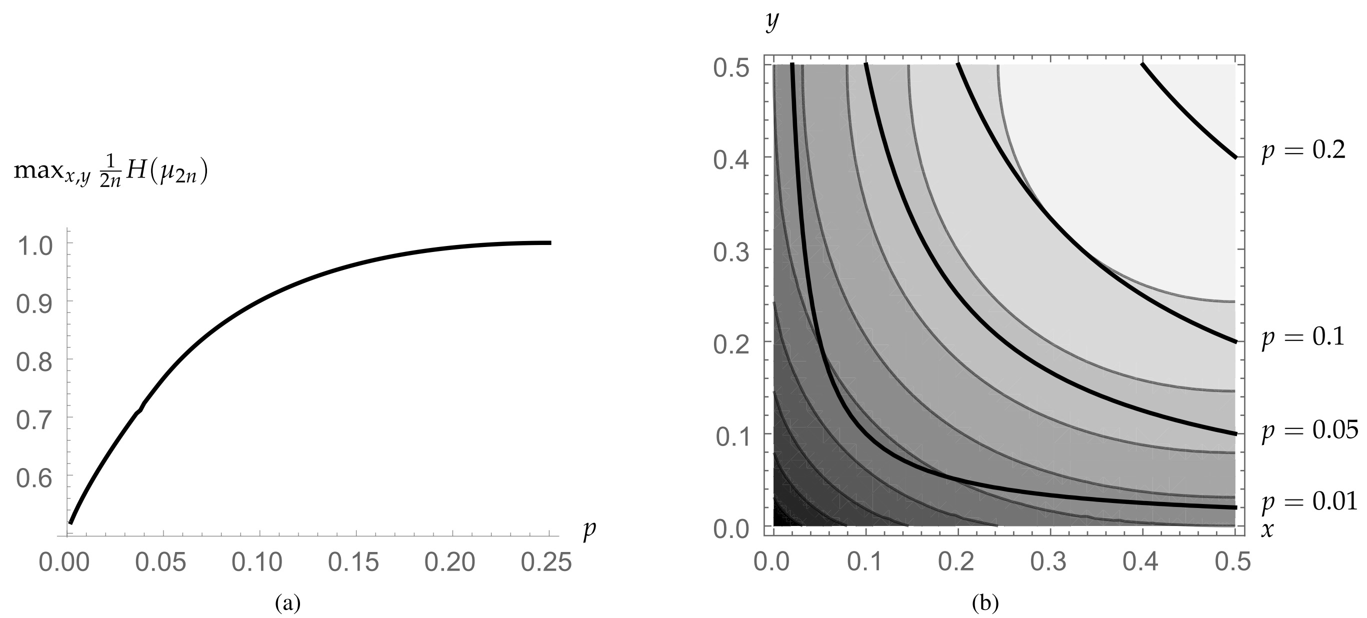

Fig. 3. A lower bound on $\limsup_{d\to\infty} \mathsf{cap}(\Gamma_{1,p}^{\boxtimes d})$ is shown in (a), where (b) shows a contour plot of $\frac{1}{2}(H_2(x)+H_2(y))$ as well as the curves $xy = p$ for $p = 0.01, 0.05, 0.1, 0.2$.

for positions with odd indices, and a Bernoulli(y) *for positions with even indices. Thus,*

$$\widehat{h_{\mathrm{ind}}}(\Gamma_{1,p}) \geqslant \max_{x,y} \frac{1}{2n} H(\mu_{2n})$$
$$= \max\left\{\frac{1}{2}(H_2(x)+H_2(y)) \;:\; 0 \leqslant x, y \leqslant 1,\; xy \leqslant p\right\}.$$

Due to monotonicity, the maximization problem always has a solution on the curve $xy = p$, which in the high range is unique $x = y = \sqrt{p}$, and in the lower range has two symmetric solutions. For example, for $p = 0.2$ the optimal solution is $x = y = \sqrt{0.2}$. However, for $p = 0.01$, the first optimal solution is $x \approx 0.454$, $y \approx 0.022$, and the symmetric solution is $x \approx 0.022$, $y \approx 0.454$. This is depicted in Figure 3. The transition between the one-solution regime and two-solution regime occurs exactly when

$$1 + (1-\sqrt{p}) \ln \frac{\sqrt{p}}{1-\sqrt{p}} = 0, \quad p \in (0, 0.25)$$

which is approximately $p \approx 0.04744$.

We note that this bound agrees with the solution for the fully constrained case, $\limsup_{d\to\infty} \mathsf{cap}(\Gamma_{k,0}^{\boxtimes d}) = \frac{1}{2}$ which was solved in [28]. We conjecture that Figure 3(a) indeed shows the exact limiting capacity. □

Appendix A
Cyclic and Non-Cyclic Capacities

The goal of this appendix is to show that the capacity, as we defined it cyclically, equals the (traditionally non-cyclic) capacity in the case of fully constrained systems. The results of this appendix together with Appendix VI imply that Theorem 34 recovers the main result of [28] about equality of limiting entropy and independence entropy, for the class of subshifts of finite type.

Definition 38: Let $d, k \in \mathbb{N}$. An element of a subclass of (traditional) fully constrained systems*, called shifts of finite type, is defined by a set $\Phi \subseteq \Sigma^{F_k^d}$ of d-dimensional words, called* forbidden patterns*. The set of all* admissible *words in $\Sigma^{F_n^d}$ is defined as*

$$\mathcal{B}_n^{\mathrm{com}}(\Phi) \triangleq \left\{x \in \Sigma^{F_n^d} \;:\; \forall \mathbf{v} \in F_{n-k}^d,\; x_{\mathbf{v}+F_k^d} \notin \Phi\right\}.$$

The (combinatorial) capacity *of Φ is defined by*

$$\mathsf{cap}^{\mathrm{com}}(\Phi) \triangleq \limsup_{n\to\infty} \frac{1}{|F_n^d|} \log_2 \left|\mathcal{B}_n^{\mathrm{com}}(\Phi)\right|.$$

□

Intuitively, a traditional fully constrained system is a set of words that do not contain any forbidden pattern *non-cyclically*. Given a (traditional) fully constrained system $\Phi \subseteq \Sigma^{F_k^d}$, we can construct a set of measures Γ_Φ defined as follows,

$$\Gamma_\Phi \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{F_k^d}) \;:\; \mu(\Sigma^{F_k^d} \setminus \Phi) = 1\right\}. \tag{43}$$

Thus, Γ_Φ is a SCS which is fully constrained in the sense of Definition 10. Since Definition 10 is more restrictive, by requiring forbidden patterns to not appear in admissible words *cyclically*, we immediately have

$$\mathcal{B}_n(\Gamma_\Phi) \subseteq \mathcal{B}_n^{\mathrm{com}}(\Phi),$$

implying also

$$\widehat{\mathsf{cap}}(\Gamma_\Phi) \leqslant \mathsf{cap}^{\mathrm{com}}(\Phi).$$

However, we now prove that the capacity of Γ_Φ does equal the (combinatorial) capacity of Φ.

Proposition 39: Let $d, k \in \mathbb{N}$. Let $\Phi \subseteq \Sigma^{F_k^d}$ be a fully constrained system as in Definition 38, and let $\Gamma_\Phi \subseteq \mathcal{P}(\Sigma^{F_k^d})$ be its corresponding fully constrained system as in Definition 10. If $\mathcal{B}_n^{\mathrm{com}}(\Phi) \neq \emptyset$ for all large enough $n \in \mathbb{N}$, then

$$\mathsf{cap}(\Gamma_\Phi) = \mathsf{cap}^{\mathrm{com}}(\Phi).$$

Proof: We first show that $\mathsf{cap}^{\mathrm{com}}(\Phi) \leqslant \mathsf{cap}(\Gamma_\Phi)$. Fix $\epsilon > 0$, and for $n \in \mathbb{N}$, $n \geqslant k$, consider the k-boundary of F_n^d which is defined as $F_n^d \setminus F_{n-k}^d$. Note that $|F_n^d \setminus F_{n-k}^d| = n^d - (n-k)^d$. Let $w \in \mathcal{B}_n^{\mathrm{com}}(\Phi)$. While w does not contain any forbidden pattern when considering the coordinates non-cyclically, it may contain some when considering the coordinates cyclically. The number of occurrences of forbidden patterns (cyclically) in w is at most $|F_n^d \setminus F_{n-k}^d| = n^d-(n-k)^d$. For all large enough n we have $\frac{n^d-(n-k)^d}{n^d} \leqslant \epsilon$, hence

$$\mathcal{B}_n^{\mathrm{com}}(\Phi) \subseteq \mathcal{B}_n(\mathbb{B}_\epsilon(\Gamma_\Phi)).$$

Thus, for every $\epsilon > 0$,

$$\mathsf{cap}^{\mathrm{com}}(\Phi) \leqslant \widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma_\Phi)).$$

Taking the limit as $\epsilon \to 0$ we obtain

$$\mathsf{cap}^{\mathrm{com}}(\Phi) \leqslant \mathsf{cap}(\Gamma_\Phi).$$

In the other direction, we now show that $\mathsf{cap}(\Gamma_\Phi) \leqslant \mathsf{cap}^{\mathrm{com}}(\Phi)$. Let $\delta_0 > 0$ and take $n_0 \in \mathbb{N}$ large enough such that

$$\frac{1}{n_0^d}\log_2\left|\mathcal{B}_{n_0}^{\mathrm{com}}(\Phi)\right| \leqslant \mathsf{cap}^{\mathrm{com}}(\Phi) + \frac{1}{3}\delta_0.$$

Denote the number of forbidden patterns by $t \triangleq |\Phi|$. Take $\delta > 0$ small enough such that both

$$\frac{t(1+\delta)}{n_0^d}H_2\left(\frac{\delta}{1+\delta}\right) \leqslant \frac{1}{3}\delta_0, \quad \text{and } t\delta\log_2|\Sigma| \leqslant \frac{1}{3}\delta_0,$$

where $H_2(\cdot)$ is the binary entropy function. Finally, for every $n \geqslant n_0$, denote $m \triangleq \lfloor n/n_0 \rfloor$, and choose any $0 < \epsilon \leqslant \delta/n_0^d$.

Consider a word $w \in \mathcal{B}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))$. We say w is made up a concatenation of m^d $F_{n_0}^d$-blocks, namely, a block is a set of positions $n_0\mathbf{v} + F_{n_0}^d$, where $\mathbf{v} \in F_m^d$, as well a boundary, namely, the set of positions $F_n^d \setminus F_{mn_0}^d$. By our choice of parameters, the number of occurrences (perhaps cyclically) of any forbidden pattern from Φ is at most

$$\epsilon|F_n^d| \leqslant \epsilon\ (m+1)^d n_0^d \leqslant \delta\ (m+1)^d.$$

This serves also as an upper bound on the number of blocks fully containing (non-cyclically) this forbidden pattern. Since there are t forbidden patterns, the number of blocks that are devoid (non-cyclically) of any forbidden pattern, is at least $m^d - t\delta(m+1)^d$. Such blocks are in fact words from $\mathcal{B}_{n_0}^{\mathrm{com}}(\Phi)$.

Fixing a specific forbidden pattern of length n_0^d, and considering each occurrence of it as a ball, we have at most $\delta(m+1)^d$ balls, which we throw into m^d+1 bins (m^d blocks, and another "virtual" bin for patterns that are not fully contained within a single block). The total number of ways to throw these balls into bins is at most $\binom{m^d+1+\delta(m+1)^d}{\delta(m+1)^d}$. Raising this to the power of t gives an upper bound on the number of ways the t forbidden patterns are dispersed among the blocks. In total we have,

$$\begin{aligned}
&|\mathcal{B}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))| \\
&\quad\leqslant \binom{m^d+1+\delta(m+1)^d}{\delta(m+1)^d}^t \\
&\qquad\cdot \left|\mathcal{B}_{n_0}^{\mathrm{com}}(\Phi)\right|^{m^d - t(m+1)^d\delta} |\Sigma|^{t\delta(m+1)^d n_0^d} |\Sigma|^{n^d-(mn_0)^d} \\
&\quad\leqslant \binom{m^d+1+\delta(m+1)^d}{\delta(m+1)^d}^t \left|\mathcal{B}_{n_0}^{\mathrm{com}}(\Phi)\right|^{m^d} \\
&\qquad\cdot |\Sigma|^{t\delta(m+1)^d n_0^d} |\Sigma|^{n^d-(mn_0)^d},
\end{aligned}$$

where the binomial coefficient follows from upper bounding the way forbidden patterns are dispersed among blocks, the following term counts the number of ways to fill blocks that do not contain (non-cyclically) any forbidden word, and the last term counts the ways to arbitrarily fill in the rest of the positions.

We now recall the well known bounds on the binomial coefficient (e.g., see [35, Lemma 7, p. 309]),

$$\frac{1}{\sqrt{8n\lambda(1-\lambda)}}2^{nH_2(\lambda)} \leqslant \binom{n}{\lambda n} \leqslant \frac{1}{\sqrt{2n\lambda(1-\lambda)}}2^{nH_2(\lambda)},$$

for all $n \in \mathbb{N}$, $0 < \lambda < 1$, and λn an integer. Thus,

$$\begin{aligned}
&\widehat{\mathsf{cap}}(\mathbb{B}_\epsilon(\Gamma_\Phi)) \\
&\quad= \limsup_{n\to\infty}\frac{1}{n^d}\log_2|\mathcal{B}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))| \\
&\quad\leqslant \frac{t(1+\delta)}{n_0^d}H_2\left(\frac{\delta}{1+\delta}\right) + \frac{1}{n_0^d}\log_2\left|\mathcal{B}_{n_0}^{\mathrm{com}}(\Phi)\right| + t\delta\log_2|\Sigma| \\
&\quad\leqslant \delta_0 + \mathsf{cap}^{\mathrm{com}}(\Phi).
\end{aligned}$$

Taking the limit as $\epsilon \to 0$, we get

$$\mathsf{cap}(\Gamma_\Phi) \leqslant \delta_0 + \mathsf{cap}^{\mathrm{com}}(\Phi).$$

Finally, since this holds for any $\delta_0 > 0$, we get the desired result,

$$\mathsf{cap}(\Gamma_\Phi) \leqslant \mathsf{cap}^{\mathrm{com}}(\Phi).$$

■

APPENDIX B
INDEPENDENCE ENTROPY FOR FULLY CONSTRAINED SYSTEMS

Here we Prove Theorem 22. We begin by recalling relevant definitions from [27]. A $\mathbb{Z}^d$ shift space X, is a subset $X \subseteq \Sigma^{\mathbb{Z}^d}$ that is closed under shifts, i.e., for all $\mathbf{v} \in \mathbb{Z}^d$, and all $x \in X$, $\sigma_{\mathbf{v}}(x) \in X$.

Definition 40: Let $d, k \in \mathbb{N}$. Given a set of forbidden words $\Phi \subseteq \Sigma^{F_k^d}$, the $\mathbb{Z}^d$ shift space over Σ defined by Φ is

$$X_\Phi \triangleq \left\{x \in \Sigma^{\mathbb{Z}^d} \ : \ \forall \mathbf{v} \in \mathbb{Z}^d, x_{\mathbf{v}+F_k^d} \notin \Phi\right\}.$$

□

Given a finite alphabet Σ, let $\tilde{\Sigma}$ denote the set of all non-empty subset of Σ, i.e.,

$$\tilde{\Sigma} \triangleq \{A \subseteq \Sigma \ : \ A \neq \emptyset\}.$$

Definition 41: Let $d \in \mathbb{N}$, $S \subseteq \mathbb{Z}^d$, and let $\tilde{x}$ be a configuration on S over $\tilde{\Sigma}$, i.e., $\tilde{x} \in \tilde{\Sigma}^S$. Denote by $\varphi(\tilde{x})$ the set of fillings of $\tilde{x}$,

$$\varphi(\tilde{x}) \triangleq \left\{ x \in \Sigma^S \ : \ \forall \mathbf{v} \in S,\ x_{\{\mathbf{v}\}} \in \tilde{x}_{\{\mathbf{v}\}} \right\}.$$

□

Definition 42: Let $d \in \mathbb{N}$, and let X be a $\mathbb{Z}^d$ shift space over Σ. We denote by $\tilde{X}$ the multi-choice shift space corresponding to X,

$$\tilde{X} \triangleq \left\{ \tilde{x} \in \tilde{\Sigma}^{\mathbb{Z}^d} \ : \ \varphi(\tilde{x}) \subseteq X \right\}.$$

We also denote by $\mathcal{B}_n(\tilde{X})$ the set of all eligible configurations on F_n^d in $\tilde{X}$, i.e.,

$$\mathcal{B}_n(\tilde{X}) \triangleq \left\{ \tilde{x}_{F_n^d} \ : \ \tilde{x} \in \tilde{X} \right\}.$$

□

Definition 43: Let $d \in \mathbb{N}$, and let X be a $\mathbb{Z}^d$ shift space. We define the combinatorial independence entropy *of X, denoted as $h_{\text{ind}}^{\text{com}}(X)$, by*

$$h_{\text{ind}}^{\text{com}}(X) \triangleq \limsup_{n\to\infty} \frac{1}{n^d} \max \left\{ \log_2 |\varphi(\tilde{w})| \ : \ \tilde{w} \in \mathcal{B}_n(\tilde{X}) \right\}.$$

□

Note that in [27] the definition of combinatorial independence entropy is slightly more general and defined over all shapes and not only on the shapes F_n^d. Finally, given a fully constrained system $\Phi \subseteq \Sigma^{F_k^n}$ (see Definition 38), its representation as a SCS is given by Γ_Φ in (43). We are now ready to prove Theorem 22.

Proof: [Proof of Theorem 22] Let $d, k \in \mathbb{N}$, and let $\Phi \subseteq \Sigma^{F_k^d}$ be a fully constrained system, with its SCS representation Γ_Φ from (43). The claim we want to prove is that

$$h_{\text{ind}}^{\text{com}}(X_\Phi) = h_{\text{ind}}(\Gamma_\Phi).$$

First, we show that $h_{\text{ind}}^{\text{com}}(X_\Phi) \leqslant h_{\text{ind}}(\Gamma_\Phi)$. For every $n \in \mathbb{N}$ choose $\tilde{w}_n \in \mathcal{B}_n(\tilde{X}_\Phi)$ which maximizes $|\varphi(\tilde{w}_n)|$. Now consider the independent measures μ_n such that $\pi_{\{\mathbf{v}\}}(\mu_n)$ is the uniform distribution over $(\tilde{w}_n)_{\{\mathbf{v}\}}$. Note that in $\mathcal{B}_n(\tilde{X})$, the forbidden patterns are considered without modulo while in $\mathcal{B}_n(\Gamma_\Phi)$ the calculation of the marginals' average uses modulo n. Therefore, if a filling in $\varphi(\tilde{w}_n)$ belongs to X_Φ, in μ_n there is perhaps a positive probability to see a forbidden pattern only in the boundaries. In F_n^d, the k-boundary is the set $F_n^d \setminus F_{n-k}^d$ of size $n^d - (n-k)^d$. Since $(n^d - (n-k)^d)/n^d \to 0$ as $n \to \infty$, we obtain that for every $\epsilon > 0$, for every $n \in \mathbb{N}$ such that $(n^d - (n-k)^d)/n^d \leqslant \epsilon$, we have that $\mu_n \in \mathbb{B}_\epsilon(\Gamma_\Phi)$. Thus,

$$\begin{aligned} \widehat{h_{\text{ind}}}(\mathbb{B}_\epsilon(\Gamma_\Phi)) &= \limsup_{n\to\infty} \sup_{\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))} \frac{1}{n^d} H(\mu) \\ &\geqslant \limsup_{n\to\infty} \frac{1}{n^d} H(\mu_n) \\ &= \limsup_{n\to\infty} \frac{1}{n^d} \log_2 |\varphi(\tilde{w}_n)| \\ &= h_{\text{ind}}^{\text{com}}(X_\Phi). \end{aligned}$$

Taking $\epsilon \to 0$ we obtain

$$h_{\text{ind}}(\Gamma_\Phi) \geqslant h_{\text{ind}}^{\text{com}}(X_\Phi).$$

We now show that $h_{\text{ind}}(\Gamma_\Phi) \leqslant h_{\text{ind}}^{\text{com}}(X_\Phi)$. Fix $\delta > 0$ and take $\delta_1 > 0$ small enough such that $\delta_1 < \frac{1}{3}\delta$. Take $n_0 \in \mathbb{N}$ large enough such that for all $n \geqslant n_0$,

$$\frac{1}{n^d} \max_{\tilde{w} \in \mathcal{B}_n(\tilde{X}_\Phi)} \left\{ \log_2 |\varphi(\tilde{w})| \right\} \leqslant h_{\text{ind}}^{\text{com}}(X_\Phi) + \frac{1}{3}\delta. \tag{44}$$

We now take $\epsilon > 0$ small enough such that all the following hold,

$$-|\Sigma| \sqrt[k^d]{n_0^d \epsilon^{\frac{1}{4}}} \log_2 \sqrt[k^d]{n_0^d \epsilon^{\frac{1}{4}}} < \frac{1}{3}\delta, \tag{45}$$

$$2^d \epsilon^{\frac{3}{4}} \log_2 |\Sigma| < \frac{1}{2}\delta_1, \tag{46}$$

$$\left| \widehat{h_{\text{ind}}}(\mathbb{B}_\epsilon(\Gamma_\Phi)) - h_{\text{ind}}(\Gamma_\Phi) \right| \leqslant \frac{1}{16}\delta_1.$$

By the definition of $\widehat{h_{\text{ind}}}(\mathbb{B}_\epsilon(\Gamma_\Phi))$ we may find $n \geqslant n_0$ large enough such that all the following hold,

$$2\left(1 - \left(1 - \frac{n_0}{n}\right)^d\right) \log_2 |\Sigma| \leqslant \frac{1}{4}\delta_1, \tag{47}$$

$$\left| \sup_{\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))} \frac{1}{n^d} H(\mu) - h_{\text{ind}}(\Gamma_\Phi) \right| \leqslant \frac{1}{8}\delta_1,$$

and there exists $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))$ for which

$$\left| \frac{1}{n^d} H(\mu) - h_{\text{ind}}(\Gamma_\Phi) \right| \leqslant \frac{1}{4}\delta_1.$$

Since $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))$, we have

$$\frac{1}{n^d} \sum_{\mathbf{v} \in F_n^d} \pi_{F_k^d + \mathbf{v}}(\mu)(\Phi) \leqslant \epsilon.$$

Denote $m \triangleq \lfloor n/n_0 \rfloor$. We now partition F_n^d into m^d blocks of shape $F_{n_0}^d$ in the natural way, $\left\{ n_0\mathbf{v} + F_{n_0}^d : \mathbf{v} \in F_m^d \right\}$, as well as a boundary $F_n^d \setminus F_{mn_0}^d$. Note that

$$\mu \cong \bigotimes_{\mathbf{v} \in F_m^d} \pi_{n_0\mathbf{v} + F_{n_0}^d}(\mu) \otimes \pi_{F_n^d \setminus F_{mn_0}^d}(\mu).$$

Since μ is independent we obtain

$$\begin{aligned} &\left| \frac{1}{(mn_0)^d} H(\pi_{F_{mn_0}^d}(\mu)) - h_{\text{ind}}(\Gamma_\Phi) \right| \\ &\quad \leqslant \left| \frac{1}{(mn_0)^d} H(\pi_{F_{mn_0}^d}(\mu)) - \frac{1}{n^d} H(\mu) \right| \\ &\qquad + \left| \frac{1}{n^d} H(\mu) - h_{\text{ind}}(\Gamma_\Phi) \right| \\ &\quad \leqslant (mn_0)^d \left(\frac{1}{(mn_0)^d} - \frac{1}{n^d} \right) \log_2 |\Sigma| \\ &\qquad + \frac{n^d - (mn_0)^d}{n^d} \log_2 |\Sigma| + \frac{1}{4}\delta_1 \\ &\quad \leqslant 2\frac{n^d - (mn_0)^d}{n^d} \log_2 |\Sigma| + \frac{1}{4}\delta_1 \\ &\quad \leqslant 2\left(1 - \left(1 - \frac{n_0}{n}\right)^d\right) \log_2 |\Sigma| + \frac{1}{4}\delta_1 \\ &\quad \leqslant \frac{1}{2}\delta_1, \end{aligned} \tag{48}$$

where the last inequality holds due to (47). Let $Z : F_m^d \to \mathbb{R}$ be a function defined by

$$Z(\mathbf{v}) \triangleq \frac{1}{n_0^d} \sum_{\mathbf{u}\in F_{n_0}^d} \pi_{F_k^d+n_0\mathbf{v}+\mathbf{u}}(\mu)(\Phi)$$

(with coordinates taken modulo n). Note that since $\mu \in \overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma_\Phi))$, we have

$$\frac{1}{n^d}\sum_{\mathbf{v}\in F_n^d} \pi_{F_k^d+\mathbf{v}}(\mu)(\Phi) \leqslant \epsilon.$$

If we now take $\mathbf{v}$ to be random uniformly distributed in F_m^d, then

$$E[Z(\mathbf{v})] = \frac{1}{m^d}\sum_{\mathbf{v}\in F_m^d}\frac{1}{n_0^d}\sum_{\mathbf{u}\in F_{n_0}^d}\pi_{F_k^d+n_0\mathbf{v}+\mathbf{u}}(\mu)(\Phi) \leqslant \left(1+\frac{1}{m}\right)^d\epsilon.$$

By Markov's inequality we have

$$\Pr\left(Z(\mathbf{v}) \geqslant \epsilon^{\frac14}\right) \leqslant \epsilon^{-\frac14} E[Z(\mathbf{v})] \leqslant \left(1+\frac{1}{m}\right)^d \epsilon^{\frac34}. \quad (49)$$

Recall that each $\mathbf{v} \in F_m^d$ may be identified with the $F_{n_0}^d$ block of F_n^d in coordinates $n_0\mathbf{v} + F_{n_0}^d$. Define,

$$\mathcal{L} \triangleq \left\{\mathbf{v}\in F_m^d \ : \ Z(\mathbf{v}) \geqslant \epsilon^{\frac14}\right\}.$$

Since $\mathbf{v}$ was distributed uniformly in F_m^d, by (49) we have,

$$|\mathcal{L}| \leqslant (m+1)^d \epsilon^{\frac34}. \quad (50)$$

It now follows that

$$\begin{aligned}
&\frac{1}{(mn_0)^d}\sum_{\mathbf{v}\in F_m^d\setminus\mathcal{L}} H(\pi_{n_0\mathbf{v}+F_{n_0}^d}(\mu)) \\
&= \frac{1}{(mn_0)^d} H(\pi_{F_{mn_0}^d}(\mu)) - \frac{1}{(mn_0)^d}\sum_{\mathbf{v}\in\mathcal{L}} H(\pi_{n_0\mathbf{v}+F_{n_0}^d}(\mu)) \\
&\overset{(a)}{\geqslant} h_{\mathrm{ind}}(\Gamma_\Phi) - \frac12\delta_1 - \frac{1}{(mn_0)^d}\sum_{\mathbf{v}\in\mathcal{L}} H(\pi_{n_0\mathbf{v}+F_{n_0}^d}(\mu)) \\
&\overset{(b)}{\geqslant} h_{\mathrm{ind}}(\Gamma_\Phi) - \frac12\delta_1 - \frac{(m+1)^d\epsilon^{\frac34}n_0^d}{(mn_0)^d}\log_2|\Sigma| \\
&\geqslant h_{\mathrm{ind}}(\Gamma_\Phi) - \frac12\delta_1 - 2^d\epsilon^{\frac34}\log_2|\Sigma| \\
&\overset{(c)}{>} h_{\mathrm{ind}}(\Gamma_\Phi) - \delta_1 \\
&> h_{\mathrm{ind}}(\Gamma_\Phi) - \frac13\delta,
\end{aligned}$$

where (a) follows from (48), (b) follows from (50), and (c) follows from (46). Since there are at most m^d summands on the left-hand side, there exists $\mathbf{v}_0 \in F_m^d \setminus \mathcal{L}$ such that

$$\frac{1}{n_0^d} H(\pi_{n_0\mathbf{v}_0+F_{n_0}^d}(\mu)) \geqslant h_{\mathrm{ind}}(\Gamma_\Phi) - \frac13\delta. \quad (51)$$

We denote by ν the independent measure $\nu \triangleq \pi_{F_{n_0}^d+n_0\mathbf{v}_0}(\mu)$.

Note that if we consider ν in a non-cyclic manner, we obtain that

$$\frac{1}{(n_0-k+1)^d}\sum_{\mathbf{u}\in F_{n_0-k+1}^d}\pi_{\mathbf{u}+F_k^d}(\nu)(\Phi) \leqslant \frac{n_0^d}{(n_0-k+1)^d}\epsilon^{\frac14},$$

and in particular, for every coordinate $\mathbf{u} \in F_{n_0-k+1}^d$, we have that $\pi_{\mathbf{u}+F_k^d}(\nu)(\Phi) \leqslant n_0^d\epsilon^{\frac14}$. Let us define

$$p \triangleq \sqrt[k^d]{n_0^d\epsilon^{\frac14}}.$$

Hence, since ν is an independent measure, if $a \in \Phi$ then there must be a coordinate $\mathbf{t} \in F_k^d$ for which $\pi_{\mathbf{u}+\mathbf{t}}(\nu)(a_\mathbf{t}) \leqslant p$.

We now construct a configuration $\tilde{\omega} \in \tilde{\Sigma}^{F_{n_0}^d}$. For every coordinate $\mathbf{u} \in F_{n_0}^d$ we take

$$\tilde{\omega}_\mathbf{u} = \left\{a\in\Sigma \ : \ \pi_{\{\mathbf{u}\}}(\nu)(a) > p\right\}.$$

By our previous observation, $\tilde{\omega} \in \mathcal{B}_{n_0}(\tilde{X}_\Phi)$ since any filling of $\tilde{\omega}$ cannot contain a forbidden word from Φ as it requires at least one position $\mathbf{u}$ such that $\pi_{\{\mathbf{u}\}}(\nu) \leqslant p$. Moreover,

$$\begin{aligned}
&\log_2|\tilde{\omega}_\mathbf{u}| \\
&\geqslant -\sum_{a\in\tilde{\omega}_\mathbf{u}}\pi_{\{\mathbf{u}\}}(\nu)(a)\log_2(\pi_{\{\mathbf{u}\}}(\nu)(a)) \\
&= H(\pi_{\{\mathbf{u}\}}(\nu)) + \sum_{a\in\Sigma\setminus\tilde{\omega}_\mathbf{u}}\pi_{\{\mathbf{u}\}}(\nu)(a)\log_2(\pi_{\{\mathbf{u}\}}(\nu)(a)) \\
&\geqslant H(\pi_{\{\mathbf{u}\}}(\nu)) + |\Sigma|\, p\log_2\ p \\
&> H(\pi_{\{\mathbf{u}\}}(\nu)) - \frac13\delta
\end{aligned}$$

where the last inequality follows from (45). Hence, using (51),

$$\begin{aligned}
\frac{1}{n_0^d}\log_2|\varphi(\tilde{\omega})| &= \frac{1}{n_0^d}\sum_{\mathbf{u}\in F_{n_0}^d}\log_2|\tilde{\omega}_\mathbf{u}| \\
&= \frac{1}{n_0^d}\sum_{\mathbf{u}\in F_{n_0}^d} H(\pi_{\{\mathbf{u}\}}(\nu)) - \frac13\delta \\
&\geqslant h_{\mathrm{ind}}(\Gamma_\Phi) - \frac23\delta.
\end{aligned}$$

Finally, using (44), this implies that,

$$h_{\mathrm{ind}}^{\mathrm{com}}(X_\Phi) \geqslant \frac{1}{n_0^d}\max_{\tilde{\omega}\in\mathcal{B}_{n_0}(\tilde{X}_\Phi)}\left\{\log_2|\varphi(\tilde{\omega})|\right\} - \frac13\delta \geqslant h_{\mathrm{ind}}(\Gamma_\Phi) - \delta.$$

Since this holds for every $\delta > 0$ we have $h_{\mathrm{ind}}(\Gamma_\Phi) \leqslant h_{\mathrm{ind}}^{\mathrm{com}}(X_\Phi)$, as claimed. ■

Acknowledgments

The authors would like to thank the associate editor and the anonymous reviewers, whose comments improved the presentation of the paper.

References

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

[2] O. Elishco, T. Meyerovitch, and M. Schwartz, "Encoding semiconstrained systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 395–399.

[3] O. F. Kurmaev, "Constant-weight and constant-charge binary run-length limited codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4497–4515, Jul. 2011.

[4] S. Kayser and P. H. Siegel, "Constructions for constant-weight ICI-free codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 1431–1435.

[5] M. Qin, E. Yaakobi, and P. H. Siegel, "Constrained codes that mitigate inter-cell interference in read/write cycles for flash memories," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 836–846, May 2014.

[6] Y. M. Chee, C. Johan, H. M. Kiah, S. Ling, T. T. Nguyen, and V. K. Vu, "Efficient encoding/decoding of capacity-achieving constant-composition ICI-free codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 205–209.

[7] Y. M. Chee, C. Johan, H. M. Kiah, S. Ling, T. T. Nguyen, and V. K. Vu, "Rates of constant-composition codes that mitigate intercell interference," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 200–204.

[8] A. Shafarenko, A. Skidin, and S. K. Turitsyn, "Weakly-constrained codes for suppression of patterning effects in digital communications," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2845–2854, Oct. 2010.

[9] A. Shafarenko, K. S. Turitsyn, and S. K. Turitsyn, "Information-theory analysis of skewed coding for suppression of pattern-dependent errors in digital communications," *IEEE Trans. Commun.*, vol. 55, no. 2, pp. 237–241, Feb. 2007.

[10] R. Karabed, D. Neuhoff, and A. Khayrallah, "The capacity of costly noiseless channels," IBM Res., New York, NY, USA, Tech. Rep. RJ 6040 (59639), Jan. 1988.

[11] A. S. Khayrallah and D. L. Neuhoff, "Coding for channels with cost constraints," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 854–867, May 1996.

[12] B. H. Marcus and R. M. Roth, "Improved Gilbert–Varshamov bound for constrained systems," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1213–1221, Jul. 1992.

[13] E. H. Lieb, "Residual entropy of square ice," *Phys. Rev. J. Arch.*, vol. 162, no. 1, pp. 162–172, 1967.

[14] R. J. Baxter, "Hard hexagons: Exact solution," *J. Phys. A, Math. Gen.*, vol. 13, pp. L61–L70, 1980.

[15] M. Schwartz and J. Bruck, "Constrained codes as networks of relations," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2179–2195, May 2008.

[16] E. Louidor and B. H. Marcus, "Improved lower bounds on capacities of symmetric 2D constraints using Rayleigh quotients," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1624–1639, Apr. 2010.

[17] N. J. Calkin and H. S. Wilf, "The number of independent sets in a grid graph," *SIAM J. Discrete Math.*, vol. 11, no. 1, pp. 54–60, 1998.

[18] P. H. Siegel and J. K. Wolf, "Bit-stuffing bounds on the capacity of 2-dimensional constrained arrays," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Aug. 1998, p. 323.

[19] H. Ito, A. Kato, Z. Nagy, and K. Zeger, "Zero capacity region of multidimensional run length constraints," *Elect. J. Combinat.*, vol. 6, no. 1, p. 33, 1999.

[20] A. Kato and K. Zeger, "On the capacity of two-dimensional run-length constrained channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1527–1540, Jul. 1999.

[21] Z. Nagy and K. Zeger, "Capacity bounds for the three-dimensional (0,1) run length limited channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 1030–1033, May 2000.

[22] E. Ordentlich and R. M. Roth, "Independent sets in regular hypergraphs and multi-dimensional runlength-limited constraints," *SIAM J. Discrete Math.*, vol. 17, no. 4, pp. 615–623, 2004.

[23] S. Halevy, J. Chen, R. M. Roth, P. H. Siegel, and J. K. Wolf, "Improved bit-stuffing bounds on two-dimensional constraints," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 824–838, May 2004.

[24] A. Sharov and R. M. Roth, "Two-dimensional constrained coding based on tiling," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1800–1807, Apr. 2010.

[25] I. Tal and R. M. Roth, "Bounds on the rate of 2-D bit-stuffing encoders," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2561–2567, Jun. 2010.

[26] M. Schwartz and A. Vardy, "New bounds on the capacity of multidimensional run-length constraints," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4373–4382, Jul. 2011.

[27] E. Louidor, B. Marcus, and R. Pavlov, "Independence entropy of $\mathbb{Z}^d$-shift spaces," *Acta Appl. Math.*, vol. 126, no. 1, pp. 297–317, 2013.

[28] T. Meyerovitch and R. Pavlov, "On independence and entropy for high-dimensional isotropic subshifts," *Proc. London Math. Soc.*, vol. 109, no. 4, pp. 921–945, 2014.

[29] T. L. Poo, P. Chaichanavong, and B. H. Marcus, "Tradeoff functions for constrained systems with unconstrained positions," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1425–1449, Apr. 2006.

[30] J.-R. Chazottes, J.-M. Gambaudo, M. Hochman, and E. Ugalde, "On the finite-dimensional marginals of shift-invariant measures," *Ergodic Theory Dyn. Syst.*, vol. 32, no. 5, pp. 1485–1500, 2012.

[31] R. Berger, "The undecidability of the domino problem," *Memoirs Amer. Math. Soc.*, no. 66, pp. 1–72, Nov. 1966.

[32] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY, USA: Springer, 1998.

[33] P. Diaconis and D. Freedman, "Finite exchangeable sequences," *Ann. Probab.*, vol. 8, no. 4, pp. 745–764, 1980.

[34] L. Lovász and B. Szegedy, "Szemerédi's lemma for the analyst," *GAFA Geometric Funct. Anal.*, vol. 17, no. 1, pp. 252–270, Apr. 2007. [Online]. Available: https://doi.org/10.1007/s00039-007-0599-6

[35] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1978.

Ohad Elishco (S'12) received his B.Sc. in mathematics and his B.Sc. in electrical engineering in 2012 from Ben-Gurion University of the Negev, Israel; the M.Sc. degree in electrical engineering in 2013 from Ben-Gurion University of the Negev. In 2013 he started his Ph.D in electrical engineering also in Ben-Gurion University. His research interests are constrained coding and dynamical systems.

Tom Meyerovitch received the B.Sc. in mathematics and computer science from from Tel-Aviv University, Israel, in 1999; the M.Sc. and Ph.D. degrees in mathematics in 2004 and 2010, respectively, also from Tel-Aviv University. In 2010-2012 he held a post doctoral position at the University of British Columbia, and the Pacific Institute for Mathematical studies in Vancouver Canada. Since 2012, he has been on the faculty of the Department of Mathematics at Ben-Gurion University, Beer-Sheva, Israel. He studies various mathematical aspects of dynamical systems, in particular symbolic dynamics.

Moshe Schwartz (M'03–SM'10) is an associate professor at the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include algebraic coding, combinatorial structures, and digital sequences.

Prof. Schwartz received the B.A. (*summa cum laude*), M.Sc., and Ph.D. degrees from the Technion – Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004 respectively, all from the Computer Science Department. He was a Fulbright post-doctoral researcher in the Department of Electrical and Computer Engineering, University of California San Diego, and a post-doctoral researcher in the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012–2014, he was a visiting scientist at the Massachusetts Institute of Technology (MIT).

Prof. Schwartz received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage.