J. Inst. Math. Jussieu (2025), 1–69
doi:10.1017/S147474802510100X

ALGEBRAIC DYNAMICAL SYSTEMS FROM LDPC CODES SATISFY A STRONG NEGATION OF THE WEAK PINSKER PROPERTY

TIM AUSTIN[1], LEWIS BOWEN[2] AND CHRISTOPHER SHRIVER[3]

[1] *Mathematics Institute, University of Warwick, U.K.*
(tim.austin@warwick.ac.uk)

[2] *Department of Mathematics, University of Texas at Austin, U.S.A.*
(lpbowen@math.utexas.edu)

[3] *Department of Mathematics, University of Texas at Austin, U.S.A.*
(christopher.shriver@math.utexas.edu)

(Received 14 May 2024; revised 15 May 2025; accepted 16 May 2025)

Abstract We construct an explicit algebraic example of a subshift of finite type over a group Γ with an invariant Markov measure which has completely positive sofic entropy (with respect to 'most' sofic approximations) and yet does not have a direct Bernoulli factor because its model spaces shatter into exponentially many clusters of sub-exponential size. The example and its analysis are related to random low-density parity-check (LDPC) codes.

Contents

Tim Austin, supported in part by NSF grant DMS-1855694
Lewis Bowen, supported in part by NSF grant DMS-2154680
Christopher Shriver, supported in part by NSF grant DMS-1937215.

Keywords: sofic entropy; LDPC codes; weak Pinsker property; Bernoulli shifts; algebraic dynamics

2020 Mathematics subject classification: Primary 37A35

1. Introduction

This paper constructs explicit dynamical systems with unusual properties related to recent work on the weak Pinsker property and shattering. The construction is explained next; the background, motivation and precise statements are developed afterwards.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Fix natural numbers d,k and let $\Gamma = \Gamma_{d,k}$ be the d-fold free product of order-k cyclic groups

$$\Gamma := \langle s_1,\dots,s_d :\ s_1^k = \dots = s_d^k = e\rangle = \underbrace{\mathbb{Z}_k * \dots * \mathbb{Z}_k}_{d},$$

where $\mathbb{Z}_k$ means $\mathbb{Z}/k\mathbb{Z}$. The set of all functions $x : \Gamma \to \mathbb{Z}_2$ is denoted $\mathbb{Z}_2^\Gamma$. This is a compact Abelian group under pointwise addition with the pointwise convergence topology. Let $X \le \mathbb{Z}_2^\Gamma$ be the closed subgroup defined by

$$X = \left\{ x \in \mathbb{Z}_2^\Gamma :\ \sum_{j=0}^{k-1} x_{gs_i^j} = 0 \quad \forall g \in \Gamma,\ i = 1,\dots,d \right\},$$

and let $\mu = m_X$ be the Haar probability measure on X.

For $g \in \Gamma$, let $T^g : \mathbb{Z}_2^\Gamma \to \mathbb{Z}_2^\Gamma$ be the continuous group automorphism given by permuting indices on the left:

$$T^g((x_h)_{h\in\Gamma}) = (x_{g^{-1}h})_{h\in\Gamma}. \tag{1}$$

The subgroup X is invariant under this action, and hence so is its Haar measure.

This state space is easily visualized in terms of the Cayley graph of Γ with its generators $s_1,\dots,s_k$. Through each group element g, each s_i generates a k-cycle. So each vertex of the Cayley graph lies in d of these k-cycles, and there are no other relations in the group, so these k-cycles are attached together into a hyper-tree. With this picture in mind, a member of X is simply an assignment of zeros and ones to the vertices of the Cayley graph such that the sum around every k-cycle is even. For this reason, and by analogy with similar constructions in coding theory, we call X a **parity check subshift**. Indeed, certain random finite parity check codes play a crucial auxiliary role later in the paper: see Section 6.

Informally stated, our main results are these:

- If $k > d \ge 3$, then the sofic entropy of the dynamical system (X,m_X,T) is $(1-d/k)\log(2)$.
- Every nontrivial factor of (X,m_X,T) has positive sofic entropy and therefore positive Rokhlin entropy (this property is called 'completely positive entropy' or 'CPE'). In fact, we prove the stronger assertion that the outer Pinsker factor of (X,m_X,T) is trivial.
- The system (X,m_X,T) is not isomorphic to a direct product of a nontrivial Bernoulli shift with another system. Combined with the previous conclusion, this is a strong negation of the weak Pinsker property.
- The system (X,m_X,T) is not weakly contained in a Bernoulli shift. It is one of the first examples that has completely positive entropy and also this property.

Next, we introduce background needed to state our main results precisely.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

1.1. Background: classical entropy theory

Kolmogorov introduced entropy theory into dynamics for the purpose of distinguishing Bernoulli shifts up to measure conjugacy. Given a standard probability space $(\mathtt{K},\kappa)$, the **Bernoulli shift over a countable group** Γ **with base space** $(\mathtt{K},\kappa)$ consists of the probability space $(\mathtt{K}^{\Gamma},\kappa^{\Gamma})$ together with the action of Γ by permuting indices as in 1. A sample of $(\mathtt{K}^{\Gamma},\kappa^{\Gamma})$ is a random $\mathtt{K}$-valued configuration $(x_g)_{g\in\Gamma}$ whose coordinates are i.i.d. with law κ.

Suppose we are given a standard probability space (X,μ) (where we have left the sigma-algebra out of the notation for simplicity). Let $\mathrm{Aut}(X,\mu)$ denote the group of all measure-preserving automorphisms of (X,μ). A **pmp** (probability-measure-preserving) action of Γ is a homomorphism $T:\Gamma\to\mathrm{Aut}(X,\mu)$. The triple (X,μ,T) is a Γ**-system**. We also refer to it as a **system** or **action** if Γ is understood.

If we are given two Γ-systems (X_i,μ_i,T_i), then a measurable map $\Phi:X_1\to X_2$ is a **factor map** if it is a.e. Γ-equivariant (this means $\Phi(T_1^g x)=T_2^g\Phi(x)$ for all $g\in\Gamma$ and μ_1-a.e. $x\in X_1$) and the pushforward measure satisfies $\Phi_*\mu_1=\mu_2$. More precisely, we allow that Φ be defined only on a subset of full measure. If Φ is invertible (after ignoring a null set), then it is a **measure-conjugacy** or **isomorphism**.

If $\Gamma=\mathbb{Z}$, then an action of the integers is given by a single transformation $T\in\mathrm{Aut}(X,\mu)$. Thus, it makes sense to consider whether two transformations are measurably conjugate.

A problem attributed to von Neumann asks whether there could be two Bernoulli shifts over the group of integers which are not measurably conjugate. To answer this, Kolmogorov defined the entropy rate of a dynamical system in the special case in which $\Gamma=\mathbb{Z}$ [43, 44]. He proved entropy is invariant under measure-conjugacy and computed entropy rates for Bernoulli shifts, thereby answering the problem in the affirmative. In fact, the entropy rate of a Bernoulli shift action is the same as the Shannon entropy of the base space. When the base space is $(\mathtt{K},\kappa)$ and $\mathtt{K}$ is countable, its **Shannon entropy** is

$$H(\kappa)=-\sum_{k\in\mathtt{K}}\kappa(\{k\})\log(\kappa(\{k\})).$$

If κ is not supported on a countable set, then its Shannon entropy is defined to be $+\infty$.

Kolmogorov's theory extends fairly directly to the case when Γ is amenable. The first published work on entropy theory for general amenable groups is due to Kieffer [41].

Since Kolmogorov's pioneering work, entropy and Bernoulli shifts have played a central role in classifying dynamical systems. For example, Sinai proved that if an ergodic action of $\mathbb{Z}$ has positive entropy, then it factors onto a Bernoulli shift of the same entropy [67]. Because entropy cannot increase under a factor map, this shows that Bernoulli factors witness entropy. Inspired by Sinai's theorem, Ornstein proved that Bernoulli shifts over the integers are isomorphic if and only if they have the same entropy [52, 53]. These results were extended to the case of amenable acting groups in [58].

Shannon entropy is easily seen to be additive under direct products, and this property is inherited by Kolmogorov's entropy rate. Naively, one might guess that any ergodic system is isomorphic to a direct product of a Bernoulli shift with a zero entropy system. This turns out to be false; counterexamples to weaker claims appear in [56, 55, 54]. If it were true for a system which was not itself isomorphic to a Bernoulli shift, then additivity implies

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

that the system would have a nontrivial (direct) factor with zero entropy. A system is said to have **completely positive entropy** (CPE) if every nontrivial factor has positive entropy. Here a 'trivial' factor is a measure-preserving system where the measure is a delta mass at a single point. This system is a factor of every other system, and it is easy to see that its entropy is zero. The paper [57] shows that for any positive number $h > 0$, there exist uncountably many pairwise non-isomorphic transformations which are CPE and have entropy h.

A factor map $\pi : (X_1,\mu_1,T_1) \to (X_2,\mu_2,T_2)$ is said to be **direct** or **split** if there is another factor map $\xi : (X_1,\mu_1,T_1) \to (X_3,\mu_3,T_3)$ so that the pair (π,ξ) together forms an isomorphism

$$(X_1,\mu_1,T_1) \to (X_2 \times X_3,\mu_2 \times \mu_3,T_2 \times T_3).$$

Note that the measure on the right-hand side is required to be the product, so in particular the factor maps π and ξ must generate independent sigma-subalgebras of subsets of X_1. While Sinai's factor theorem shows the existence of Bernoulli factors, it does not say anything about the existence of *direct* Bernoulli factors.

In the 1970s, Thouvenot defined a system to have the weak Pinsker property (WPP) if for every $\epsilon > 0$, it is isomorphic to a direct product of a Bernoulli shift with a system of entropy less than ϵ [69]. In other words, a system has the WPP if its entropy is witnessed by *direct* Bernoulli factors. Thouvenot asked whether every ergodic transformation has the WPP. The first author recently proved that this is indeed the case [5]. Moreover, the statement holds whenever the acting group Γ is amenable.

1.2. Background: sofic entropy theory

The second author constructed a system without the WPP in the special case when the group Γ is a free group of sufficiently high rank [17]. To explain, we need to pause for a moment to discuss entropy theory when the acting group is not amenable.

An example due to Ornstein and Weiss in [58] suggested it might not be possible to extend entropy theory to non-amenable groups. However, this changed with the introduction of sofic entropy theory [11]. The new theory applies to all sofic groups, which is a class of groups containing amenable and linear groups, for example. It is unknown whether all countable groups are sofic. Sofic entropy theory is reviewed in §2.2.

A sofic approximation to a group Γ is a sequence Σ of partial actions on finite sets which approximates the action of the group on itself by left-translations. To be precise, $\Sigma = (\sigma_n)_{n\in\mathbb{N}}$ where $\sigma_n : \Gamma \to \mathrm{Sym}(V_n)$, V_n are finite sets, $\mathrm{Sym}(V_n)$ is the symmetric group on V_n and the sequence is required to satisfy for all $g,h,f \in \Gamma$ such that f is not the identity,

$$\begin{aligned}1 &= \lim_{n\to\infty} |V_n|^{-1}|\{v \in V_n : \ \sigma_n(gh)v = \sigma_n(g)\sigma_n(h)v\}| \\ 0 &= \lim_{n\to\infty} |V_n|^{-1}|\{v \in V_n : \ \sigma_n(f)v = v\}|.\end{aligned}$$

A group is called **sofic** if it admits a sofic approximation. The sofic entropy of a system (X,μ,T) (defined in Section 2.2) depends a priori on a choice of sofic approximation, although for many actions where it has been computed, it has been shown not to.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Many classical results extend to the sofic setting. For example, the sofic entropy of a Bernoulli shift action is equal to the Shannon entropy of the base. Sofic entropy is a measure-conjugacy invariant, and so two Bernoulli shifts with different sofic entropy are not isomorphic. In recent work, Seward completed the converse direction: for any countable group Γ, if two Bernoulli shifts over Γ have the same *base space Shannon entropy*, then they are measurably conjugate [68, 13, 66]. This converse does not depend on sofic entropy, which might not even be defined.

In a series of works generalizing Krieger's Theorem [63, 64], Seward introduced Rokhlin entropy. To define it, suppose we are given an action $T : \Gamma \to \mathrm{Aut}(X,\mu)$ and a countable measurable partition $\mathcal{P}$ of X. We say the partition is **generating** if the smallest sigma-algebra containing it which is also $T(\Gamma)$-invariant is the sigma-algebra of all measurable sets (up to sets of measure zero). Then Rokhlin entropy is defined to be the infimum of $H_\mu(\mathcal{P})$ over all generating partitions, where

$$H_\mu(\mathcal{P}) = -\sum_{P \in \mathcal{P}} \mu(P) \log(\mu(P))$$

is the **Shannon entropy** of $(\mathcal{P},\mu)$.

It is immediate that Rokhlin entropy is a measure-conjugacy invariant. Moreover, it upper bounds sofic entropy. In fact, it is unknown whether Rokhlin entropy equals sofic entropy whenever the latter is not minus infinity (which can happen). However, the only known method for computing a lower bound to Rokhlin entropy uses sofic entropy. For example, it is unknown how to compute the Rokhlin entropy of Bernoulli shift actions, except in the case when Γ is assumed to be sofic.

In a different paper [65], Seward generalized Sinai's factor theorem: every ergodic system with positive Rokhlin entropy factors onto a Bernoulli shift with the same entropy.

However, other structural results about classical Kolmogorov–Sinai entropy break down outside the world of amenable groups. For example, Ornstein and Weiss' example shows that sofic entropy can *increase* under a factor map. In fact, recent work of the second author shows that if Γ is an arbitrary non-amenable group, then every Bernoulli shift over Γ factors onto every Bernoulli shift over Γ [15]. For example, Bernoulli shifts of small entropy factor onto Bernoulli shifts of infinite entropy.

Although Bernoulli shifts themselves have been classified, there is no known substitute for broader 'Ornstein theory', which provides necessary and sufficient conditions for general ergodic processes to be isomorphic to Bernoulli shifts. Moreover, some specific counterexamples show that the story must change substantially for some non-amenable groups. For example, when G has property (T), Popa and Sasyk [60] have given simple examples of factors of Bernoulli shifts that are not isomorphic to Bernoulli shifts.

It is also known that the weak Pinsker property does not hold for all non-amenable groups and for the main non-amenable notions of entropy. The first counterexample appeared in [17]. While that counterexample does not have the WPP, it might still admit some direct Bernoulli factors. In other words, the system might be measurably conjugate to the direct product of a Bernoulli shift with another system, but one cannot choose the other system to have entropy less than ϵ if $\epsilon > 0$ is chosen low enough. In this work, we present a new counterexample which does not admit *any* nontrivial Bernoulli factors.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Here, a Bernoulli shift $(\mathtt{K}^\Gamma, \kappa^\Gamma, T)$ is said to be **trivial** if κ is supported on a single point in $\mathtt{K}$. So a trivial Bernoulli shift is measurably conjugate to the trivial system which consists of Γ acting on a single point.

1.3. Main results

Recall from the introduction that Γ is the d-fold free power of $\mathbb{Z}_k$, $X \subset (\mathbb{Z}_2)^\Gamma$ is a certain closed 'parity check' subgroup and $T : \Gamma \to \mathrm{Aut}(X, m_X)$ is the canonical shift action by automorphisms.

Theorem A. *Let $k > d \geq 3$. Then there exists a sofic approximation $\Sigma = (\sigma_n)_n$ to Γ such that*

$$\mathrm{h}_\Sigma(X, m_X, T) = (1 - d/k)\log 2$$

and along which the outer Pinsker factor of (X, m_X, T) is trivial. In particular, (X, m_X, T) has completely positive sofic entropy along Σ.

Since sofic entropy is always bounded above by Rokhlin entropy, the last part of this theorem has the following immediate corollary:

Corollary 1.1. *The outer Rokhlin Pinsker factor of (X, m_X, T) is trivial, so it has completely positive Rokhlin entropy.*

Theorem B. *If $k > d \geq 3$, then the system (X, m_X, T) has no nontrivial direct Bernoulli factors.*

Remark. Theorems A and B should hold in the more general setting in which $\Gamma = \Gamma_1 * \cdots * \Gamma_d$, where each Γ_i is a group of order k. That is, we do not need to require that each Γ_i is cyclic. The proofs are essentially the same.

Remark. The **weak Pinsker entropy** of a system (X, μ, T) is defined to be the supremum of the Shannon entropies $H(\mathtt{K}, \kappa)$ over all direct Bernoulli factors of the form $\Gamma \curvearrowright (\mathtt{K}, \kappa)^\Gamma$. This concept was introduced in [18]. So Theorem B implies that (X, m_X, T) has zero weak Pinsker entropy.

Finally, our methods also give the following.

Theorem C. *If $k > d \geq 3$, then the system (X, m_X, T) is not weakly contained in a Bernoulli system.*

For a definition of weak containment of measure-preserving systems, see, for example, [19].

Theorems B and C are consequences of the system having totally shattered microstate spaces along some sofic approximation: see Corollary 5.5.

Together, Theorems A and C show that (X, m_X, T) has completely positive sofic entropy along some sofic approximation, and hence also completely positive Rokhlin entropy, but is not weakly contained in a Bernoulli shift. The authors are not aware of other systems for which both of these properties have been verified previously.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

1.4. Probabilistic versions of the main theorems

Both Theorem A and Theorem B will be derived as corollaries of probabilistic theorems. To explain, we say that a permutation of a set V is **k-uniform** if it consists entirely of k-cycles: that is, its cycle type is $[k^{n/k}]$. Recall that $\Gamma = \Gamma_{d,k} = \langle s_1, \ldots, s_d : s_1^k = \cdots = s_d^k = e \rangle$. A homomorphism $\sigma : \Gamma \to \mathrm{Sym}(V)$ is **k-uniform** if for each generator s_i, the image $\sigma(s_i)$ is k-uniform. Let $\mathbb{P}_n$ be the uniform distribution on the set $\mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V_n))$ of k-uniform homomorphisms. We will always assume n is chosen so that k divides n (since otherwise there are no k-uniform permutations).

We infer the existence of sofic approximations with the desired properties from the next proposition, which follows immediately from a more precise estimate in [2, Lemma 3.1].

Proposition 1.2. *There are subsets $\Omega_n^{\mathrm{sofic}}$ of $\mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V_n))$ with $\mathbb{P}_n(\Omega_n^{\mathrm{sofic}}) \to 1$ and such that $(\sigma_n)_n$ is a sofic approximation to Γ whenever $\sigma_n \in \Omega_n^{\mathrm{sofic}}$ for all n.*

With Proposition 1.2 in hand, Theorem A is a corollary of the following probabilistic version.

Theorem 1.3. *There are subsets $\Omega_n' \subseteq \Omega_n^{\mathrm{sofic}}$ with $\mathbb{P}_n(\Omega_n') \to 1$ and such that, if $\sigma_n \in \Omega_n'$ for every n, then*

a. *the sofic entropy of (X, m_X, T) along any subsequence of $(\sigma_n)_n$ equals*
$$(1 - d/k)\log 2;$$

b. *every nontrivial factor of (X, m_X, T) inherits positive sofic entropy along $(\sigma_n)_n$.*

Part (b) of this theorem refers to ‘inherited’ sofic entropy. This quantity has previously been referred to as ‘extension’ sofic entropy, ‘outer’ sofic entropy, or ‘sofic entropy in the presence’. The outer Pinsker factor along $(\sigma_n)_n$ is the largest factor of (X, m_X, T) for which this quantity vanishes, so part (b) above asserts that the outer Pinsker factor of (X, m_X, T) along $(\sigma_n)_n$ is trivial. We recall these definitions and discuss our choice of terminology in Subsection 2.3 below.

Remark. It is tempting to summarize part (b) of Theorem 1.3 as ‘(X, m_X, T) has completely positive entropy with high probability according to $\mathbb{P}_n$’. But we must be careful because this summary hides an important detail about the order of quantifiers. For part (b) of Theorem 1.3, the high-probability subsets Ω_n' are found a priori, and then *any* factor of (X, m_X, T) has positive sofic entropy along *any* choice of σ_n from those subsets. One could instead take ‘completely positive entropy with high probability’ to mean that for every factor of (X, m_X, T), there is a sequence of high-probability sets Ω_n' (depending on the factor) within which one sees the desired positive sofic entropy. The latter conclusion is formally weaker, and we do not know whether there exist examples that satisfy one but not the other. Since the collection of all factors may be uncountable, one cannot use a simple diagonal argument to prove that these two formulations are equivalent. (If one identifies factors with their conditional expectation operators, then their induced strong operator topology is separable, but the semi-continuity properties of sofic entropy still do not obviously combine with approximations in this topology to enable a more complicated diagonal argument.)

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Theorem B is also a corollary of a probabilistic assertion about randomly-chosen σ_n and the geometry of the space of microstates. We will prove that the space of microstates shatters in a strong sense, which roughly speaking means that it splits into a union of exponentially many well-separated clusters, each of which has sub-exponential size. The precise definition is given in §5, and the precise statement of the result for the parity-check subshift (Theorem 6.1, part (3)) is formulated in §6.1 below. It was essentially already known from [17] that this phenomenon is incompatible with having a direct Bernoulli factor. We give a formal proof of this incompatibility in §5.

Our focus on the geometry and on shattering is inspired by similar ideas from statistical physics on random sparse graphs [1]. As far as we know, the term 'shattering' first appears in [22]. In other works, this phenomenon is called dynamical replica symmetry breaking [29, 45].

1.5. Outline of the paper

This paper is divided into two parts. In Part I, we consider general symbolic dynamical systems with a focus on the case in which Γ is a free product of finite cyclic groups.

- Section 2 is a review of sofic group theory and sofic entropy.
- Section 3 introduces Kikuchi entropy and annealed entropy for actions of $\Gamma_{d,k}$. These entropies generalize the F-functional and the f-invariant from [10] and are strongly related to the first moment method in statistical physics. This is the main tool for proving Theorem 1.3 part (a).
- Section 4 relates completely positive inherited entropy to a version of uniform mixing for model spaces that we call 'property M'. These general notions provide the background and tools needed to prove Theorem 1.3 part (b).
- Section 5 defines totally shattered microstate spaces and shows how this property prevents both direct Bernoulli factors and weak containment in a Bernoulli shift.

Part II focuses on the parity check sub-shifts which appear in the main theorems:

- Section 6 discusses random (finite) LDPC codes.
- Section 7 proves that for a typical sequence of these codes, the sequence of uniform measures on the codebooks has property M.
- Section 8 proves that typically these uniform measures converge locally and empirically to the Haar measure. Since the sequence also has property M, this gives Theorem 1.3(b) as an application of Theorem 4.3.
- Section 9 proves that the sofic entropy of the Haar measure is $(1-d/k)\log 2$ over most sofic approximations (Theorem 1.3(a)).
- Section 10 proves that the Haar measure typically has totally shattered microstate spaces (Theorem 6.1(3)). This implies Theorem B (the Haar measure has no direct Bernoulli factors) via Corollary 5.5(2).

1.6. Notational conventions

We use the following standard notation for approximate comparison of functions. Let f and g be real-valued functions on the same domain S, and let A be any set of additional parameters in specifying these functions. Then,

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

- In case g is nonnegative, we write $f = O_A(g)$ if there is a positive constant c, depending possibly on A but on nothing else, such that $|f| \le cg$.
- In case g is nonnegative and $S = \mathbb{N}$, we write $f = o_A(g)$ if $g(n)$ is strictly positive for all sufficiently large n and $f(n)/g(n) \to 0$, where this convergence may be bounded by a function that tends to 0 and depends only on A.
- For any S, we write $f \lesssim_A g$ or $g \gtrsim_A f$ if (i) both functions are nonnegative, and (ii) there is a positive constant c, depending possibly on A but on nothing else, such that $f \le cg$. Note that this is similar to writing '$f = O_A(g)$', except here it is part of the assertion that f and g are both nonnegative, in case this is not obvious.

Finally, if $(\Omega_n, \mathbb{P}_n)$ is a sequence of probability spaces indexed by $n \in \mathbb{N}$, we write '$o_p(1)$' as a placeholder for any sequence of random variables X_n on these spaces such that

$$\mathbb{P}_n(|X_n| > \varepsilon) \to 0 \qquad \forall \varepsilon > 0.$$

So this is essentially an analog of $o(1)$ for convergence in probability.

Part I
General systems

2. Preliminaries

2.1. Sofic groups

A **sofic approximation** to a countable group Γ is a sequence $\Sigma = \{\sigma_n\}_{n\in\mathbb{N}}$ of maps

$$\sigma_n : \Gamma \to \mathrm{Sym}(V_n),$$

where V_n are finite sets, $\mathrm{Sym}(V_n)$ is the symmetric group on V_n, and the sequence is required to be asymptotically homomorphic and free in the following sense: For every $g, h \in \Gamma$, we require that the homomorphism equation $\sigma_n(gh)v = \sigma_n(g)\sigma_n(h)v$ holds for asymptotically all $v \in V_n$:

$$1 = \lim_{n\to\infty} |V_n|^{-1} |\{v \in V_n : \ \sigma_n(gh)v = \sigma_n(g)\sigma_n(h)v\}|.$$

For every non-identity element $g \in \Gamma \setminus \{1_\Gamma\}$, we require that the percentage of points fixed by $\sigma_n(g)$ tends to zero:

$$0 = \lim_{n\to\infty} |V_n|^{-1} |\{v \in V_n : \ \sigma_n(g)v = v\}|.$$

A group Γ is **sofic** if it admits a sofic approximation.

If Γ admits a finite generating set S, then it is common to visualize a map σ_n as above in terms of the labeled and directed graph $G(\sigma_n) = (V_n, E_n)$ it induces. The edges of this graph are pairs of the form $(v, \sigma_n(s)v)$ for $s \in S$ and $v \in V_n$, and the label of this pair is s. Then Σ is a sofic approximation precisely when this sequence of graphs $G(\sigma_n)$ Benjamini-Schramm converges to the (labeled and directed) Cayley graph of Γ with respect to S [27].

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

It is an exercise to check that all amenable groups and all residually finite groups are sofic. Because finitely generated linear groups are sofic and direct unions of sofic groups are also sofic, it follows that all linear groups are sofic. It is an open problem whether all countable groups are sofic.

The class of sofic groups was introduced implicitly by Gromov [35] and explicitly by Weiss [73]. For further background on sofic groups, see [59, 21].

2.2. Sofic entropy

This section defines sofic entropy for subshifts using the formulation from [14]. See also [16] or [40] for more comprehensive references.

Let Γ denote a countable group and let $\mathtt{A}$ be a finite set (called the **alphabet**). Let $\mathtt{A}^\Gamma$ be the set of all functions $x : \Gamma \to \mathtt{A}$. We write either x_g or $x(g)$ for the value of x on $g \in \Gamma$, whichever is most convenient. We endow $\mathtt{A}^\Gamma$ with the pointwise convergence topology, under which it is compact and metrizable.

Let $\mathrm{Prob}(\mathtt{A}^\Gamma)$ denote the space of all Borel probability measures on $\mathtt{A}^\Gamma$, which we endow with the weak* topology. In this topology, a sequence of Borel probability measures $(\mu_n)_{n\in\mathbb{N}}$ converges to a measure μ_∞ if and only if for every continuous function $f : \mathtt{A}^\Gamma \to \mathbb{R}$, $\lim_{n\to\infty} \int f \; d\mu_n = \int f \; d\mu_\infty$. An equivalent characterization uses cylinder sets which are defined as follows. Given a finite subset $F \subset \Gamma$ and $x : F \to \mathtt{A}$, let $C(x,F)$ be the set of all functions $y : \Gamma \to \mathtt{A}$ such that $y(f) = x(f)$ for all $f \in F$. Then $(\mu_n)_{n\in\mathbb{N}}$ converges to μ_∞ if and only if for every such F and x, $\lim_{n\to\infty} \mu_n(C(x,F)) = \mu_\infty(C(x,F))$. This is because the cylinder sets $C(x,F)$ form a sub-basis for the topology on $\mathtt{A}^\Gamma$.

Let $T = (T^g)_{g\in\Gamma}$ be the shift action on $\mathtt{A}^\Gamma$ defined by $T^g x(f) = x(g^{-1}f)$ for $x \in \mathtt{A}^\Gamma$. This induces an action on $\mathrm{Prob}(\mathtt{A}^\Gamma)$ by pushforwards. The set of all shift-invariant Borel probability measures on $\mathtt{A}^\Gamma$ will be denoted $\mathrm{Prob}^\Gamma(\mathtt{A}^\Gamma)$. If μ is a shift-invariant Borel probability measure on $\mathtt{A}^\Gamma$, then the system $(\mathtt{A}^\Gamma,\mu,T)$ is called a **shift Γ-system**. We define here the sofic entropy of such systems.

Given $\sigma : \Gamma \to \mathrm{Sym}(V)$, $v \in V$ and $x : V \to \mathtt{A}$, the **pullback name of x at v** is the labeling $\Pi^\sigma_v(x) \in \mathtt{A}^\Gamma$ defined by

$$\Pi^\sigma_v(x)(g) = x_{\sigma(g^{-1})v} \quad \forall g \in \Gamma.$$

For the sake of building some intuition, note that when σ is a homomorphism, the map $v \mapsto \Pi^\sigma_v(x)$ is Γ-equivariant (in the sense that $\Pi^\sigma_{\sigma(g)v}(x) = T^g \Pi^\sigma_v(x)$). In particular, $\Pi^\sigma_v(x) \in \mathtt{A}^\Gamma$ is periodic. In general, we think of $\Pi^\sigma_v(x)$ as an approximate periodic point.

The **empirical measure of** $x : V \to \mathtt{A}$ is

$$P^\sigma_x = |V|^{-1} \sum_{v\in V} \delta_{\Pi^\sigma_v(x)} \in \mathrm{Prob}(\mathtt{A}^\Gamma),$$

where, for $y \in \mathtt{A}^\Gamma$, $\delta_y \in \mathrm{Prob}(\mathtt{A}^\Gamma)$ is the Dirac measure concentrated on $\{y\}$. For example, if σ is a homomorphism, then P^σ_x is a Γ-invariant measure supported on the Γ-orbits of the pullback names $\Pi^\sigma_v(x)$.

Given an open set $\mathcal{O} \subset \mathrm{Prob}(\mathtt{A}^\Gamma)$, a map $x : V \to \mathtt{A}$ is called an **$(\mathcal{O},\sigma)$-microstate** if $P^\sigma_x \in \mathcal{O}$. Typically, we take $\mathcal{O}$ to be a small neighborhood of μ, in which case we consider

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

$(\mathcal{O},\sigma)$-microstates to be 'good microstates for μ'. Let $\Omega(\sigma,\mathcal{O}) \subset \mathtt{A}^V$ denote the set of all $(\mathcal{O},\sigma)$-microstates.

Let $\mu \in \operatorname{Prob}(\mathtt{A}^\Gamma)$ be Γ-invariant and let $\Sigma = (\sigma_i : \Gamma \to \operatorname{Sym}(V_i))_{i\in\mathbb{N}}$ be a sofic approximation to Γ. We say that the system $(\mathtt{A}^\Gamma,\mu,T)$ **has microstates along** Σ if for every neighbourhood $\mathcal{O}$ of μ,

$$\Omega(\sigma_n,\mathcal{O}) \neq \emptyset \quad \text{for all sufficiently large } n.$$

More generally, an arbitrary measure-preserving Γ-system **has microstates along** Σ if every shift-system factor of it has microstates along Σ.

The **Σ-entropy** of the action $(\mathtt{A}^\Gamma,\mu,T)$ is defined by

$$h_\Sigma(\mathtt{A}^\Gamma,\mu,T) := \inf_{\mathcal{O}\ni\mu} \limsup_{i\to\infty} |V_i|^{-1} \log|\Omega(\sigma_i,\mathcal{O})|, \tag{2}$$

where the infimum is over all open neighborhoods of μ in $\operatorname{Prob}(\mathtt{A}^\Gamma)$. We abbreviate this to $h_\Sigma(\mu)$ if the other data are clear from the context. This number depends on the action $(\mathtt{A}^\Gamma,\mu,T)$ only up to measure conjugacy [11]. It therefore defines an invariant for any abstract measure-preserving system that can be represented up to measure conjugacy by a shift system with a finite alphabet or, equivalently, that has a finite generating partition. If the system does not have microstates along any subsequence of Σ, then we declare that the Σ-entropy is $-\infty$.

2.3. Factor maps and inherited entropy

The second part of Theorem A concerns the entropy not only of (X,m_X,T) but also of all its factors. Since X is a shift-invariant subset of $\mathbb{Z}_2^\Gamma$, the Σ-entropy of (X,m_X,T) is an instance of formula (2). But this system may have factors that are not measure conjugate to shift systems, so that formula (2) does not apply.

Sofic entropy can be generalized to measure-preserving systems on standard measurable spaces in various ways: see, for instance, [16, Subsection 2.4] and [3, Subsection 3.1] for formulations and proofs of their equivalence. However, rather than repeat these in detail here, we need only recall how they are controlled by another entropy notion that does permit us to reduce our work to the study of shift systems.

Towards defining this, consider a factor map between two shift systems on finite alphabets, say

$$\Phi : (\mathtt{A}^\Gamma,\mu) \to (\mathtt{B}^\Gamma,\nu), \tag{3}$$

where μ and ν are both invariant under the shift actions of Γ on their respective spaces. Rather than counting good microstates for μ or ν separately, we can ask how many of the good microstates for ν can be lifted to good microstates for μ – that is, how many good microstates ν 'inherits' through the map Φ. To make this precise, consider the graphical joining

$$\lambda := \int_{\mathtt{A}^\Gamma} \delta_{(x,\Phi(x))}\, d\mu(x). \tag{4}$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

This is invariant under the shift action of Γ on $(\mathtt{A}\times\mathtt{B})^{\Gamma}$. Let proj_i be the coordinate projection from $(\mathtt{A}\times\mathtt{B})^{V_i}$ to $\mathtt{B}^{V_i}$. Finally, define the **inherited Σ-entropy of Φ** to be

$$h_{\Sigma}(\mu,T\,;\Phi) := \inf_{\mathcal{O}\ni\lambda}\limsup_{i\to\infty}|V_i|^{-1}\log|\mathrm{proj}_i[\Omega(\sigma_i,\mathcal{O})]|. \tag{5}$$

If Φ is the identity, then this is easily checked to coincide with $h_{\Sigma}(\mu)$.

Like sofic entropy itself, inherited sofic entropy can be generalized to factor maps between arbitrary measure-preserving systems. This general notion is not new, but our use of the term 'inherited' is new. Indeed, the idea behind this quantity is already implicit in Kerr's original approach to defining sofic entropy itself for general measure-preserving systems (see [39] or [16, Subsubsection 2.4.2]). It was formulated and studied explicitly by Hayes in [38, 37], who refers to it as 'entropy in the presence' in recognition of a parallel usage in Voiculescu's theory of free entropy [70]. It has also been studied by other authors under various names, including 'outer sofic entropy' and 'extension sofic entropy'. It is reviewed in [16, Subsection 11.1], which gives more complete references. Despite this history, we do propose the new name 'inherited' entropy because this seems to capture the idea behind the definition better.

Starting from (5), one can define entropy for a factor map between general systems by carefully inserting a supremum over generating partitions of the lower system and an infimum over generating partitions of the upper system. However, the proof that sofic entropy itself is invariant under measure conjugacy can be adapted directly to the quantity in (5), showing that it is an invariant of Φ, where we consider two factor maps to be equivalent if they appear downwards in a commuting square whose horizontal arrows are conjugacies. As a corollary, (5) coincides with the abstract inherited entropy in the case of two shift systems. A more general version of this argument can be found in [38, Proposition 2.9].

One crucial advantage to working with inherited sofic entropy is its monotonicity under factor maps. The following lemma is a special case of parts (iii) and (iv) of [38, Proposition 2.10].

Lemma 2.1. *If*

$$(X,\mu,T)\overset{\Pi}{\to}(Y,\nu,S)\overset{\Phi}{\to}(Z,\theta,R)$$

are factor maps, then

$$h_{\Sigma}(\mu,T\,;\Phi\circ\Pi)\le\min\big\{h_{\Sigma}(\nu,S\,;\Phi),\ h_{\Sigma}(\mu,T\,;\Pi)\big\}.$$

In particular, taking Π (resp. Φ) to be the identity, we obtain

$$h_{\Sigma}(\mu,T\,;\Phi)\le h_{\Sigma}(\mu,T)\quad (\textit{resp.}\quad h_{\Sigma}(\mu,T\,;\Pi)\le h_{\Sigma}(\mu,S)).$$

Inspired by this property, one defines the **outer Σ-Pinsker factor** of a measure-preserving system to be the largest factor whose inherited Σ-entropy is zero. A routine argument shows that a unique maximal such factor exists. Hayes' paper [38] develops this story as well, although it had been considered in unpublished work previously. As a result, the assertion that every nontrivial factor of a measure-preserving system has

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

positive inherited Σ-entropy is equivalent to the assertion that the outer Σ-Pinsker factor of that system is trivial. In addition, by the last inequality in Lemma 2.1, if a system has this property, then it also has completely positive Σ-entropy. The observations together explain our formulation of the second part of Theorem A.

Starting from Lemma 2.1, Hayes develops enough properties of the outer Pinsker factor to turn it into a valuable tool in the study of sofic entropy in general. For instance, these properties are crucial to his proof that a certain large class of systems of algebraic origin all have completely positive sofic entropy [37]. Our reasons for using inherited entropy in the proof of Theorem A are similar, but the details of our proof are essentially disjoint from those of Hayes, and our LDPC system is not among the systems of algebraic origin that he considers in that reference.

Among its useful consequences, Lemma 2.1 gives the following:

Corollary 2.2. *Let (X,μ,T) be a measure-preserving system. If every nontrivial factor map of (X,μ,T) to a shift system has positive inherited Σ-entropy, then every nontrivial factor map of (X,μ,T) has positive inherited Σ-entropy.*

Proof. Let Π be a factor map to another nontrivial system (Y,ν,S). Then (Y,ν) has a nontrivial partition into two measurable subsets. By acting on this partition using S, we define a further factor map of the form

$$\Phi : (Y,\nu,S) \to (\{0,1\}^{\Gamma},\theta,\text{shift}),$$

where θ is not a Dirac measure. Now our hypothesis gives that $h_{\Sigma}(\mu,T\,;\Phi\circ\Pi) > 0$, and this implies that $h_{\Sigma}(\mu,T\,;\Pi) > 0$ by Lemma 2.1. □

Corollary 2.2 can simplify many of the technicalities involved in a proof of completely positive inherited entropy by letting us restrict our attention to factor maps between shift spaces. Given such a map, say

$$\Phi : (\mathtt{A}^{\Gamma},\mu) \to (\mathtt{B}^{\Gamma},\nu),$$

it is uniquely determined by its coordinate at the identity of Γ, which is an arbitrary measurable map $\phi : \mathtt{A}^{\Gamma} \to \mathtt{B}$. Starting from ϕ, we write ϕ^{Γ} for the factor map that it induces, which is given by

$$\phi^{\Gamma}(x)(\gamma) = \phi(\gamma^{-1}\cdot x). \tag{6}$$

When working with such a map Φ, a further simplification is often necessary. If D is a finite subset of Γ, then the map ϕ above is called ***D*-local** if the image $\phi(x)$ depends only on the coordinates of x indexed by D – equivalently, if ϕ factorizes into maps

$$\mathtt{A}^{\Gamma} \to \mathtt{A}^{D} \to \mathtt{B},$$

where the first map is the coordinate projection. We say that ϕ is **local** if it is D-local for some finite set D,and apply the same terminology to the whole of ϕ^{Γ}. This is equivalent to ϕ^{Γ} being a continuous map for the product topologies on our shift spaces.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

If $\mathbf{x} \in \mathtt{A}^V$ is some good microstate for μ over a map $\sigma\colon \Gamma \to \mathrm{Sym}(V)$, we might attempt to send it to a good microstate for ν using the map $\phi^\sigma : \mathtt{A}^V \to \mathtt{B}^V$ defined by

$$\phi^\sigma(\mathbf{x})(v) = \phi(\Pi^\sigma_v \mathbf{x}), \tag{7}$$

since the empirical distribution of $\phi^\sigma(\mathbf{x})$ would then be $\phi^\Gamma_* P^\sigma_{\mathbf{x}}$. Since $P^\sigma_{\mathbf{x}}$ is close to μ, one would hope this would be close to ν. This argument is correct in case ϕ^Γ_* acts continuously on measures, which in turn holds if ϕ is local. In that case, ϕ^σ also has the following form of quantitative continuity.

Lemma 2.3. *Suppose* $\mathtt{A}$,$\mathtt{B}$ *are finite sets,* $\phi\colon \mathtt{A}^\Gamma \to \mathtt{B}$ *is D-local, and* $\sigma\colon \Gamma \to \mathrm{Sym}(V)$*. If* $\mathbf{x}$ *and* $\mathbf{y}$ *are elements of* $\mathtt{A}^V$ *that disagree in exactly one coordinate, then* $\phi^\sigma(\mathbf{x})$ *and* $\phi^\sigma(\mathbf{y})$ *disagree in at most* $|D|$ *coordinates.*

Equivalently, this asserts that ϕ^σ is $|D|$-Lipschitz for the 'normalized Hamming metrics' on $\mathtt{A}^V$ and $\mathtt{B}^V$. This point of view is introduced and used for an application of Lemma 2.3 in Subsection 5.1.

Proof. Suppose that $\mathbf{x}(u) \neq \mathbf{y}(u)$, and let $v \in V$. Since ϕ is D-local, we can have $\phi(\Pi^\sigma_v\mathbf{x}) \neq \phi(\Pi^\sigma_v\mathbf{y})$ only if there exists $\gamma \in D$ such that $\mathbf{x}(\sigma(\gamma^{-1})v) \neq \mathbf{y}(\sigma(\gamma^{-1})v)$, and hence only if u appears in the set $\sigma(D^{-1}) \cdot v$. Since σ is an action by permutations, this holds if and only v lies in $\sigma(D^{-1})^{-1} \cdot u$, and that set has cardinality at most $|D^{-1}| = |D|$. □

The arguments about ϕ^σ above can fail for general measurable factor maps, but this problem can be overcome by approximating these in measure by local factor maps. We say that a sequence of maps

$$\psi_m : \mathtt{A}^\Gamma \to \mathtt{B} \quad (m = 1,2,\dots)$$

is a **local approximating sequence** to ϕ if each ψ_m is local and

$$\mu\{\psi_m \neq \phi\} \to 0. \tag{8}$$

(Note that this notion implicitly also depends on the measure μ.) These are the special case for shift spaces of the 'almost Lipschitz approximating sequences' introduced and used in [3] to study factor maps between more general measure-preserving systems. At a few points in the sequel, we refer to [3] for properties that we need in our special case. For instance, informally, if ψ is a good enough local approximation to ϕ, then ψ^σ sends very good microstates for μ to fairly good microstates for $\phi^\Gamma_*\mu$ [3, Lemma 4.10].

3. Bethe–Kikuchi entropy and annealed calculations

Throughout this section, we set $\Gamma = \Gamma_{d,k} = \mathbb{Z}_k * \cdots * \mathbb{Z}_k$ equal to the free product of d copies of $\mathbb{Z}_k = \mathbb{Z}/k\mathbb{Z}$.

The proof of Theorem A part (a) (and its probabilistic version Theorem 1.3) relies on a first moment calculation. These kinds of computations have a long and interesting history in statistical physics and more recently appeared in the entropy theory of actions of the free group, where, for example, they were used to answer a long-standing open problem on the classification of Bernoulli shifts over free groups [10]. This history is reviewed in §3.3.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

In the next section, we introduce the uniformly random sofic approximation to Γ. This will be our main focus, and we will only obtain results about deterministic sofic approximations indirectly as a consequence of the analysis of these random ones.

In Section 3.2, we introduce Kikuchi entropy, which has historical roots in physics [42] and is a version of the functional F of [10] adapted from the free group to groups Γ which are free products of finite groups. This entropy is a first approximation to the annealed entropy which in recent ergodic-theory research was called the f-invariant [10]. These entropies are key ingredients for proving Theorem 1.3 part (a).

3.1. Random sofic approximations

Instead of directly constructing sofic approximation sequences $\sigma_n : \Gamma \to \mathrm{Sym}(V_n)$, we will construct probability measures on the space of homomorphisms from Γ to $\mathrm{Sym}(kn)$. These measures were used for the same purpose in [2].

Let V be a finite set whose size is divisible by k. Let us say that a permutation of V is k**-uniform** if it consists entirely of k-cycles: that is, its cycle type is $[k^{n/k}]$. Consider a d-tuple of k-uniform permutations $(\sigma(s_1),\dots,\sigma(s_d))$. Then $\sigma(s_i^k)$ equals the identity permutation for each i, and so the tuple $(\sigma(s_i),\dots,\sigma(s_d))$ is the image of the generators $(s_1,\dots,s_d)$ under a homomorphism $\sigma : \Gamma \to \mathrm{Sym}(V)$. We call such a homomorphism k**-uniform** if the images of these generators are all k-uniform permutations. Denote the set of k-uniform homomorphisms into $\mathrm{Sym}(V)$ by $\mathrm{Hom}_{\mathrm{unif}}(\Gamma,\mathrm{Sym}(V))$. Note that for arbitrary members of $\mathrm{Hom}(\Gamma,\mathrm{Sym}(V))$, the cycle sizes of the images of the generators of Γ must be factors of k.

Given a homomorphism $\sigma \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma,\mathrm{Sym}(V))$, consider the collection of all the orbits of the individual maps $\sigma_i = \sigma(s_i)$ – that is, all the subsets of the form

$$\{\sigma(s_i^j)(v) : 0 \le j < k\} \tag{9}$$

for $v \in V$ and $i \in [d]$. Taken together, these may be regarded as a hyper-graph on V. Because we consider a k-uniform homomorphism, this hyper-graph is k-uniform in the usual sense in combinatorics: this is the origin of the terminology. However, it could happen that two σ_i and σ_j give some vertex the same orbit. In this eventuality, it is better to think of the family of sets (9) as a *multi-hyper-graph*, in that we count each hyper-edge with this multiplicity.

For each n, we set $V_n := \{1,2,\dots,n\}$. If k divides n, then we let $\mathbb{P}_n$ be the uniform distribution on $\mathrm{Hom}_{\mathrm{unif}}(\Gamma,\mathrm{Sym}(V_n))$. The sofic approximations that appear in Theorem A are obtained at random from these distributions and are shown to have all the desired properties with high probability.

In order to take this approach, Proposition 1.2 above is a basic prerequisite. It allows us to focus on typical properties of the microstate spaces, knowing that random homomorphisms will be good sofic approximation maps with high probability. It is implied by the more precise estimate in [2, Lemma 3.1], where our $\mathbb{P}_n$ is called '$\mathbb{P}_n^u$'. Note that attention is restricted to even n in that reference, but that restriction is unnecessary for the case of $\mathbb{P}_n^u$; it is included there only for the sake of the other probability distribution $\mathbb{P}_n^p$ that is covered by the same lemma.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

3.2. Formula for Kikuchi entropy

Let $\mathtt{A}$ be a finite set and $\Gamma = (\mathbb{Z}_k)^{*d}$ as above. Given a Γ-invariant measure $\mu \in \mathrm{Prob}(\mathtt{A}^\Gamma)$, we define the **edge weight** $W_\mu(\cdot;\cdot)\colon \mathtt{A}^{\mathbb{Z}_k} \times [d] \to [0,1]$ by

$$W_\mu(\mathbf{a};i) = \mu\{\mathbf{x} \in \mathtt{A}^\Gamma : \mathbf{x}(s_i^j) = \mathbf{a}(j)\ \forall 0 \le j < k\}.$$

Each $W(\cdot;i)$ is a probability measure on $\mathtt{A}^{\mathbb{Z}_k}$ which records the statistics of μ on the hyper-edge $\{s_i, s_i^2, \ldots, s_i^{d-1}\}$. Also define the **vertex weight**

$$W_\mu(\mathtt{a}) = \mu\{\mathbf{x} \in \mathtt{A}^\Gamma : \mathbf{x}(e) = \mathtt{a}\}$$

for $\mathtt{a} \in \mathtt{A}$. This probability measure on $\mathtt{A}$ records the single-site statistics of μ. It is determined by the edge weight $W_\mu(\cdot;\cdot)$.

More generally, an abstract **weight** W is a d-tuple of probability vectors $W(\cdot;1), \ldots, W(\cdot;d)$ on $\mathtt{A}^{\mathbb{Z}_k}$ such that there is a probability vector $W(\cdot)$ on $\mathtt{A}$ satisfying the consistency condition

$$W(\mathtt{a}_0) = \sum_{\mathbf{a} \in \mathtt{A}^{\mathbb{Z}_k} : \mathbf{a}(j) = \mathtt{a}_0} W(\mathbf{a};i)$$

for every $i \in [d]$, $j \in \mathbb{Z}_k$ and $\mathtt{a}_0 \in \mathtt{A}$.

For any weight W, we define the **Kikuchi entropy** of W by

$$\mathrm{H}_{\mathrm{K}}(W) := (1-d)\,\mathrm{H}(W(\cdot)) + \frac{1}{k} \sum_{i \in [d]} \mathrm{H}(W(\cdot;i)),$$

where $\mathrm{H}(\cdot)$ denotes Shannon entropy, and for a Γ-invariant measure μ on $\mathtt{A}^\Gamma$, we abbreviate $\mathrm{H}_{\mathrm{K}}(W_\mu)$ to $\mathrm{H}_{\mathrm{K}}(\mu)$. The functional H_{K} appears in our work because it gives the upper exponential growth rate of the expected number of microstates whose averaged hyper-edge marginals are approximately specified by μ. This is an analog of [9, Theorem 1.4]. Given a microstate $\mathbf{x} \in \mathtt{A}^V$ and $\sigma \in \mathrm{Hom}(\Gamma, \mathrm{Sym}(V))$, let $W_{\mathbf{x},\sigma}$ be the weight corresponding to $P_{\mathbf{x}}^\sigma \in \mathrm{Prob}(\mathtt{A}^\Gamma)$, the empirical distribution of $\mathbf{x}$ over σ. Also, given two weights W, W', let

$$\|W - W'\| := \max_{1 \le i \le d} \max_{\mathbf{a} \in \mathtt{A}^{\mathbb{Z}_k}} |W(\mathbf{a};i) - W'(\mathbf{a};i)|.$$

Proposition 3.1. *We have*

$$\mathrm{H}_{\mathrm{K}}(\mu) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{nk} \log \mathbb{E}_{\sigma_n \sim \mathbb{P}_{nk}} |\{\mathbf{x} \in \mathtt{A}^{nk} : \|W_{\mathbf{x},\sigma_n} - W_\mu\| < \varepsilon\}|.$$

The proof of this proposition shows that we also obtain $\mathrm{H}_{\mathrm{K}}(\mu)$ if $\limsup$ is replaced with $\liminf$ on the right-hand side.

Proof. Let $\mathbb{Z}_k$ act on $\mathtt{A}^{\mathbb{Z}_k}$ in the usual way: $\pi\mathbf{a}(j) = \mathbf{a}(j-\pi)$ for $\pi \in \mathbb{Z}_k$, $\mathbf{a} \in \mathtt{A}^{\mathbb{Z}_k}$ and $j \in \mathbb{Z}_k$. We will say that two elements $\mathbf{a}, \mathbf{b} \in \mathtt{A}^{\mathbb{Z}_k}$ are equivalent if they are in the same $\mathbb{Z}_k$-orbit. Let $[\mathbf{a}] \subset \mathtt{A}^{\mathbb{Z}_k}$ denote the equivalence class of $\mathbf{a}$. For a weight W, write

$$W([\mathbf{a}];i) := \sum_{\mathbf{b} \in [\mathbf{a}]} W(\mathbf{b};i).$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

A weight W is **cyclically invariant** if $W(\mathbf{a};i) = W(\mathbf{b};i)$ for every equivalent pair $\mathbf{a},\mathbf{b}$ in $\mathtt{A}^{\mathbb{Z}_k}$. For example, since μ is Γ-invariant, for each $i \in [d]$, the probability measure $W_\mu(\cdot;i) \in \mathrm{Prob}(\mathtt{A}^{\mathbb{Z}_k})$ is cyclically invariant.

We first calculate $\mathbb{E}|\{\mathbf{x} \in \mathtt{A}^{kn} : W_{\mathbf{x},\sigma_n} = W\}|$ for an arbitrary cyclically invariant weight W with denominator kn (i.e., such that $kn \cdot W([\mathbf{a}];i) \in \mathbb{Z}$ for each $i,\mathbf{a}$). Letting

$$N_n = |\mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(kn))| = \left(\frac{(kn)!}{n!k^n}\right)^d,$$

we have

$$\begin{aligned}\mathbb{E}|\{\mathbf{x} \in \mathtt{A}^{kn} : W_{\mathbf{x},\sigma_n} = W\}| &= \frac{1}{N_n} \sum_{\sigma \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma,\, \mathrm{Sym}(kn))} |\{\mathbf{x} \in \mathtt{A}^{kn} : W_{\mathbf{x},\sigma} = W\}| \\ &= \frac{1}{N_n} \sum_{\mathbf{x} \in \mathtt{A}^{kn}} |\{\sigma \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(kn)) : W_{\mathbf{x},\sigma} = W\}|.\end{aligned}$$

Now a labeling $\mathbf{x} \in \mathtt{A}^{kn}$ admits at least one σ that gives the correct weight W if and only if it has the correct 'vertex statistics'; that is, $\frac{1}{kn}|\{i \in [kn] : \mathbf{x}(i) = \mathtt{a}\}| = W(\mathtt{a})$ for all $\mathtt{a} \in \mathtt{A}$. There are $\exp\{kn(\mathrm{H}(W(\cdot)) + o(1))\}$ such $\mathbf{x}$, and each admits the same number of σ. From now on, fix one such $\mathbf{x}$.

For $i \in [d]$, let G_i be the set of k-uniform permutations $\pi \in \mathrm{Sym}(kn)$ such that if $W(\cdot;\pi)$ is the probability vector on $\mathtt{A}^{\mathbb{Z}_k}$ given by

$$W(\mathbf{a};\pi) = (kn)^{-1}|\{v \in [kn] :\ \mathbf{a}(t) = \mathbf{x}(\pi^t(v))\ \forall 0 \le t < k\}|,$$

then $W(\cdot;\pi) \equiv W(\cdot;i)$. Any k-uniform homomorphism σ with $W_{\mathbf{x},\sigma} = W$ is determined by the permutations $\sigma(s_i)$ which must be in G_i. So the number of such homomorphisms is $\prod_{i=1}^d |G_i|$.

Let $G(\mathbf{x})$ be the set of permutations $g \in \mathrm{Sym}(kn)$ which fix $\mathbf{x}$ in the sense that $\mathbf{x}(v) = \mathbf{x}(gv)$ for all $v \in [kn]$. Observe that $G(\mathbf{x})$ acts on G_i by conjugation. This means that if π_i is a fixed permutation in G_i and $g \in G(\mathbf{x})$, then $g\pi_i g^{-1} \in G_i$. Moreover, this action is transitive. So

$$|G_i| = \frac{|G(\mathbf{x})|}{|\mathrm{Stab}(\pi_i)|}$$

where $\mathrm{Stab}(\pi_i)$ is the set of $g \in G(\mathbf{x})$ with $g\pi_i g^{-1} = \pi_i$. Observe

$$|G(\mathbf{x})| = \prod_{\mathtt{a} \in \mathtt{A}} (kn \cdot W(\mathtt{a}))!.$$

There are two mechanisms by which a $g \in G(\mathbf{x})$ can stabilize π_i. Either g can permute k-cycles with the same labels or it can rotate a given labeled k-cycle. Therefore,

$$|\mathrm{Stab}(\pi_i)| = \prod_{[\mathbf{a}] \in \mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k} (n \cdot W([\mathbf{a}];i))! \left(\frac{k}{|[\mathbf{a}]|}\right)^{n \cdot W([\mathbf{a}];i)}.$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Putting everything together, we get

$$\mathbb{E}|\{\mathbf{x} \in \mathtt{A}^{kn} : W_{\mathbf{x},\sigma_n} = W\}| = \frac{e^{kn(\mathrm{H}(W(\cdot))+o(1))}\left(n!k^n\prod_{\mathtt{a}\in\mathtt{A}}(kn\cdot W(\mathtt{a}))!\right)^d}{(kn)!^d\prod_{i\in[d]}\prod_{[\mathbf{a}]\in\mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k}(n\cdot W([\mathbf{a}];i))!\left(\frac{k}{|[\mathbf{a}]|}\right)^{n\cdot W([\mathbf{a}];i)}}.$$

Applying Stirling's approximation $\log n! = n\log n - n + o(n)$, the logarithm of this is

$$kn\left(\mathrm{H}(W(\cdot)) + d\sum_{\mathtt{a}\in\mathtt{A}} W(\mathtt{a})\log W(\mathtt{a}) - \frac{1}{k}\sum_{i\in[d]}\sum_{[\mathbf{a}]} W([\mathbf{a}];i)\log\frac{W([\mathbf{a}];i)}{|[\mathbf{a}]|}\right) + o(n).$$

By cyclic invariance of each $W(\cdot;i)$, we have $W(\mathbf{a};i) = \frac{W([\mathbf{a}];i)}{|[\mathbf{a}]|}$, so this gives

$$\mathbb{E}|\{\mathbf{x} \in \mathtt{A}^{kn} : W_{\mathbf{x},\sigma_n} = W\}| = e^{nkF(W)+o(n)},$$

where

$$F(W) := (1-d)\mathrm{H}(W(\cdot)) + \frac{1}{k}\sum_{i\in[d]}\mathrm{H}(W(\cdot;i)).$$

Now, since the number of denominator-n weights W grows polynomially in n, it follows that for any $\varepsilon > 0$,

$$\lim_{n\to\infty}\frac{1}{kn}\log\mathbb{E}|\{\mathbf{x}\in\mathtt{A}^{kn} : \|W_{\mathbf{x},\sigma_n} - W_\mu\| < \varepsilon\}| = \sup\{F(W) : \|W - W_\mu\| < \varepsilon\},$$

and taking ε to 0 gives the claimed formula, by continuity of F. □

The **annealed entropy** of μ is defined by

$$\mathrm{h}_{\mathrm{ann}}(\mu) = \inf_{\mathcal{O}\ni\mu}\limsup_{n\to\infty}\frac{1}{nk}\log\mathbb{E}_{\sigma_n\sim\mathbb{P}_{nk}}|\Omega(\sigma_n,\mathcal{O})|,$$

where the infimum is over all open neighborhoods of μ.

To emphasize the relationship between $\mathrm{H}_{\mathrm{K}}(\mu)$ and $\mathrm{h}_{\mathrm{ann}}(\mu)$, let

$$\mathcal{O}_\epsilon(\mu) = \{\nu\in\mathrm{Prob}(\mathtt{A}^\Gamma) : \ \|W_\nu - W_\mu\| < \epsilon\}.$$

Then $\mathcal{O}_\epsilon(\mu)$ is an open neighborhood of μ and Proposition 3.1 becomes

$$\mathrm{H}_{\mathrm{K}}(\mu) = \inf_{\epsilon>0}\limsup_{n\to\infty}\frac{1}{nk}\log\mathbb{E}_{\sigma_n\sim\mathbb{P}_{nk}}|\Omega(\sigma_n,\mathcal{O}_\epsilon(\mu))|.$$

In particular, $\mathrm{h}_{\mathrm{ann}}(\mu) \le \mathrm{H}_{\mathrm{K}}(\mu)$.

The next proposition shows that $\mathrm{H}_{\mathrm{K}}(\mu)$ is an upper bound for sofic entropy with respect to 'most' sofic approximations.

Proposition 3.2. *Let μ be a Γ-invariant Borel probability measure on $\mathtt{A}^\Gamma$. Then there are subsets $\Omega_n' \subseteq \Omega_n^{\mathrm{sofic}}$ with*

$$\lim_{n\to\infty}\mathbb{P}_n(\Omega_n') = 1$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

and such that if $\Sigma=\{\sigma_n\}_{n=1}^{\infty}$ *satisfies* $\sigma_n\in\Omega'_{i_n}$ *for some increasing sequence* $(i_n)_n$ *with* $k\mid i_n$ *for all n, then*

$$h_\Sigma(\mathtt{A}^\Gamma,\mu,T)\le \mathrm{H}_{\mathrm{K}}(\mu).$$

Proof. Given a positive integer n divisible by k, $\varepsilon>0$, and $\sigma:\Gamma\to\mathrm{Sym}(n)$, let $N_{n,\varepsilon}(\sigma)$ be $|\Omega(\sigma,\mathcal{O}_\epsilon(\mu))|$. Think of $N_{n,\varepsilon}$ as a random variable with respect to the uniform measure on the space of k-uniform homomorphisms from Γ to $\mathrm{Sym}(n)$. In addition, let

$$\mathrm{H}_{\mathrm{K},\varepsilon}(\mu):=\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}[N_{n,\varepsilon}],$$

where here and below we always restrict n to be a multiple of k.

For each ε and n as above, let $\Omega'_{n,\epsilon}$ be the set of those $\sigma\in\Omega_n^{\mathrm{sofic}}$ that satisfy $N_{n,\epsilon}(\sigma)\le e^{\sqrt{n}}\mathbb{E}[N_{n,\varepsilon}]$, and let $\Omega'_n:=\Omega'_{n,1}\cap\cdots\cap\Omega'_{n,1/n}$. Then Markov's inequality gives

$$\mathbb{P}_n(\Omega'_n)\ge 1-\sum_{m=1}^{n}\mathbb{P}_n\big(N_{n,1/m}(\sigma)>e^{\sqrt{n}}\mathbb{E}[N_{n,1/m}]\big)\ge 1-ne^{-\sqrt{n}}\to 1.$$

Finally, suppose that $\Sigma=\{\sigma_n\}_{n=1}^{\infty}$ satisfies $\sigma_n\in\Omega'_{i_n}$ for some increasing sequence $(i_n)_n$ with $k\mid i_n$ for all n. Then we also have $\sigma_n\in\Omega'_{i_n,1/m}$ whenever $i_n\ge m$. Fixing m and letting $n\to\infty$, it follows that

$$h_\Sigma(\mathtt{A}^\Gamma,\mu,T)\le\limsup_{n\to\infty}\frac{1}{i_n}\log N_{i_n,1/m}(\sigma_n)\le\limsup_{n\to\infty}\frac{1}{i_n}\log\mathbb{E}[N_{i_n,1/m}]\le\mathrm{H}_{\mathrm{K},1/m}(\mu),$$

where the first inequality uses again the fact that $\mathcal{O}_{1/m}(\mu)$ is an open neighborhood of μ for every positive integer m. Now letting $m\to\infty$, Proposition 3.1 gives $\mathrm{H}_{\mathrm{K}}(\mu)=\lim_{m\to\infty}\mathrm{H}_{\mathrm{K},1/m}(\mu)$, so this completes the proof. □

3.3. A brief history

The most basic method for analyzing the behaviour of a random sofic approximation is the first moment method. Our first indication of the typical number of microstates for (X,m_X,T) over σ_n chosen from $\mathbb{P}_n$ is given by the expectation of that number.

In the analogous setting of actions of free groups, such averages have been studied intensely in recent years. In [9], the exponential growth rate of the expected number of good microstates was shown to coincide with an invariant of systems previously introduced by the second author in [10], where it was used to solve the isomorphism problem for finite-state Bernoulli actions of free groups. In those and several subsequent papers, this invariant was called the ‘f-invariant’. Here, we propose a new term instead: we refer to this quantity as ‘annealed entropy’.

In work of the second author, the f-invariant was obtained as a limit of functionals referred to as F, which are annealed entropies of Markov approximations. As explained farther below, this quantity first appeared in refinements of work of Kikuchi [42]. For this reason, we call it Kikuchi entropy. In later sections, it is used to prove Theorem 1.3.

The reason for the name ‘annealed entropy’ is a connection to statistics and statistical physics. During the last forty years, very similar first-moment calculations for various

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

configurations over large sparse random graphs have become a central feature of the analysis of 'graphical models' in those disciplines. Often, the use of such averages can be seen as a first attempt to find the value for a 'typical' random graph. In such settings, the first moment is referred to as an annealed average: see, for instance, the usage in [50, Section IV.1] or [49, Section 5.4]. Its use as a prediction of typical behaviour is called the 'Bethe ansatz' (or sometimes the 'replica symmetric' approximation in reference to a phenomenology in the study of spin glasses that we do not explain here: see, for instance, [50, Chapter I] or [49, Chapter 8][1]).

In fact, the origins of these quantities lie even farther back in the statistical physics literature. In the general setting for random graphical models studied in statistics, the leading order exponents in first moment calculations are given by quantities called 'Bethe' or 'Kikuchi' entropy.

The first of these terms refers to foundational work by Bethe [8]. He estimated the free energy of a certain model of an alloy on a two-dimensional lattice by a recursive expansion that retained nearest-neighbour interactions but ignored the effect of loops in the lattice graph. A more mathematical description is that the two-dimensional lattice is approximated by an infinite regular tree, and this is why such trees are now often called 'Bethe lattices' in statistical physics.

In [42], Kikuchi expanded on Bethe's ideas by proposing a more careful expansion that respects slightly more of the lattice structure. In modern terms, this can be understood as an approximation to the lattice by a hyper-tree rather than a simple tree. While Bethe argued mostly in terms of free energies, Kikuchi's paper includes various explicit formulae for entropy estimates, and these evolved over time into the quantities studied in statistical inference today. See [42, Equations (A.7) and (C1.6)] for early intimations of these modern formulae. Bethe's and Kikuchi's approximation methods can also be found in physics surveys from closer to that time such as Section III in Burley's contribution [26, Chapter 9].

These formulae were brought explicitly into statistical theory around 2000 by Yedidia and various co-authors in a series of technical reports: see, in particular, [77, 78] and the further references given there. While these references continued to emphasize free energy more than entropy, they do cover both: the explicit formula for Bethe entropy is [77, Formula (1.32)], for example.

The term 'annealed entropy' (instead of 'f-invariant') emphasizes the connection between these two fields. To be more precise, the annealed entropy of a measure-preserving action of a free group is defined as an infimum of the values of a more elementary quantity over Markov approximations to the action. This more elementary quantity, denoted by 'F' in [10, 9], has precisely the same formula as Bethe entropy. Thus, the same formula for an entropy-like quantity was discovered independently and then used for very similar first-moment calculations in both fields.

Bethe entropy has by now become a textbook topic in statistical inference with graphical models: see, for instance, [71, Section 4] or [49, Chapter 14, especially Subsection 14.2.4].

[1] Indeed, the second author has previously also suggested the term 'replica-symmetric entropy' [14, Subsection 7.3], but we feel 'annealed' reflects the general nature of this quantity better.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Some of the theory of these models has also been analyzed rigorously in the probability literature. For example, [24] proves that a first-moment quantity asymptotically agrees with the typical free energy for ferromagnetic Potts models over sparse graph sequences, justifying the ‘Bethe ansatz’ for these models.

Whereas Bethe entropy can be understood as a functional of a probability distribution over a tree, the extension to Kikuchi entropy allows an underlying graph that is a hyper-tree, which is a hyper-graph $G = (V,E)$ such that there is some tree with vertex set V whose subgraphs induced by hyper-edges of G are all connected.

For the first-moment calculations we need below, the exponent is given by the analogous annealed entropy for an action of Γ, the d-fold free power of $\mathbb{Z}_k$, rather than a free group. It turns out that this could again be defined as an infimum of a more elementary quantity over Markov approximations, where now the more elementary quantity is the Kikuchi entropy associated to the Cayley hyper-tree of Γ. We do not work this out completely here. Instead, we give a more direct formula for the annealed entropy and only show that it is bounded above by the Kikuchi entropy (see discussion following Proposition 3.1).

4. Completely positive entropy, local convergence and model mixing

In order to prove the completely positive entropy (CPE) statement in Theorem 1.3, we will use a variant of the main result of [7]. That paper proves that if a model measure sequence locally and empirically converges to the target measure and is uniformly model mixing, then the system is CPE. The local and empirical convergence result needed to prove CPE will also help us establish the lower bound in Theorem 1.3, part (a). We review these concepts here.

As in previous sections, let $\mathtt{A}$ be a finite set, Γ be a countable group and μ be a Γ-invariant Borel probability measure on $\mathtt{A}^\Gamma$. We also let $\Sigma = (\sigma_n)_n$ be a sofic approximation, where $\sigma_n : \Gamma \to \mathrm{Sym}(V_n)$ for each n.

Given a probability measure κ on $\mathtt{A}^{V_n}$ and a vertex $v \in V_n$, the **localization of κ at v** is the probability measure

$$\mathrm{Loc}(\kappa, v) = (\Pi_v^{\sigma_n})_* \kappa = \sum_{\mathbf{x} \in \mathtt{A}^{V_n}} \kappa(\mathbf{x}) \delta_{\Pi_v^{\sigma_n}(\mathbf{x})} \in \mathrm{Prob}(\mathtt{A}^\Gamma).$$

This is the law of the pull-back name of a κ-random sample, as viewed from a fixed $v \in V_n$. This measure depends on the homomorphism σ_n, but we will usually leave that dependence implicit. If we want to specify σ_n, we use the notation $\mathrm{Loc}_{\sigma_n}(\kappa, v)$.

A **model measure sequence** is a sequence $(\mu_n)_n$ of probability measures μ_n on $\mathtt{A}^{V_n}$. The sequence $(\mu_n)_n$ is said to converge to μ **locally and empirically** if for every open neighborhood $\mathcal{O}$ of μ in $\mathrm{Prob}(\mathtt{A}^\Gamma)$,

$$1 = \lim_{n\to\infty} |V_n|^{-1} |\{v \in V_n : \ \mathrm{Loc}(\mu_n, v) \in \mathcal{O}\}|$$
$$1 = \lim_{n\to\infty} \mu_n(\{\mathbf{x} \in \mathtt{A}^{V_n} : \ P_{\mathbf{x}}^{\sigma_n} \in \mathcal{O}\}).$$

Below, we will sometimes refer to the first equality holding for every $\mathcal{O}$ as *local convergence* and the second as *empirical convergence* in order to be explicit about which property is relevant.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

If the measures μ_n and/or the maps σ_n are random with law $\mathbb{P}_n$, then we say the sequence converges locally and empirically in probability to μ if the same limits hold in probability. Explicitly, for every open neighborhood $\mathcal{O}$ of μ and every $\varepsilon > 0$,

$$1 = \lim_{n\to\infty} \mathbb{P}_n \left\{ |V_n|^{-1} |\{v \in V_n : \ \mathrm{Loc}(\mu_n, v) \in \mathcal{O}\}| > 1 - \varepsilon \right\}$$
$$1 = \lim_{n\to\infty} \mathbb{P}_n \left\{ \mu_n(\{\mathbf{x} \in \mathtt{A}^{V_n} : \ P_{\mathbf{x}}^{\sigma_n} \in \mathcal{O}\}) > 1 - \varepsilon \right\}.$$

It will be convenient to reformulate local convergence in probability in terms of total variation distance between marginals. To make this precise, we need notation for the marginals.

Given a finite set $B \subset \Gamma$ and a probability measure ν on $\mathtt{A}^\Gamma$, let ν_B be the probability measure on $\mathtt{A}^B$ equal to the pushforward of ν under the projection map $\mathtt{A}^\Gamma \to \mathtt{A}^B$. This is the marginal of ν on B.

Let d_{TV} denote total variation distance. Because the sets of the form $\mathcal{O}(B,\varepsilon,\mu) = \{\nu \in \mathrm{Prob}(\mathtt{A}^\Gamma) : \ d_{\mathrm{TV}}(\nu_B, \mu_B) < \varepsilon\}$ form a neighborhood basis for the topology at μ, it follows that a sequence of random measures $\mu_n \in \mathrm{Prob}(\mathtt{A}^{V_n})$ converges locally in probability to a measure $\mu \in \mathrm{Prob}(\mathtt{A}^\Gamma)$ if and only if for every finite $B \subset \Gamma$ and $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_n \left\{ \frac{1}{|V_n|} \left| \left\{ v \in V_n : d_{\mathrm{TV}}\big(\mathrm{Loc}(\mu_n, v)_B, \mu_B\big) > \varepsilon \right\} \right| > \varepsilon \right\} = 0. \tag{10}$$

Versions of the next lemma have appeared several times before: for instance, inside the proof of [12, Theorem 4.1], or explicitly as [36, Lemma 5.4] or [3, Corollary 5.7]. We include a proof for completeness.

Lemma 4.1. *If a sequence of random measures $(\mu_n)_n$ converges locally in probability to an ergodic measure $\mu \in \mathrm{Prob}^\Gamma(\mathtt{A}^\Gamma)$ over some random sequence of homomorphisms, then it converges locally and empirically in probability to μ.*

Proof. For each n, let $\theta_n \in \mathrm{Prob}(\mathrm{Prob}^\Gamma(\mathtt{A}^\Gamma))$ denote the law of

$$\frac{1}{|V_n|} \sum_{v \in V_n} \mathrm{Loc}_{\sigma_n}(\mu_n, v),$$

where (σ_n, μ_n) are jointly distributed as given. As stated, θ_n is supported on Γ-imvariant measures because each σ_n is a homomorphism, and therefore, the empirical measure $P_{\mathbf{x}}^{\sigma_n}$ is invariant, for any $\mathbf{x} \in \mathtt{A}^{V_n}$.

Passing to a subsequential limit if necessary, the sequence $(\theta_n)_n$ converges weakly to some $\theta \in \mathrm{Prob}(\mathrm{Prob}^\Gamma(\mathtt{A}^\Gamma))$. We first show the barycenter of θ must be μ: given a continuous function $g \in C(\mathtt{A}^\Gamma)$,

$$\begin{aligned} \iint g(\mathbf{z})\, \nu(d\mathbf{z})\, \theta(d\nu) &= \lim_{n\to\infty} \iint g(\mathbf{z})\, \nu(d\mathbf{z})\, \theta_n(d\nu) \\ &= \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{|V_n|} \sum_{v \in V_n} \int g(\Pi_v^{\sigma_n} \mathbf{x})\, \mu_n(d\mathbf{x}) \right]. \end{aligned}$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Now given $\varepsilon > 0$, let $\mathcal{O} \ni \mu$ be an open neighborhood such that if $\nu \in \mathcal{O}$, then $\int g\, d\nu$ is within ε of $\int g\, d\mu$. Then we control the expectation above by dividing up the terms based on whether μ_n looks like μ near v:

$$\mathbb{E}\left[\frac{1}{|V_n|}\sum_{v\in V_n}\int g(\Pi_v^{\sigma_n}\mathbf{x})\,\mu_n(d\mathbf{x})\right] = \frac{1}{|V_n|}\mathbb{E}\left[\sum_{\substack{v\in V_n\\ \mathrm{Loc}(\mu_n,v)\in\mathcal{O}}}\int g(\Pi_v^{\sigma_n}\mathbf{x})\,\mu_n(d\mathbf{x})\right] + \frac{1}{|V_n|}\mathbb{E}\left[\sum_{\substack{v\in V_n\\ \mathrm{Loc}(\mu_n,v)\notin\mathcal{O}}}\int g(\Pi_v^{\sigma_n}\mathbf{x})\,\mu_n(d\mathbf{x})\right].$$

The magnitude of the second term is bounded by

$$\max|g|\cdot\mathbb{E}\left[\frac{1}{|V_n|}|\{v\in V_n : \mathrm{Loc}(\mu_n,v)\notin\mathcal{O}\}|\right]$$

which goes to 0 as n goes to infinity by definition of local convergence in probability. By choice of $\mathcal{O}$, the first term is within ε of $\int g\, d\mu$ for large n. Since $\varepsilon > 0$ was arbitrary, this proves that the (subsequential) limit must have barycenter μ.

Since m is ergodic, the only possible subsequential limit with barycenter μ is δ_μ, so this is the true limit. This implies that for any $\varepsilon > 0$ and open $\mathcal{O} \ni \mu$,

$$\mathbb{P}\big\{\mu_n(\Omega(\sigma_n,\mathcal{O})) > 1-\varepsilon\big\} \to 1.$$

This is because

$$\mathbb{P}\big\{1-\mu_n(\Omega(\sigma_n,\mathcal{O})) > \varepsilon\big\} \le \frac{\mathbb{E}\big[1-\mu_n(\Omega(\sigma_n,\mathcal{O}))\big]}{\varepsilon} = \frac{1-\mathbb{E}[\mathbb{E}_{\mathbf{x}\sim\mu_n}[\mathbf{1}_{P_{\mathbf{x}}^{\sigma_n}\in\mathcal{O}}]]}{\varepsilon} = \frac{1-\theta_n(\mathcal{O})}{\varepsilon} \to 0$$

using Markov's inequality, the tower law of expectation and the portmanteau theorem. □

4.1. Property M

To define notions of model mixing, we will impose distance functions on the finite sets V which form a given sofic approximation. For this purpose, we will assume Γ is finitely generated and let $E \subset \Gamma$ be a finite symmetric generating subset. For $g \in \Gamma$, let $|g|$ be the word-length of g, which is the length of the shortest word in E representing g. Given $\sigma : \Gamma \to \mathrm{Sym}(V)$, define distance in V by

$$d_\sigma(v,w) = \min\{|g| : \ g \in \Gamma, \sigma(g)v = w\}.$$

If there does not exist g with $\sigma(g)v = w$, then we set $d_\sigma(v,w) = +\infty$. If σ is not a homomorphism, then d_σ may fail to satisfy the triangle inequality.

A subset $S \subset V$ is **r-separated** if $d_\sigma(v,w) > r$ for every pair of distinct $v,w \in S$.

Suppose a model measure sequence $(\mu_n)_n$ converges locally and empirically to μ. We say the sequence is **uniformly model mixing (umm)** if for every finite $F \subset \Gamma$ and every $\epsilon > 0$, there is some $r < \infty$ and a sequence of finite subsets $W_n \subset V_n$ such that

$$|W_n| = (1-o(1))|V_n|,$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

and if $S \subset W_n$ is r-separated, then

$$\mathrm{H}((\mu_n)_{\sigma_n^F(S)}) \geq |S|(\mathrm{H}(\mu_F) - \epsilon),$$

where

- μ_F is the probability measure on $\mathtt{A}^F$ which is the pushforward of μ under the projection map $\mathtt{A}^\Gamma \to \mathtt{A}^F$;
- $\sigma_n^F(S) = \{\sigma_n(f)s : \ f \in F, s \in S\}$;
- $(\mu_n)_{\sigma_n^F(S)}$ is the probability measure on $\mathtt{A}^{\sigma_n^F(S)}$ which is the pushforward of μ_n under the projection map $\mathtt{A}^{V_n} \to \mathtt{A}^{\sigma_n^F(S)}$.

This is a microstates analog of uniform mixing, introduced by Rudolph and Weiss in [62] for actions of an amenable group; see also [74, Definition 10], where the name 'uniform mixing' appears for the first time. The main result of [7] is that if $(\mu_n)_n$ locally and empirically converges to μ and is uniformly model mixing with respect to a fixed sofic approximation Σ, then the system $(\mathtt{A}^\Gamma, \mu, T)$ has completely positive entropy with respect to Σ, in analogy with a corresponding result of [62].

Unfortunately, we do not know whether the parity check sub-shifts of Theorem 1.3 are uniformly model mixing. Instead, we define a weaker version of model mixing which suffices.

Definition 4.2 (Property M). Suppose $(\mu_n)_n$ is a model measure sequence and $\mu \in \mathrm{Prob}^\Gamma(\mathtt{A}^\Gamma)$ is an invariant measure. We say the sequence has **property M** if for every $\epsilon > 0$ and $0 < r < \infty$, there is a sequence of subsets $S_n \subset V_n$ such that

$$\liminf \frac{|S_n|}{|V_n|} > 0$$

and

$$\mathrm{H}((\mu_n)_{\sigma_n^{B_r}(S_n)}) \geq |S_n|(\mathrm{H}(\mu_{B_r}) - \epsilon) \tag{11}$$

for all n, where $B_r = \mathrm{B}(r, e) \subset \Gamma$ denotes the ball of radius r centered at the identity. In applications, μ will be a limit of the sequence $(\mu_n)_n$, but we do not impose any such requirement for the definition.

In contrast with uniform model-mixing, we only require S_n to have asymptotically positive density in V_n, and there is no uniform lower bound on this density across different choices of ϵ, r. This density could be much smaller than $|B_r|^{-1}$, for example. We also do not require the sets S_n to be separated, although the lower bound (11) does usually imply a kind of approximate separation anyway.

We suspect that other variants of uniform model mixing could be used in a similar way to prove completely positive entropy, so Definition 4.2 is not an attempt at optimal generality. This is why we have chosen a rather bland name for Property M, although it is convenient in our work below.

Here is the main result of this section:

Theorem 4.3. *As above, let μ be a Γ-invariant probability measure on $\mathtt{A}^\Gamma$, $\Sigma = (\sigma_n : \Gamma \to \mathrm{Sym}(V_n))_n$ a sofic approximation, and $(\mu_n)_n$ a model measure sequence. Assume $(\mu_n)_n$*

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

converges to μ locally and empirically along Σ and has property M. Then every nontrivial factor of $(\mathtt{A}^{\Gamma},\mu,T)$ inherits positive Σ-entropy.

Proof. By Corollary 2.2, it suffices to consider factor maps to other shift systems. For these, the proof is based on the proof of [7, Theorem 1.2].

Step 1. We start by establishing some general entropy inequalities. For two or more jointly distributed random variables $X_1,\dots,X_k$, define the total correlation

$$\mathrm{TC}(X_1;\cdots;X_k)=\left(\sum_{i=1}^{k}\mathrm{H}(X_i)\right)-\mathrm{H}(X_1,\dots,X_k).$$

This is a generalization of mutual information to more than two random variables, introduced in [72]. It can also be recursively defined by setting $\mathrm{TC}(X_1;X_2)=\mathrm{I}(X_1;X_2)$ and for $k\geq 3$

$$\mathrm{TC}(X_1;\cdots;X_k)=\mathrm{TC}(X_1;\cdots;X_{k-1})+\mathrm{I}(X_1,\dots,X_{k-1};X_k).$$

Using this recursion and the data-processing inequality [23, Theorem 2.8.1], it can be shown by induction on k that if f is any function and $Y_i=f(X_i)$ for each i, then

$$\mathrm{TC}(Y_1;\cdots;Y_k)\leq \mathrm{TC}(X_1;\cdots;X_k). \tag{12}$$

This inequality has also appeared in [6, Lemma 4.3]. Note that the total correlation does not depend on the order in which the random variables are listed. Below, we will refer to the total correlation of a collection of random variables $\{X_i : i\in I\}$ indexed by a finite set I using the notation $\mathrm{TC}(\{X_i : i\in I\})$, since fixing an ordering would unnecessarily complicate notation.

The **Rokhlin distance** between random variables α,β which are defined on the same probability space is defined by $\mathrm{d}^{\mathrm{Rok}}_{\mu}(\alpha,\beta)=\mathrm{H}_{\mu}(\alpha|\beta)+\mathrm{H}_{\mu}(\beta|\alpha)$. This satisfies the triangle inequality, and it equals zero if and only if α and β generate the same partition up to null sets. This distance can be used to control total correlation via the bound

$$\begin{aligned}&|\mathrm{TC}(X_1;\cdots;X_k)-\mathrm{TC}(Y_1;\cdots;Y_k)|\\&\qquad\leq|\mathrm{H}(X_1,\dots,X_k)-\mathrm{H}(Y_1,\dots,Y_k)|+\sum_{i=1}^{k}|\mathrm{H}(X_i)-\mathrm{H}(Y_i)|\\&\qquad\leq 2\sum_{i=1}^{k}\big(\mathrm{H}(X_i|Y_i)+\mathrm{H}(Y_i|X_i)\big)=2\sum_{i=1}^{k}\mathrm{d}^{\mathrm{Rok}}(X_i,Y_i).\end{aligned}$$

Step 2. Since $(\mu_n)_n$ locally and empirically converges to μ, if $S_n\subset V_n$ satisfies $\liminf\frac{|S_n|}{|V_n|}>0$, then

$$\frac{1}{|S_n|}\sum_{v\in S_n}\mathrm{H}\left((\mu_n)_{\sigma_n^{B_r}(v)}\right)=\mathrm{H}(\mu_{B_r})+o(1).$$

So property M implies that for every $r,\varepsilon > 0$, there is a sequence of subsets $S_n \subset V_n$ such that $\liminf \frac{|S_n|}{|V_n|} > 0$ and

$$\frac{1}{|S_n|}\,\mathrm{TC}\left(\{(\mathbf{x}_n)^{\sigma_n^{B_r}(v)} : v \in S_n\}\right) \le \varepsilon$$

for all large n, where $\mathbf{x}_n$ is a random sample of μ_n and the projections $\{(\mathbf{x}_n)^{\sigma_n^{B_r}(v)} : v \in S_n\}$ are jointly distributed in the natural way (from a common sample of $\mathbf{x}_n$).

Step 3. Now let $\phi\colon \mathtt{A}^\Gamma \to \mathtt{B}$ be a measurable map into a finite set $\mathtt{B}$ generating a factor map ϕ^Γ as in (6). If this factor is nontrivial, then $\mathrm{H}_\mu(\phi) > 0$. We want to show that the property M assumption on μ_n implies that

$$h_\Sigma(\mu;\phi^\Gamma) > 0.$$

Let λ be the graphical joining of the factor map ϕ^Γ as in (4).

Fix $r \in \mathbb{N}$ and a $\mathrm{B}(r,e)$-local function $\psi\colon \mathtt{A}^\Gamma \to \mathtt{B}$ which approximates ϕ closely enough in measure that $\mathrm{d}_\mu^{\mathrm{Rok}}(\psi,\phi) < \frac{1}{8}\mathrm{H}_\mu(\phi)$.

Now with $\varepsilon = \frac{1}{8}\mathrm{H}_\mu(\phi)$ and this r, let $(S_n)_n$ be the sequence of subsets of V_n given by property M. Since ψ is $\mathrm{B}(r,e)$-local, the data-processing inequality (12) above implies that

$$\mathrm{TC}\left(\{\psi^{\sigma_n}(\mathbf{x}_n)(v) : v \in S_n\}\right) \le \mathrm{TC}\left(\{(\mathbf{x}_n)^{\sigma_n^{B_r}(v)} : v \in S_n\}\right) \le \varepsilon |S_n|,$$

where $\mathbf{x}_n$ is a random sample of μ_n and $\psi^{\sigma_n}\colon \mathtt{A}^{V_n} \to \mathtt{B}^{V_n}$ is defined as in (7).

Step 4. Let $(\psi_m)_m$ be a local approximating sequence sequence to ϕ, meaning that (8) holds, and hence, $\mathrm{d}_\mu^{\mathrm{Rok}}(\phi,\psi_m)$ also converges to 0. Since the Rokhlin distance satisfies the triangle inequality, there is some $M \in \mathbb{N}$ such that if $m \ge M$, then $\mathrm{d}_\mu^{\mathrm{Rok}}(\psi,\psi_m) < \frac{1}{8}\mathrm{H}_\mu(\phi)$.

Given an open neighborhood $\mathcal{O} \ni \lambda$, by [3, Prop. 4.10], there is some open neighborhood $\mathcal{U} \ni \mu$ and some $m \ge M$ such that, for all large enough n, the map $(\mathrm{id}_{\mathtt{A}^{V_n}}, \psi_m^{\sigma_n})$ sends $\mathcal{U}$-microstates to $\mathcal{O}$-microstates. Let us also assume m is large enough that $\mathrm{H}_\mu(\psi_m) \ge \frac{1}{2}\mathrm{H}_\mu(\phi)$.

Now fix some R such that ψ and ψ_m are both $\mathrm{B}(R,e)$-local. By local convergence of μ_n to μ, for any $\delta > 0$, the fraction of $v \in S_n$ for which the local marginal $\mathrm{Loc}(\mu_n,v)_{B_R}$ is within total variation distance δ of μ_{B_R} is $1 - o(1)$. For the rest of the $v \in S_n$, the term in the sum below has the upper bound $2\log|\mathtt{B}|$:

$$\begin{aligned}
&\tfrac{1}{|S_n|}\left|\mathrm{TC}\left(\{\psi^{\sigma_n}(\mathbf{x}_n)(v) : v \in S_n\}\right) - \mathrm{TC}\left(\{\psi_m^{\sigma_n}(\mathbf{x}_n)(v) : v \in S_n\}\right)\right| \\
&\qquad \le \frac{2}{|S_n|}\sum_{v \in S_n}\left(\mathrm{H}_{\mathrm{Loc}(\mu_n,v)}(\psi|\psi_m) + \mathrm{H}_{\mathrm{Loc}(\mu_n,v)}(\psi_m|\psi)\right) \\
&\qquad \le 2\mathrm{d}_\mu^{\mathrm{Rok}}(\psi_m,\psi) + o(1).
\end{aligned}$$

Hence,

$$\tfrac{1}{|S_n|}\,\mathrm{TC}\left(\{\psi_m^{\sigma_n}(\mathbf{x}_n)(v) : v \in S_n\}\right) \le \frac{3}{8}\mathrm{H}_\mu(\phi) + o(1). \tag{13}$$

By empirical convergence, for large n, the model measure μ_n is mostly supported on $\mathcal{U}$-microstates. So $(\mathrm{id}_{\mathtt{A}^{V_n}}, \psi_m^{\sigma_n})_*\mu_n$ is mostly supported on $\mathcal{O}$-microstates, and using Fano's inequality, we see that

$$\frac{1}{|V_n|}\,\mathrm{H}(\psi_m^{\sigma_n}{}_*\mu_n) \leq \frac{1}{|V_n|}\log|\mathrm{proj}_n[\Omega(\sigma_n,\mathcal{O})]| + o(1).$$

The total correlation bound (13) gives

$$\begin{aligned}\mathrm{H}(\psi_m^{\sigma_n}{}_*\mu_n) = \mathrm{H}(\psi_m^{\sigma_n}(\mathbf{x}_n)) &\geq \mathrm{H}(\{(\psi_m^{\sigma_n}(\mathbf{x}_n)(v) : v \in S_n\}) \\ &\geq \sum_{v\in S_n}\mathrm{H}((\psi_m^{\sigma_n}(\mathbf{x}_n)(v)\}) - |S_n|\big(\tfrac{3}{8}\,\mathrm{H}_\mu(\phi) + o(1)\big).\end{aligned}$$

Since $(\mu_n)_n$ converges locally to μ and ψ_m is a local function,

$$\frac{1}{|S_n|}\sum_{v\in S_n}\mathrm{H}((\psi_m^{\sigma_n}(\mathbf{x}_n)(v)) = \mathrm{H}_\mu(\psi_m) + o(1) \geq \tfrac{1}{2}\,\mathrm{H}_\mu(\phi) + o(1).$$

So

$$\frac{1}{|V_n|}\log|\mathrm{proj}_n[\Omega(\sigma_n,\mathcal{O})]| \geq \frac{|S_n|}{|V_n|}\big(\tfrac{1}{8}\,\mathrm{H}_\mu(\phi) + o(1)\big) + o(1),$$

and for every $\mathcal{O} \ni \lambda$,

$$\liminf_{n\to\infty}\frac{1}{|V_n|}\log|\mathrm{proj}_n[\Omega(\sigma_n,\mathcal{O})]| \geq \left(\liminf_{n\to\infty}\frac{|S_n|}{|V_n|}\right)\tfrac{1}{8}\,\mathrm{H}_\mu(\phi).$$

Since we chose $(S_n)_n$ independently of $\mathcal{O}$, and $\liminf_{n\to\infty}\frac{|S_n|}{|V_n|} > 0$, taking the infimum over $\mathcal{O}$ completes the proof. □

5. Shattering

Let $(\mathtt{A}^\Gamma,\mu,T)$ be a shift Γ-system and $\Sigma = (\sigma_n)_n$ a sofic approximation, where $\sigma_n : \Gamma \to \mathrm{Sym}(V_n)$. On each model space $\mathtt{A}^{V_n}$, we have the normalized Hamming distance defined by

$$\mathrm{d}^{(V_n)}(\mathbf{x},\mathbf{y}) = |V_n|^{-1}|\{v \in V_n :\ \mathbf{x}(v) \neq \mathbf{y}(v)\}|.$$

In this section, we derive ergodic-theoretic consequences from the following phenomenon, which is at the heart of our study of parity check shifts.

Definition 5.1. The shift system has **totally shattered microstate spaces along** Σ if (i) it has microstates along Σ, and (ii) there exists a $\delta > 0$ for which the following holds. For every $\varepsilon > 0$, there exist a weak* neighbourhood $\mathcal{O}$ of μ and a positive integer n_0 such that, for any $n \geq n_0$ and any two microstates $\mathbf{x},\mathbf{y} \in \Omega(\sigma_n,\mathcal{O})$, we have

$$\text{either}\quad \mathrm{d}^{(V_n)}(\mathbf{x},\mathbf{y}) \geq \delta \quad\text{or}\quad \mathrm{d}^{(V_n)}(\mathbf{x},\mathbf{y}) < \varepsilon.$$

We refer to any such δ as a **shatter distance** for the system along Σ.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

5.1. Consequences of shattering

In this section, fix a sofic approximation $\Sigma = (\sigma_n)_n$ by homomorphisms, and assume that $(\mathtt{A}^\Gamma, \mu, T)$ has totally shattered microstate spaces along Σ.

Theorem 5.2. *For every $\varepsilon > 0$, there is a neighbourhood $\mathcal{O}$ of μ for which the following holds. Let $(\mathtt{B}^\Gamma, \nu, T)$ be another shift system that has microstates along Σ, let $(\mathtt{K}^\Gamma, \kappa^\Gamma, T)$ be a Bernoulli shift, and let*

$$\phi : \mathtt{B}^\Gamma \times \mathtt{K}^\Gamma \to \mathtt{A}$$

be a measurable map. If

$$\phi^\Gamma_*(\nu \times \kappa^\Gamma) \in \mathcal{O}, \tag{14}$$

then

$$(\nu \times \kappa^\Gamma \times \kappa^\Gamma)\{(y, z, z') : \ \phi(y, z) \neq \phi(y, z')\} < \varepsilon. \tag{15}$$

Intuitively, the conclusion is that if ϕ^Γ is approximately a factor map onto $(\mathtt{A}^\Gamma, \mu, T)$, then it must be approximately independent from the second coordinate in $\mathtt{B}^\Gamma \times \mathtt{K}^\Gamma$. There are many ways to capture the latter assertion precisely, but (15) turns out to be convenient during the proof.

The proof of Theorem 5.2 has much in common with the main proof in [33]. In that paper, Gamarnik and Sudan used a relative of shattering to prove an a.a.s. upper bound on the maximum size of an independent set on a random regular graph that can be constructed using a local algorithm. The property they use there is now called the 'overlap gap property' and is actually a little weaker than being totally shattered. See [32] for a recent survey. The reference [46, Section 4] explains how the absence of an approximating local algorithm for certain combinatorial problems implies that a resulting limit process is not weakly contained in a Bernoulli shift.

Our proof of Theorem 5.2 needs a couple of known facts about microstate spaces for a product in which one factor is Bernoulli. We recall these as separate lemmas before starting the proof.

Lemma 5.3. *Let $(\mathtt{B}^\Gamma, \nu, T)$ be a shift system and let $(\mathtt{L}^\Gamma, \lambda^\Gamma, T)$ be a Bernoulli shift. Assume that $\mathbf{y}_n \in \mathtt{B}^{V_n}$ is a sequence such that $P^{\sigma_n}_{\mathbf{y}_n} \to \nu$. Then*

$$\lambda^{V_n}\big\{\mathbf{z} : \ P^{\sigma_n}_{(\mathbf{y}_n, \mathbf{z})} \in \mathcal{O}\big\} \to 1$$

for every neighbourhood $\mathcal{O}$ of $\nu \times \lambda^\Gamma$. symbol

Lemma 5.3 is well known as folklore in the study of sofic entropy, and a full proof can be found inside the proofs of some of its consequences in the literature. The earliest and perhaps easiest to extract is inside the proof of the lower bound in [11, Theorem 8.1].

The next lemma is more specialized but was also used in the second author's previous counterexample to the weak Pinsker conjecture for some sofic groups: it is a special case of [17, Proposition 7.9]. It refers to 'hereditary' neighbourhoods of a shift-invariant measure. We do not repeat the definition of these here; the only property we need is that they form a basis for the weak* topology.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Lemma 5.4. *Let $\mathcal{U} \subset \mathcal{V}$ be open neighborhoods of $\nu \times \kappa^\Gamma$ with $\mathcal{U}$ hereditary and $\mathcal{V}$ containing the closure of $\mathcal{U}$. For every $\delta > 0$, if n is large enough, then for every $\mathbf{z},\mathbf{z}' \in \mathtt{K}^{V_n}$ and $\mathbf{y} \in \mathtt{B}^{V_n}$, if $(\mathbf{y},\mathbf{z}), (\mathbf{y},\mathbf{z}')$ are in $\Omega(\sigma_n,\mathcal{U})$, then they are δ-connected within $\Omega(\sigma_n,\mathcal{V})$ (this means there are $\mathbf{w}_1,\ldots,\mathbf{w}_k \in \Omega(\sigma_n,\mathcal{V})$ with $\mathbf{w}_1 = (\mathbf{y},\mathbf{z})$, $\mathbf{w}_k = (\mathbf{y},\mathbf{z}')$ and $d^{(V_n)}(\mathbf{w}_i,\mathbf{w}_{i+1}) < \delta$ for all i). symbol*

Proof of Theorem 5.2. Let δ be a shatter distance for the system along Σ. Let $\varepsilon > 0$ be small enough that $2\varepsilon < \delta$, and now let $\mathcal{O}_1$ and n_0 be a neighbourhood and positive integer as promised by Definition 5.1 for this choice of ε. Lastly, choose a smaller neighbourhood $\mathcal{O}$ of μ whose closure is contained in $\mathcal{O}_1$.

Let $\phi : \mathtt{B}^\Gamma \times \mathtt{K}^\Gamma \to \mathtt{A}$ be a measurable map such that $\phi^\Gamma_*(\nu \times \kappa^\Gamma) \in \mathcal{O}$. In the rest of the proof, we show that (15) holds for this $\mathcal{O}$ and with 3ε in place of ε.

Step 1. Since $\mathtt{A}$ is finite and ϕ is measurable, for any $\eta > 0$, there is an approximating map

$$\psi : \mathtt{B}^\Gamma \times \mathtt{K}^\Gamma \to \mathtt{A}$$

such that

$$(\nu \times \kappa^\Gamma)\{\psi \neq \phi\} < \eta \tag{16}$$

and such that (i) ψ is a local map and (ii) ψ depends on the coordinates in $\mathtt{K}^\Gamma$ only through some finite measurable partition $\mathcal{P}$ of $\mathtt{K}$. If we choose η sufficiently small in terms of $\mathcal{O} \subset \mathcal{O}_1$ and ε, then (14) and (16) imply that $\psi^\Gamma_*(\nu \times \kappa^\Gamma)$ still lies in $\mathcal{O}$, and also (16) implies that the desired conclusion (15) holds for ϕ with error 3ε if it holds for ψ with error 2ε.

By replacing $\mathtt{K}$ with the set of cells $\mathcal{P}$, we have therefore reduced our work to the case when $\mathtt{K}$ is finite and ϕ is F-local for some finite subset F of Γ. We assume this for the rest of the proof and do not refer to ψ again.

Having made these assumptions, let us note that the set

$$\Delta := \{(y,z,z') : \ \phi(y,z) \neq \phi(y,z')\}$$

is closed and open in $\mathtt{B}^\Gamma \times \mathtt{K}^\Gamma \times \mathtt{K}^\Gamma$.

Step 2. Since ϕ is local, pushing forward by the equivariant map ϕ^Γ acts continuously on probability measures. We may therefore choose a neighbourhood $\mathcal{V}$ of $\nu \times \kappa^\Gamma$ such that $\phi^\Gamma_*[\mathcal{V}] \subset \mathcal{O}$. Since σ_n is a homomorphism, we have $P^{\sigma_n}_{\phi^{\sigma_n}(\mathbf{x})} = \phi^\Gamma_* P^{\sigma_n}_{\mathbf{x}}$, so this implies that

$$\phi^{\sigma_n}[\Omega(\sigma_n,\mathcal{V})] \subset \Omega(\sigma_n,\mathcal{O}) \quad \text{for every } n. \tag{17}$$

In addition, since hereditary neighbourhoods form a basis, we may let $\mathcal{U}$ be a hereditary neighbourhood of $\nu \times \kappa^\Gamma$ whose closure is contained in $\mathcal{V}$.

Step 3. By assumption, there is a sequence $\mathbf{y}_n \in \mathtt{B}^{V_n}$ such that $P^{\sigma_n}_{\mathbf{y}_n} \to \nu$. Fix this for the rest of the proof.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

For each n, consider $\mathbf{z}_n$ and $\mathbf{z}'_n \in \mathtt{K}^{V_n}$ drawn independently at random according to κ^{V_n}, and consider the event

$$E := \big\{(\mathbf{z}_n,\mathbf{z}'_n) \in \mathtt{K}^{V_n} \times \mathtt{K}^{V_n} : \ (\mathbf{y}_n,\mathbf{z}_n),(\mathbf{y}_n,\mathbf{z}'_n) \in \Omega(\sigma_n,\mathcal{U})\big\}.$$

By Lemma 5.3, we have

$$(\kappa^{V_n} \times \kappa^{V_n})(E) \to 1. \tag{18}$$

When E occurs, we can draw the following additional conclusions:

- The points $(\mathbf{y}_n,\mathbf{z}_n)$ and $(\mathbf{y}_n,\mathbf{z}'_n)$ are $(\delta/|F|)$-connected within $\Omega(\sigma_n,\mathcal{V})$ for all sufficiently large n (not depending on the specific values of $\mathbf{y}_n$, $\mathbf{z}_n$ or $\mathbf{z}'_n$), by Lemma 5.4.
- Since ϕ is F-local, the map ϕ^{σ_n} is $|F|$-Lipschitz for the normalized Hamming metrics (Lemma 2.3), so $\phi^{\sigma_n}(\mathbf{y}_n,\mathbf{z}_n)$ and $\phi^{\sigma_n}(\mathbf{y}_n,\mathbf{z}'_n)$ both lie in $\Omega(\sigma_n,\mathcal{O})$ and are δ-connected within that set, by the previous conclusion and (17).
- By total shattering and our choice of $\mathcal{O}$, we can now deduce that
$$\mathrm{d}^{(V_n)}\big(\phi^{\sigma_n}(\mathbf{y}_n,\mathbf{z}_n),\phi^{\sigma_n}(\mathbf{y}_n,\mathbf{z}'_n)\big) < \varepsilon \tag{19}$$
for all sufficiently large n. Indeed, if n is large enough and
$$\phi^{\sigma_n}(\mathbf{y}_n,\mathbf{z}_n) = \mathbf{w}_1, \ \ldots, \ \mathbf{w}_l = \phi^{\sigma_n}(\mathbf{y}_n,\mathbf{z}'_n)$$
is a sequence in $\Omega(\sigma_n,\mathcal{O})$ with all consecutive distances less than δ, then total shattering implies that all of these distances are actually less than ε. Then the triangle inequality implies that $\mathrm{d}^{(V_n)}(\mathbf{w}_1,\mathbf{w}_3) < 2\varepsilon$, which is still less than δ. We may therefore invoke total shattering again to conclude that in fact $\mathrm{d}^{(V_n)}(\mathbf{w}_1,\mathbf{w}_3) < \varepsilon$. Now a simple induction shows that in fact $\mathrm{d}^{(V_n)}(\mathbf{w}_1,\mathbf{w}_i) < \varepsilon$ for every i, giving (19) when $i = l$.

Unpacking the definitions of normalized Hamming metric and empirical distribution, the left-hand side of (19) is equal to

$$\frac{1}{|V_n|}\sum_{v\in V_n} 1_{\{\phi(\Pi_v^{\sigma_n}\mathbf{y}_n,\Pi_v^{\sigma_n}\mathbf{z}_n)\neq\phi(\Pi_v^{\sigma_n}\mathbf{y}_n,\Pi_v^{\sigma_n}\mathbf{z}'_n)\}} = P^{\sigma_n}_{(\mathbf{y}_n,\mathbf{z}_n,\mathbf{z}'_n)}(\Delta).$$

Therefore, in view of the conclusions above, (18) implies that

$$(\kappa^{V_n} \times \kappa^{V_n})\big\{P^{\sigma_n}_{(\mathbf{y}_n,\mathbf{z}_n,\mathbf{z}'_n)}(\Delta) < \varepsilon\big\} \to 1. \tag{20}$$

However, since Δ is a closed and open set, another appeal to Lemma 5.3 (this time for the larger product $\nu \times \kappa^{\Gamma} \times \kappa^{\Gamma}$) shows that

$$(\kappa^{V_n} \times \kappa^{V_n})\big\{P^{\sigma_n}_{(\mathbf{y}_n,\mathbf{z}_n,\mathbf{z}'_n)}(\Delta) > (\nu \times \kappa^{\Gamma} \times \kappa^{\Gamma})(\Delta) - \epsilon\big\} \to 1. \tag{21}$$

The limits (20) and (21) can hold simultaneously only if

$$(\nu \times \kappa^{\Gamma} \times \kappa^{\Gamma})(\Delta) < 2\varepsilon.$$

Since ε was arbitrary, this completes the proof. □

Theorem 5.2 has several more streamlined consequences.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Corollary 5.5. *Suppose that $(\mathtt{A}^\Gamma,\mu,T)$ has totally shattered microstate spaces along Σ.*

1. *If (Y,θ,S) is any probability-preserving Γ-system that has microstates along Σ (recall the definition of this property from Subsection 2.2), and $(\mathtt{K}^\Gamma,\kappa^\Gamma,T)$ is a Bernoulli shift, then any factor map*
$$(Y\times\mathtt{K}^\Gamma,\theta\times\kappa^\Gamma,S\times T)\to(\mathtt{A}^\Gamma,\mu,T)$$
factorizes through the coordinate projection $Y\times\mathtt{K}^\Gamma\to Y$ up to agreement a.e.
2. *The system $(\mathtt{A}^\Gamma,\mu,T)$ has no nontrivial direct Bernoulli factors.*
3. *The system $(\mathtt{A}^\Gamma,\mu,T)$ is not weakly contained in a Bernoulli shift unless it is actually a trivial system, meaning that $\mu=\delta_{(\dots,a,a,\dots)}$ for some $a\in\mathtt{A}$.*

Proof. *Part 1.* Let the factor map in question be ξ^Γ for some measurable map $\xi:Y\times\mathtt{K}^\Gamma\to\mathtt{A}$. Let $\varepsilon>0$, and let $\mathcal{O}$ be the neighbourhood of μ given by Theorem 5.2 for this ε.

By approximating the level sets of ξ by measurable rectangles, for any $\eta>0$, we can choose (i) a factor map π from (Y,θ,S) to a finite-alphabet shift system $(\mathtt{B}^\Gamma,\nu,T)$ and (ii) a map $\phi:\mathtt{B}^\Gamma\times\mathtt{K}^\Gamma\to\mathtt{A}$ such that
$$(\theta\times\kappa^\Gamma)\{(y,z):\ \phi(\pi(y),z)\neq\xi(y,z)\}<\eta. \tag{22}$$
Provided we choose η small enough, the inequality (22) implies that
$$\phi_*^\Gamma(\nu\times\kappa^\Gamma)\in\mathcal{O}.$$
By shrinking further if necessary, we may also assume that $\eta<\varepsilon$.

Fix a choice of η with these properties, and consider the resulting maps π and ϕ. Since (Y,θ,S) has microstates along Σ, so does its factor $(\mathtt{B}^\Gamma,\nu,T)$. By our choice of η, we have arranged that this factor system and the map ϕ satisfy (14). Therefore, Theorem 5.2 tells us that they also satisfy (15). Combined with (22), this gives
$$\begin{aligned}(\theta\times\kappa^\Gamma\times\kappa^\Gamma)&\{(y,z,z'):\ \xi(y,z)\neq\xi(y,z')\}\\ &\le 2\eta+(\nu\times\kappa^\Gamma\times\kappa^\Gamma)\{(y_1,z,z'):\ \phi(y_1,z)\neq\phi(y_1,z')\}<2\eta+\varepsilon<3\varepsilon.\end{aligned}$$
Since ε is arbitrary, the left-hand side here must actually equal 0, so in fact,
$$\xi(y,z)=\xi(y,z')\quad\text{for }(\theta\times\kappa^\Gamma\times\kappa^\Gamma)\text{-a.e. }(y,z,z').$$
Therefore, possibly after adjusting on a set of measure zero, ξ depends on only the first coordinate in $Y\times\mathtt{K}^\Gamma$.

Part 2. If
$$\xi^\Gamma:(Y,\theta,S)\times(\mathtt{K}^\Gamma,\kappa^\Gamma,T)\to(\mathtt{A}^\Gamma,\mu,T)$$
is an isomorphism, then the system (Y,θ,S) is a factor of $(\mathtt{A}^\Gamma,\mu,T)$, and so it inherits the property of having microstates along Σ. Therefore, Part 1 shows that, up to agreement a.e., ξ depends on only the first coordinate in $Y\times\mathtt{K}^\Gamma$. Since ξ^Γ is an isomorphism, this is possible only if the Bernoulli factor $(\mathtt{K}^\Gamma,\kappa^\Gamma,T)$ is trivial.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Part 3. Towards a contradiction, suppose that $(\mathtt{K}^\Gamma, \kappa^\Gamma, T)$ is a Bernoulli shift and that $\phi_m : \mathtt{K}^\Gamma \to \mathtt{A}$ is a sequence of measurable maps such that

$$\phi^\Gamma_{m*}\kappa^\Gamma \to \mu. \tag{23}$$

Apply Theorem 5.2 by inserting a trivial one-point system in the place of $(\mathtt{B}^\Gamma, \nu, T)$. The conclusion is that

$$(\kappa^\Gamma \times \kappa^\Gamma)\{(z,z') : \ \phi_m(z) \neq \phi_m(z')\} \to 0.$$

However, this implies that the distribution of ϕ_m is converging to δ_a for some $a \in \mathtt{A}$. Combined with (23), it follows that $\mu = \delta_{(\ldots,a,a,\ldots)}$. □

5.2. Shattering and Bernoulli splittings

Suppose the shift system $(\mathtt{A}^\Gamma, \mu, T)$ has totally shattered microstate spaces along Σ. Note as soon as $\varepsilon < \delta/2$, if $\mathcal{O}$ is the neighborhood of μ given by the definition, then the relation '$\mathrm{d}^{(V_n)} < \delta$' restricted to the microstate space $\Omega(\sigma_n, \mathcal{O})$ is an equivalence relation, so it partitions the microstate space into small-diameter, well-separated clusters.

In this section, we give a second proof of Corollary 5.5(2) and sketch a third. Both of these approaches are based on the clusters of microstates: one uses the number of clusters, which is one of the sofic homological invariants introduced in [17], and the other uses the sizes of clusters. Both can be compared to the use of the 'overlap gap property' introduced in [33] to prove an a.a.s. upper bound on the size of an independent set on a random regular graph, when the independent set is required to be constructed from a local algorithm.

The size of clusters can be used as follows: if a direct Bernoulli factor exists, it can be shown that its entropy rate is a uniform lower bound for the exponential size of all microstate clusters, while if a system has totally shattered microstate spaces, then all its clusters are of subexponential size. This implies Corollary 5.5(2).

In the rest of the section, we give a proof using the number of clusters. First, we give relevant definitions.

If $\mathcal{O}$ is a subset of a metric space and $x, y \in \mathcal{O}$, we say x, y are **δ-connected within** $\mathcal{O}$ if there is a sequence of points $z_1, \ldots, z_l \in \mathcal{O}$ with $z_1 = x$, $z_l = y$, and $\mathrm{d}(z_i, z_{i+1}) < \delta$ for each i. This defines an equivalence relation on $\mathcal{O}$, which we denote $\mathrm{cl}(\mathcal{O}, \delta)$.

Given a map $\sigma : \Gamma \to \mathrm{Sym}(V)$, a labeling $\mathbf{x} \in \mathtt{A}^V$, an open set $\mathcal{O} \subset \mathrm{Prob}(\mathtt{A}^\Gamma)$, and $\delta > 0$, let

$$\mathrm{cl}(\sigma, \mathbf{x}, \mathcal{O}, \delta) = [x]_{\mathrm{cl}(\Omega(\sigma,\mathcal{O}),\delta)} = \{\mathbf{z} \in \mathtt{A}^V : \mathbf{z} \text{ is } \delta\text{-connected to } \mathbf{x} \text{ within } \Omega(\sigma, \mathcal{O})\},$$

where $\Omega(\sigma, \mathcal{O}) \subset \mathtt{A}^V$ has the normalized Hamming metric $\mathrm{d}^{(V)}$. In particular, $\mathrm{cl}(\sigma, \mathbf{x}, \mathcal{O}, \delta) \subset \Omega(\sigma, \mathcal{O})$.

Now if in addition to the above we have some $\mathcal{O}' \subset \mathcal{O}$, the quotient

$$\Omega(\sigma_n, \mathcal{O}') \Big/ \mathrm{cl}(\Omega(\sigma_n, \mathcal{O}), \delta)$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

is the set of clusters of $\mathcal{O}'$-microstates that are δ-connected within $\Omega(\sigma_n,\mathcal{O})$. We define

$$b_{0,\Sigma}(\mu) = \sup_{\mathcal{O}\ni\mu}\sup_{\delta>0}\inf_{\mu\in\mathcal{O}'\subset\mathcal{O}}\limsup_{n\to\infty}\frac{1}{|V_n|}\log\left|\Omega(\sigma_n,\mathcal{O}')\big/\mathrm{cl}(\Omega(\sigma_n,\mathcal{O}),\delta)\right|.$$

Informally, we first set coarseness parameters $\mathcal{O},\delta$ which divide microstate spaces into clusters. We consider the exponential growth rate of the number of these clusters which contain $\mathcal{O}'$-good microstates for μ. At first, it might seem more natural to consider directly the growth rate of the number of δ-connected clusters within a single $\Omega(\sigma_n,\mathcal{O})$:

$$\limsup_{n\to\infty}\frac{1}{|V_n|}\log\left|\Omega(\sigma_n,\mathcal{O})\big/\mathrm{cl}(\Omega(\sigma_n,\mathcal{O}),\delta)\right|,$$

then take δ to 0 and in some sense $\mathcal{O}$ to μ. But there is no monotonicity in $\mathcal{O}$: shrinking the neighborhood of μ removes some microstates from $\Omega(\sigma_n,\mathcal{O})$, but this can both remove some clusters and/or break a cluster into multiple pieces. Considering pairs $\mathcal{O}'\subset\mathcal{O}$ is one natural way around this. See also the discussion of a related definition of connected model spaces in [4, Section 2.2].

In [17], $b_{0,\Sigma}(\mu)$ is called the 0th Betti number of μ. If X is totally disconnected and $\mu\in\mathrm{Prob}^{\Gamma}(X^{\Gamma})$, then $b_{0,\Sigma}(\mu)$ is a measure-conjugacy invariant [17, Corollary 4.2].

It follows directly from the definition that $b_{0,\Sigma}(\mu)\le \mathrm{h}_{\Sigma}(\mu)$ (a similar inequality holds for higher-dimensional sofic homology theories [17, Lemma 7.13]).

Lemma 5.6. *If* $(\mathtt{A}^{\Gamma},\mu)$ *has totally shattered microstate spaces over* Σ*, then* $b_{0,\Sigma}(\mu)=\mathrm{h}_{\Sigma}(\mu)$.

Proof. Recall that in general $b_{0,\Sigma}(\mu)\le\mathrm{h}_{\Sigma}(\mu)$, so we only have to prove that having totally shattered microstate spaces implies the reverse inequality.

Let $\delta>0$ be as in the definition of totally shattered microstate spaces. Given $\varepsilon<\delta/2$, there exists a neighborhood U of μ such that for all large n, every $\mathbf{x},\mathbf{y}\in\Omega(\sigma_n,U)$ have $\mathrm{d}^{(V_n)}(\mathbf{x},\mathbf{y})\in[0,\varepsilon)\cup[\delta,\infty)$. In particular, for every $\mathbf{x}\in\Omega(\sigma_n,U)$,

$$\mathrm{cl}(\sigma_n,\mathbf{x},U,\varepsilon)\subseteq\mathrm{B}(\varepsilon,\mathbf{x}),$$

where $\mathrm{B}(\varepsilon,\mathbf{x})$ is the radius-ε ball around $\mathbf{x}$. Thus,

$$|\mathrm{cl}(\sigma_n,\mathbf{x},U,\varepsilon)|\le|\mathrm{B}(\varepsilon,\mathbf{x})|\le\exp\big(|V_n|(\mathrm{H}(\varepsilon)+\varepsilon\log|\mathtt{A}|+o(1))\big).$$

This leads to a lower bound on the number of clusters for all $\mathcal{O}'\subset U$:

$$\left|\Omega(\sigma_n,\mathcal{O}')\big/\mathrm{cl}(\Omega(\sigma_n,U),\varepsilon)\right|\ge|\Omega(\sigma_n,\mathcal{O}')|\exp\big(-|V_n|(\mathrm{H}(\varepsilon)+\varepsilon\log|\mathtt{A}|+o(1))\big).$$

Hence,

$$\inf_{\mu\in\mathcal{O}'\subset U}\limsup_{n\to\infty}\frac{1}{|V_n|}\log\left|\Omega(\sigma_n,\mathcal{O}')\big/\mathrm{cl}(\Omega(\sigma_n,U),\varepsilon)\right|\ge\mathrm{h}_{\Sigma}(\mu)-\big(\mathrm{H}(\varepsilon)+\varepsilon\log|\mathtt{A}|\big),$$

and taking the supremum over $\varepsilon>0$ and $U\ni\mu$ completes the proof. □

Alternate proof of Corollary 5.5(2). To obtain a contradiction, suppose that $(\mathtt{A}^{\Gamma},\mu,T)$ is measurably conjugate to the direct product of a nontrivial Bernoulli shift $(\mathtt{K}^{\Gamma},\kappa^{\Gamma},T)$ with another shift system $(\mathtt{B}^{\Gamma},\nu,S)$ where $\mathtt{B}$ is a compact metrizable space. We may assume

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

$\mathtt{B}$ is totally disconnected without loss of generality because any dynamical system is measurably conjugate to a system of this form; this assumption is used in results of [17] cited below.

Sofic entropy is additive under taking direct products with a Bernoulli shift [11, Theorem 8.1], so

$$\mathrm{h}_\Sigma(\mathtt{A}^\Gamma,\mu,T) = \mathrm{H}(\mathtt{K},\kappa) + \mathrm{h}_\Sigma(\mathtt{B}^\Gamma,\nu,S) > \mathrm{h}_\Sigma(\mathtt{B}^\Gamma,\nu,S). \tag{24}$$

Theorem 7.8 of [17] implies that the 0-dimensional sofic homology theories of $(\mathtt{A}^\Gamma,\mu,T)$ and $(\mathtt{B}^\Gamma,\nu,S)$ are equivalent. In particular, this implies that the exponential rate of growth of the number of clusters in the microstate spaces of the two actions are the same. In the notation of [17], this means

$$b_{0,\Sigma}(\mathtt{A}^\Gamma,\mu,T) = b_{0,\Sigma}(\mathtt{B}^\Gamma,\nu,S) \le \mathrm{h}_\Sigma(\mathtt{B}^\Gamma,\nu,S),$$

where the last inequality holds by [17, Lemma 7.13].

Because $(\mathtt{A}^\Gamma,\mu,T)$ has totally shattered microstate spaces, by Lemma 5.6, we have $b_{0,\Sigma}(\mathtt{A}^\Gamma,\mu,T) = \mathrm{h}_\Sigma(\mathtt{A}^\Gamma,\mu,T)$. Combined with the previous inequality, this implies

$$\mathrm{h}_\Sigma(\mathtt{A}^\Gamma,\mu,T) = b_{0,\Sigma}(\mathtt{A}^\Gamma,\mu,T) \le \mathrm{h}_\Sigma(\mathtt{B}^\Gamma,\nu,S),$$

which contradicts (24). This contradiction finishes the proof.

PART II
Parity check subshifts

In this part, we fix natural numbers d,k with $k > d \ge 3$ and let $\Gamma = \Gamma_{d,k}$ be the d-fold free product of order-k cyclic groups:

$$\Gamma := \langle s_1,\ldots,s_d :\ s_1^k = \cdots = s_d^k = e\rangle = \underbrace{\mathbb{Z}_k * \cdots * \mathbb{Z}_k}_{d}.$$

Let $X \le \mathbb{Z}_2^\Gamma$ be the closed subgroup defined by

$$X = \left\{ x \in \mathbb{Z}_2^\Gamma :\ \sum_{j=0}^{k-1} x_{gs_i^j} = 0\ \forall g \in \Gamma,\ i = 1,\ldots,d \right\},$$

and let $\mu = m_X$ be the Haar probability measure on X.

If F is any subset of Γ, let X_F be the image of X under the coordinate projection map $\mathbb{Z}_2^\Gamma \to \mathbb{Z}_2^F$, and let m_F be the pushforward of m_X under this projection. We mostly use these notations when F is B_r, the ball of radius r centered at the identity in the Cayley graph of Γ, where we use $\{s_i^j :\ 1 \le i \le d, 1 \le j \le k-1\}$ for the generating set.

6. LDPC codes and measures on microstate spaces

6.1. The use of LDPC codes

Our proofs of the main theorems are considerably simplified by using the special structure of the system as a subgroup of $\mathbb{Z}_2^\Gamma$. We largely do so via the corresponding finitary codes

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

constructed over the sofic approximations. Given a k-uniform homomorphism $\sigma : \Gamma_{d,k} \to \mathrm{Sym}(V)$, let

$$X_\sigma := \left\{ \mathbf{x} \in \mathbb{Z}_2^V : \sum_{j=0}^{k-1} x_{\sigma_i^j(v)} = 0 \ \forall v \in V,\ i = 1,\dots,d \right\}.$$

Let μ_σ be the uniform distribution on X_σ.

These are the obvious finitary analogs of X and m_X themselves. It turns out that these finitary constructs can be used as better and better approximations to the infinitary system and measure, and their linear structure makes them easier to analyze than the 'looser' sets $\Omega(\sigma, U)$ that are used to define sofic entropy.

In fact, sets such as X_σ are classical objects in coding theory. They are linear codes over the field $\mathbb{Z}_2$, each defined by a collection of parity-check constraints. Since each of those parity-checks involves only k vertices, and k is fixed as n grows, these are examples of low-density parity-check ('LDPC') codes. Such codes were introduced in Gallager's PhD thesis [31, 30]. After many years of relative neglect, they were rediscovered independently by MacKay in the late 1990s [47], and they are now a textbook family of codes with desirable properties. A good basic reference for their theory is [48, Chapter 47], and a more dedicated treatment is [61]. In fact, the essence of the parity-check subshift X itself already appears in those sources too, playing the role of an 'idealized limit code' on which to investigate the performance of local decoding algorithms: see, for instance, [48, Figure 47.11] and the discussion around it.

6.2. Outline of the rest of the paper

Our use of the measures μ_σ to analyze the system (X, m_X, T) rests on the following main results. As before, we confine the index n to multiples of k. Abbreviate X_{σ_n} to X_n and μ_{σ_n} to μ_n.

Recall that $\mathbb{P}_n$ is the uniform probability measure on $\mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(n))$ which is the set of k-uniform homomorphisms $\sigma : \Gamma \to \mathrm{Sym}(n)$.

Theorem 6.1. *There are subsets*

$$\Omega_n' \subseteq \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V_n))$$

such that $\mathbb{P}_n(\Omega_n') \to 1$ and the following holds. If $\sigma_n \in \Omega_n'$ for each n, then

1. $(\mu_n)_n$ *has property M;*
2. $(\mu_n)_n$ *converges locally and empirically to m_X;*
3. (X, m_X, T) *has totally shattered microstate spaces along $\Sigma = (\sigma_n)_n$.*

Our proof of Theorem 6.1 relies on the linear structure of the codes X_σ.

In the next few sections, we introduce notation to formulate Proposition 7.1 which, roughly speaking, rules out near-cancellations among parity-check words of a typical $\sigma_n \sim \mathbb{P}_n$. Section §7.1 proves Theorem 6.1(1) from Proposition 7.1. The rest of §7 proves Proposition 7.1.

Item (2) of Theorem 6.1 is proven in §8. Its proof relies on item (1). Item (3) is proven in §10. Its proof does not refer to items (1) or (2). Theorem B is an immediate consequence of item (3) and Corollary 5.5(2).

Theorem 1.3 is proven in §9. Part (b) of that theorem follows readily from items (1) and (2) of Theorem 6.1 together with Theorem 4.3. Part (a) (which computes the sofic entropy value) uses item (2) and the Bethe-Kikuchi entropy theory of §3.

6.3. Random factor graphs

In this subsection, fix a size n that is divisible by k and suppress it from the notation: thus, for instance, $\mathbb{P}$ stands for $\mathbb{P}_n$. Most of our work towards Theorem 6.1 consists of estimates of various probabilities under $\mathbb{P}$.

Several of these estimates involve a sum or union bound over possible subfamilies of the set of all hyper-edges of the form (9). We need to be able to move the sum outside an expectation, and for this purpose, the sum must be over a range which is fixed, not random. For this reason, it is convenient to augment the information in σ_n with a labeling of the family of hyper-edges (9) by a fixed index set.

Let $E_1, \ldots, E_d$ be disjoint sets, each of size n/k, and let $E := E_1 \cup \cdots \cup E_d$. Taking some terminology from coding theory, we refer to the elements of E as **check nodes**: this is explained further in Subsection 6.4 below. Let $\sigma \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V))$ with $|V| = n$. Fix $i \in [d]$, and consider that V is partitioned into the orbits of the generator $\sigma(s_i)$. Each orbit corresponds to a hyper-edge as in (9). Since each hyper-edge has size k, there are n/k of them, and so there exists a bijection between E_i and this family of hyper-edges. Let us choose such a bijection uniformly and independently at randomly for each i, and record the result as a subset $H \subseteq E \times V$: a pair $(e,v) \in E_i \times V$ lies in H if e is attached by the i^{th} bijection to the hyper-edge that contains v.

We regard H as a bipartite graph on the disjoint union of E and V. As such, each check node in E has exactly k neighbours in V, and each vertex in V has exactly one neighbour in each of the subsets E_i (and thus d neighbours in total). It follows that each intersection $H \cap (E_i \times V)$ is equivalent to a partition of V into parts labeled by E_i. In the sequel, we borrow some more terminology from coding theory and refer to any such bipartite graph H on E and V as a **factor graph**: see, for instance, [48, Sections 26.1 and 47.2]. Beware that this is actually a slight deviation from standard usage, which would not insist that H be a union of the partitions $H \cap (E_i \times V)$, but here we do take this as part of the definition of a 'factor graph'.

More generally, if $F \subseteq E$, then a **partial factor graph** on F and V is a bipartite graph $M \subseteq F \times V$ such that every check node in F has precisely k neighbours in V and every vertex in V is joined to at most one check node in each intersection $F \cap E_i$. Equivalently, there exists a factor graph H such that $M = H \cap (F \times V)$. In particular, if $F = E_i$, then a partial factor graph is simply a partition of V into k-sets that are labeled by E_i.

If H is a factor graph on E and V and $F \subseteq E$, then the **vertex neighbourhood** of F is the set

$$\mathtt{Vert}(H;F) := \{v \in V : \ (e,v) \in H \text{ for some } e \in F\}.$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

We also use the notation $\texttt{Vert}(M;F)$ for a partial factor graph M in the same way.

Similarly, the **check-node neighbourhood** of $U\subseteq V$ is the set

$$\texttt{Check}(H;U) := \{e\in E:\ (e,v)\in H \text{ for some } v\in V\}.$$

We may iterate these definitions to define neighbourhoods with larger radii. In general, for $F\subseteq E$ or $U\subseteq V$, we define

$$\begin{aligned} \texttt{Vert}_1(H;F) &:= \texttt{Vert}(H;F),\\ \texttt{Check}_1(H;U) &:= \texttt{Check}(H;U),\\ \texttt{Check}_1(H;F) &:= \texttt{Check}_1\big(\texttt{Vert}_1(H;F)\big)\\ \text{and}\quad \texttt{Vert}_1(H;U) &:= \texttt{Vert}_1\big(\texttt{Check}_1(H;U)\big). \end{aligned}$$

Then for integers $r>1$, we make the recursive definitions

$$\texttt{Vert}_r(H;F) := \texttt{Vert}_1(H;\big(\texttt{Vert}_{r-1}(H;F)\big),$$

and the same with U in place of F or $\texttt{Check}$ in place of $\texttt{Vert}$. In graph theoretic terms, if $F\subseteq E$, then $\texttt{Vert}_r(H;F)$ is the set of all vertices whose graph distance in H is at most $2r-1$ from F, $\texttt{Check}_r(H;F)$ is the set of all check nodes whose graph distance in H is at most $2r$ from F, and similarly for subsets of V.

These neighbourhoods are compatible with the Cayley graph neighbourhoods $B_r = \mathrm{B}(e,r)$ induced by the word metric on Γ (with respect to the generating set $\{s_i^j:\ 1\le i\le d, 1\le j\le k-1\}$). Specifically, if H arises from σ through the construction above, and $U\subseteq V$, then

$$\sigma^{B_r}(U) := \{\sigma^g(u):\ g\in B_r,\ u\in U\} = \texttt{Vert}_r(H;U).$$

Now consider $\sigma\sim\mathbb{P}$ and generate H from it as described above. Then H results from two random steps: the choice of σ and then the choice of a bijection for each i. Let $\tilde{\mathbb{P}}$ be the joint distribution of (σ,H) after this construction. This is a probability measure on

$$\tilde{\Omega} := \mathrm{Hom}_{\mathrm{unif}}(\Gamma,\mathrm{Sym}(V))\times\{0,1\}^{E\times V}.$$

It is a coupling of $\mathbb{P}$ to the law of H described above, and it has the following simple properties:

- Given σ, the conditional distribution of H is uniform over $((n/k)!)^d$ choices of bijections.
- Given H, the conditional distribution of σ is uniform over all choices of cyclic orderings for each hyper-edge of the multi-hyper-graph: there are $((k-1)!)^{dn/k}$ such choices for any H.
- Under the marginal distribution of H, the intersections $H\cap(E_i\times V)$ are independent as i varies.

For (σ,H) drawn from $\tilde{\mathbb{P}}$, the underlying multi-hyper-graph may be read off from either coordinate: it is the multi-set of all $\sigma(s_i)$ orbits for $i\in[d]$, and it is also the collection of all vertex neighbourhoods of the check nodes according to H. The labeling of these hyper-edges by the fixed set E that is given by H is convenient for union bounds and other forms

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

of counting. Most of our probabilistic estimates concerning these hyper-graphs later refer to $\tilde{\mathbb{P}}$ rather than $\mathbb{P}$.

Remark. The random hyper-graphs on V that arise from $\sigma \sim \mathbb{P}$ are k-uniform and d-regular.

Models of such random hyper-graphs already have an established place in the literature on probabilistic combinatorics: see, for instance, [76, Subsection 3.5] and the references given there. However, most of that literature is dedicated to a *uniform* random choice of such a hyper-graph, and this is not the same as the distribution that results from a random choice of σ. The point is that if the hyper-graph is generated by a homomorphism σ, then its hyper-edges can be classified according to which generator s_i gave rise to them. For each single i, the corresponding hyper-edges form a partition of V into k sets. In general, a k-uniform d-regular hyper-graph need not be a union of partitions.

In case $k = 2$, the difference here is between a uniformly random d-regular graph and the sum of d independent random matchings. In this case, the difference has been studied in some depth, with the outcome that these models are 'contiguous': see, for instance, [34] or[76, Subsection 4.3]. The relation of contiguity is strong enough to allow us to transfer most phenomena of interest from one model to the other.

However, for $k > d \geq 3$, it turns out that contiguity fails. While we have not found a reference for this fact in the literature, it follows fairly easily from some other standard results, so we explain this in Appendix A.

For uniformly random k-uniform d-regular hyper-graphs, estimates on the typical behaviour of the resulting LDPC codes are widely available in the coding theory literature. For example, the typical rate of the resulting LDPC code is known very exactly from [51] (we cite this in our proof of non-contiguity in Appendix A). These known results would include most of the facts we need here, were it not for the difference in the underlying random hyper-graph model. However, we have not found the analogous estimates for our distributions $\mathbb{P}$, so we must develop them here from scratch as necessary. Nevertheless, the conclusions generally look the same as in those previous works, and we have been guided by those throughout.

6.4. Parity-check matrices

To prove the desired properties of the random measures μ_{σ_n}, we make careful use of the linear structure of the codes X_n. As is standard in coding theory, this structure is conveniently summarized by a parity-check matrix.

To introduce this point of view, we start with some more notation. If a ground set A is understood and $B \subseteq A$, then $\mathbf{e}_B$ denotes the mod-2 indicator function of B: this is the element of $\mathbb{Z}_2^A$ with entries equal to $\mathtt{1}$ precisely at the indices in B.

Now consider a vertex set V of size n which is a multiple of k, and let $E = E_1 \cup \cdots \cup E_d$ as in Subsection 3.1. Let H be a factor graph on E and V, and turn it into the $(E \times V)$-matrix $\mathbf{H} = \mathbf{e}_H$, which also takes values in $\mathbb{Z}_2$. If this generates the same hyper-graph as $\sigma \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V))$, then our code is given by

$$X_\sigma = \{\mathbf{x} \in \mathbb{Z}_2^V : \ \mathbf{Hx} = \mathbf{0}\} = \ker \mathbf{H}.$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

In this role, $\mathbf{H}$ is called the **parity-check matrix** (for an introduction, see, for instance, [20, Chapter 9], [23, Section 7.11] or [48, Chapter 1], especially the discussion of Exercise 1.9). This is why we refer to the elements of E as 'check nodes': each corresponds to a row of $\mathbf{H}$, which every codeword $\mathbf{x}$ must be orthogonal to. The representation of a code using a set of connections between vertices and check nodes is a common example of a 'factor graph representation' [48, Section 47.2] – hence our use of the term 'factor graph' for H.

Because X_σ is a linear subspace of $\mathbb{Z}_2^V$, it may equivalently be specified via its dual code

$$X_\sigma^\perp := \{\mathbf{y} \in \mathbb{Z}_2^V : \langle \mathbf{y}, \mathbf{x} \rangle = 0 \pmod 2 \ \forall \mathbf{x} \in X_\sigma\}.$$

Below, we refer to elements of $X_\sigma^\perp$ as **parity checks**. By the construction of X_σ, this $X_\sigma^\perp$ is precisely the linear subspace of $\mathbb{Z}_2^V$ spanned by the rows of $\mathbf{H}$ – that is, by the vectors $\mathbf{e}_{\mathtt{Vert}(H;\{e\})}$ for $e \in E$. More succinctly,

$$X_\sigma^\perp = \operatorname{img}(\mathbf{H}^\mathsf{T}).$$

Theorem 6.1 (1,2) are proved by counting relations or 'near relations' among the rows of $\mathbf{H}$. The second of these theorems requires the more complicated calculation. It is based on Proposition 7.1 below. With that proposition in hand, we can also prove Theorem 6.1(2); see Section 8.

7. Property M

In this section, we again suppress the subscript n from our notation, as in Subsection 6.3. As before, this index will tend to ∞ along multiples of k.

The main part of the proof of Theorem 6.1 parts (1) and (2) is another, more technical proposition. Put roughly, it rules out most 'near cancellations' among the parity-check words of a typical (σ, H) drawn at random. However, for the application to Theorem 6.1, we need such a result not only when $(\sigma, H) \sim \tilde{\mathbb{P}}$ but also after conditioning (σ, H) on a small fraction of the vertices and check nodes.

We formulate this technical proposition next. For each $i \in [d]$, let $F_i \subseteq E_i$ and let $W_i := \mathtt{Vert}(H; F_i)$. Let $w_i := |W_i| = k|F_i|$. Let $F := F_1 \cup \cdots \cup F_d$, $W := W_1 \cup \cdots \cup W_d$, and $w := |W|$; see Figure 1. Then

$$\max\{w_1, \ldots, w_d\} \leq w \leq w_1 + \cdots + w_d.$$

Let M be a partial factor graph on F and V, and let $\tilde{\mathbb{P}}^M$ be the distribution obtained by conditioning $\tilde{\mathbb{P}}$ on the event that $H \cap (F \times V) = M$. The key to Theorem 6.1 is that if the sets F_i are small enough, then after this conditioning, the rest of H is very unlikely to create many new parity checks that involve only vertices inside W. Roughly speaking, this means that if $\mathbf{x} \in \mathbb{Z}_2^W$ satisfies all of the parity checks arising from F, then, with high probability, it admits an extension satisfying all of the parity checks.

Here and in the rest of the paper, if $F \subseteq E$ and $U \subseteq V$, then $\mathbf{H}_{F \times U}$ denotes the submatrix of $\mathbf{H}$ indexed by these sets, and we use analogous notation for the transpose $\mathbf{H}^\mathsf{T}$ and for vectors indexed by either E or V. We identify $\mathbb{Z}_2^{E \setminus F}$ as a subspace of $\mathbb{Z}_2^E$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

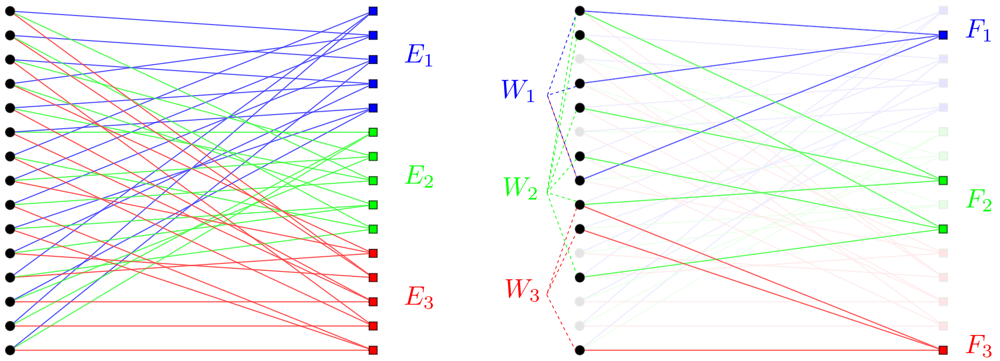

Figure 1. Diagrams showing full factor graph H (left) and partial factor graph M (right). The square vertices on the right of each graph are the check nodes, colored according to their membership in the sets E_1, E_2, E_3. The distribution $\tilde{\mathbb{P}}^M$ draws a new pair (σ, H) conditioned on the edges on the right being present.

in the obvious way, and so any $\mathbf{y} \in \mathbb{Z}_2^{E\setminus F}$ may be written as a tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_d)^\mathsf{T}$ with $\mathbf{y}_i \in \mathbb{Z}_2^{E_i\setminus F_i}$.

Proposition 7.1 ('Few additional checks inside W'). *For every $K > 0$ and $\varepsilon > 0$, the following holds for all sufficiently small δ (depending on d, k, ε and K). Let F_i, W_i, w_i and M be as above. If*

$$\delta n \le w_1 + \cdots + w_d \le K\delta n, \tag{25}$$

then

$$\tilde{\mathbb{P}}^M\Big(\exists \mathbf{y} \in \mathbb{Z}_2^{E\setminus F} \text{ such that } \min\{|\mathbf{y}_i|, |\mathbf{e}_{E_i\setminus F_i} - \mathbf{y}_i|\} \ge \varepsilon\delta\frac{n}{k} \text{ for some } i \\ \text{and } (\mathbf{H}^\mathsf{T}\mathbf{y})_{V\setminus W} = \mathbf{0}\Big) \to 0$$

as $n \to \infty$. The rate of convergence depends only on d, k and δ and is independent of δ provided δ is small enough and also bounded away from zero.

We show how Proposition 7.1 implies Theorem 6.1 in the next subsection, and then we prove Proposition 7.1 itself in Subsection 7.2.

Remark. Proposition 7.1 has a similar flavor to the analysis of the satisfiability threshold for the random combinatorial problem known as random XORSAT. But our setting has the additional complication that we must analyse conditional probabilities given the behaviour of the random factor graph H on the small sets F_i, which seems to make our situation less 'homogeneous'. The random XORSAT model is treated in [49, Chapter 18] using the paradigm of 'belief propagation'. It is possible that that approach could also be brought to bear in the situation in Proposition 7.1, possibly leading to an alternative proof, but we have not pursued this idea further.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

7.1. Property M from few additional checks

To deduce Theorem 6.1 from Proposition 7.1, we also need two simpler calculations that we present as separate lemmas.

The first concerns the following construction. Fix a subset R of V and let $r>0$. For any factor graph $H \subset E \times V$, we can form the neighbourhood $F := \mathtt{Check}_r(H;R)$, and then the intersection $M := H \cap (F \times V)$ is a partial factor graph on F and V. Let us call the pair (F,M) **possible with respect to** (R,r) if they can arise from a factor graph in this way.

Lemma 7.2. *Let R and r be as above, let (F,M) be possible with respect to (R,r), and let $(\sigma,H) \sim \tilde{\mathbb{P}}$. If the event*

$$\{H: \ H \cap (F \times V) = M\}$$

occurs, then so does the event

$$\{H: \ \mathtt{Check}_r(H;R) = F\}.$$

If we condition on the former event, then the subsets $H \cap ((E_i \setminus F) \times V)$ are independent for different i, and each is a uniform random labeled partition of $V \setminus \mathtt{Vert}(M;E_i \cap F)$ into k-sets.

Proof. First, the definition of the neighbourhood $\mathtt{Check}_r(H;R)$ depends only on those edges of the bipartite graph H that connect this neighbourhood to V. Since we are told that (F,M) is possible with respect to (R,r), all those edges must already be visible in M, and so knowing that $H \cap (F \times V) = M$ is enough to tell us that $\mathtt{Check}_r(H;R) = F$.

Now the conditional probability in question is $\tilde{\mathbb{P}}^M$, as introduced previously. Since $\tilde{\mathbb{P}}$ is a uniform distribution, $\tilde{\mathbb{P}}^M$ is the uniform distribution over all factor graphs for which the event holds. However, if the event holds, then (i) $H \cap (F \times V)$ is uniquely determined, (ii) each $H \cap ((E_i \setminus F) \times V)$ must consist of a labeled partition of $V \setminus \mathtt{Vert}(M;E_i \cap F)$ into k-sets, and (iii) any tuple of such labeled partitions is still possible. So this conditional distribution is simply the uniform distribution over the Cartesian product set of tuples of such labeled partitions. This implies the desired joint distribution for these sets. □

The next lemma may be well known in coding theory, but we have not found a convenient reference.

Lemma 7.3. *Let A be any finite nonempty index set, $\delta < 1/3$, and let Y be a linear subspace of $\mathbb{Z}_2^A$ such that*

$$\text{either} \quad |\mathbf{y}| \le \delta|A| \quad \text{or} \quad |\mathbf{y}| \ge (1-\delta)|A| \quad \text{for every} \quad \mathbf{y} \in Y. \tag{26}$$

Then $\dim Y \le 2\delta|A| + 1$.

Proof. Let Z be the subset of all $\mathbf{y} \in Y$ for which $|\mathbf{y}| \le \delta|A|$. We prove that Z is a linear subspace, $\dim Y/Z \le 1$, and $\dim Z \le 2\delta|A|$.

First, Z clearly contains $\mathbf{0}$, and if $\mathbf{y},\mathbf{y}' \in Z$, then $\mathbf{y}+\mathbf{y}' \in Y$ and

$$|\mathbf{y}+\mathbf{y}'| \le 2\delta|A|.$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Since $2\delta < 1-\delta$, by (26), this forces $\mathbf{y}+\mathbf{y}' \in Z$, so Z is a subspace.

Next, if $\mathbf{y},\mathbf{y}' \in Y\setminus Z$, then

$$|\mathbf{y}-\mathbf{y}'| \le |(\mathtt{1},\dots,\mathtt{1})-\mathbf{y}|+|(\mathtt{1},\dots,\mathtt{1})-\mathbf{y}'| \le 2\delta|A|,$$

so again we must in fact have $\mathbf{y}-\mathbf{y}' \in Z$. Therefore, $\dim Y/Z \le 1$.

Finally, call an index $i \in A$ **proper** if some member of Z is nonzero in this coordinate. Now choose $\mathbf{z}$ from Z uniformly at random. If i is proper, then the coordinate z_i is equally likely to be 0 or 1, so the expectation of $|\mathbf{z}|$ is half the number of proper coordinates. Therefore, the number of proper coordinates is at most $2\delta|A|$, and this number is an upper bound on $\dim Z$. □

Property M (Theorem 6.1) follows from the following result, which is also used to prove local and empirical convergence:

Theorem 7.4. *Given $r \in \mathbb{N}$ and $\eta > 0$, for all small enough $\delta > 0$, the following holds: if for each n, $R_n \subset V_n$ is a subset of size $\lceil \delta n\rceil$, then with high probability as $n \to \infty$,*

$$\mathrm{H}\big((\mu_n)_{\sigma_n^{B_r}(R_n)}\big) \ge (1-\eta)\,\mathrm{H}\big(m_{B_r}\big)|R_n|.$$

Proof of Theorem 7.4 assuming Proposition 7.1. Beware that we continue to suppress n from subscripts where possible. It should be understood that the following argument and construction are carried out for each n that is divisible by k.

Fix $\varepsilon > 0$. For a small positive δ to be specified shortly, let R be any fixed choice of a subset of V of size $\lceil \delta n\rceil$, and now consider the following three subsets of (σ,H) in $\tilde{\Omega}$:

i. (Few additional checks) (σ,H) is in Ω^1 if for

any $\mathbf{y} \in \mathbb{Z}_2^{E\setminus \mathtt{Check}_r(H;R)}$ such that $(\mathbf{H}^{\mathsf{T}}\mathbf{y})_{V\setminus \mathtt{Vert}_r(H;R)} = \mathbf{0}$ also has

$$\min\{|\mathbf{y}_i|,|\mathbf{e}_{E_i\setminus F_i}-\mathbf{y}_i|\} < \varepsilon\delta\frac{n}{k} \qquad \text{for every } i \in [d],$$

where $F_i = E_i \cap \mathtt{Check}_r(H;R)$.

ii. (Most vertices in R well-separated) (σ,H) is in Ω^2 if the set

$$S_1 := \big\{v \in R:\ \mathtt{Vert}_r(H;v)\cap \mathtt{Vert}_r(H;R\setminus v) = \emptyset\big\}$$

has $|S_1| > (1-\varepsilon)|R|$.

iii. (Most vertices in R not close to any short loops) (σ,H) is in Ω^3 if the set

$$S_2 := \big\{v \in R:\ \text{the orbit map } B_r \to \mathtt{Vert}_r(H;v) \text{ is injective}\big\}$$

has $|S_2| > (1-\varepsilon)|R|$. (Here, the orbit map sends $\gamma \in B_r \subset \Gamma$ to $\sigma(\gamma)v$.)

Let $S := S_1 \cap S_2$ and $\Omega' := \Omega^1\cap\Omega^2\cap\Omega^3$. Think of S as the result of expurgating the ‘bad’ vertices from R and Ω' as the event that there are only a few of these.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

In the remainder of the proof, we show that, provided δ was chosen small enough, we have $\tilde{\mathbb{P}}(\Omega') \to 1$ and the set R satisfies the desired entropy bound.

Step 1: $\tilde{\mathbb{P}}(\Omega') \to 1$. First, we have $\tilde{\mathbb{P}}(\Omega^2) \to 1$ as $n \to \infty$ for any sufficiently small δ in terms of d, k, r and ε. To see this, observe that although we are fixing R and choosing (σ, H) at random, we obtain the same distribution on $|S_1|$ if we choose R uniformly at random among all subsets of V of cardinality $\lceil \delta n \rceil$, independently of (σ, H). Thus, it suffices to show that on any d-regular k-uniform hyper-graph on n vertices, if a subset $R \subset V$ of $\lceil \delta n \rceil$ vertices is chosen uniformly at random, then the probability that there are more than $\varepsilon\delta n$ vertices of R which are $\leq 2r$ distance from another vertex in R tends to zero as $n \to \infty$.

If some number, say x, of vertices have already been chosen, then the probability that the next vertex is not within distance $2r$ of the previously selected vertices is at least

$$1 - \frac{(kd)^{2r} x}{n - x} \geq 1 - \frac{(kd)^{2r} \lceil \delta n \rceil}{n - \lceil \delta n \rceil} \geq 1 - (kd)^{2r} \frac{\delta}{1 - \delta}.$$

This is because the number of vertices in the $(2r)$-neighbourhood of a given vertex is at most $(kd)^{2r}$. Thus, $\tilde{\mathbb{P}}(\Omega^2)$ is at least the probability that in $\lceil \delta n \rceil$ Bernoulli trials with success probability $1 - (kd)^{2r} \delta/(1-\delta)$, there are at least $(1-\varepsilon)\lceil \delta n \rceil$ successes. This occurs with overwhelming probability as $n \to \infty$ as long as $\varepsilon > (kd)^{2r} \delta/(1-\delta)$, which we may assume by choosing δ sufficiently small.

Second, for any $\delta > 0$, we have $\tilde{\mathbb{P}}(\Omega^3) \to 1$ as $n \to \infty$ as an immediate corollary of Proposition 1.2.

So now let us show that $\tilde{\mathbb{P}}(\Omega^1) \to 1$ as $n \to \infty$. This is our application of Proposition 7.1. We make contact with that proposition by conditioning on the partial factor graph $H \cap (\mathtt{Check}_r(H; R) \times V)$ and using the law of total probability.

Let (F, M) be a possible pair with respect to (R, r) as in Lemma 7.2, and let $\tilde{\mathbb{P}}^M$ be the result of conditioning $\tilde{\mathbb{P}}$ on the event $H \cap (F \times V) = M$, as previously. Let $W_i := \mathtt{Vert}(M; F_i)$ and let $W := W_1 \cup \cdots \cup W_d$. By the law of total probability, $\tilde{\mathbb{P}}(\tilde{\Omega} \setminus \Omega^1)$ is equal to the sum

$$\sum_{\text{possible } F, M} \tilde{\mathbb{P}}\big(H \cap (F \times V) = M\big) \cdot \tilde{\mathbb{P}}^M \Big(\exists \mathbf{y} \in \mathbb{Z}_2^{E \setminus F} \text{ such that}$$
$$\min\{|\mathbf{y}_i|, |\mathbf{e}_{E_i \setminus F_i} - \mathbf{y}_i|\} \geq \varepsilon\delta \frac{n}{k} \text{ for some } i \text{ and } (\mathbf{H}^{\mathsf{T}} \mathbf{y})_{V \setminus W} = \mathbf{0}\Big).$$

Let $K := (dk)^{r+1} + 1$. Our construction of the possible pair (F, M) gives that

$$\delta n \leq |R| \leq |W| \leq \sum_{i=1}^{d} |W_i| \leq (dk)^{r+1} |R| \leq K\delta n$$

when $\delta n > 1$, and so the condition (25) is satisfied for this value of K by the second factor in every term of the sum above. Therefore, by Proposition 7.1, if δ is small enough in terms of d, k, ε and r, then this sum is a convex combination of quantities that converge to 0 at a rate depending only on d, k and δ, and hence, so does the whole expression.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

This proves that $\tilde{\mathbb{P}}(\Omega') \to 1$ for any sufficiently small δ. Fix such a δ for the rest of the proof.

Step 2. To finish the proof, we show that if $\sigma \in \Omega'$, then (i), (ii) and (iii) imply that S has the properties required to witness property M as in Definition 4.2.

First, S_1 is $(2r)$-separated by property (ii), and hence, so is S.

Second, properties (ii) and (iii) together give

$$\frac{|S|}{|V|} \geq (1-2\varepsilon)\frac{|R|}{|V|} \geq (1-2\varepsilon)\delta,$$

which is uniformly positive in n provided $\varepsilon < 1/2$.

Finally, we must prove (11). Because μ_σ is the uniform distribution on the linear subspace X_σ of $\mathbb{Z}_2^V$, the projection of μ_σ over the subset of vertices $\sigma^{B_r}(S) = \mathtt{Vert}_r(H;S)$ is the uniform distribution on the linear space

$$Z := \{\mathbf{x}_{\mathtt{Vert}_r(H;S)} : \ \mathbf{x} \in X_\sigma\}.$$

Therefore, after ignoring a factor of $\log 2$, we need a lower bound on $\dim Z$.

By property (iii), each of the neighbourhoods $\mathtt{Vert}_r(H;v)$ for $v \in S$ is a bijective copy of $B_r \subseteq \Gamma$, and by property (ii), these neighbourhoods are disjoint. Therefore, Z is naturally identified with a linear subspace of

$$\prod_{v \in S} \mathbb{Z}_2^{\mathtt{Vert}_r(H;v)} \cong (\mathbb{Z}_2^{B_r})^S.$$

As a linear subspace of $\mathbb{Z}_2^{\mathtt{Vert}_r(H;S)}$, Z is determined by its dual code $Z^\perp$ – that is, its own collection of parity-check words in $\mathbb{Z}_2^{\mathtt{Vert}_r(H;S)}$ [48, Section 13.10]. Since Z is the projection of X_σ, each word in $Z^\perp$ becomes a word in $X_\sigma^\perp$ when it is extended by 0 to $V \setminus \mathtt{Vert}_r(H;S)$. Therefore, $Z^\perp$ is identified with the set of all mod-2 sums of the rows of $\mathbf{H}$ that vanish on $V \setminus \mathtt{Vert}_r(H;S)$, and so the rank-nullity formula gives

$$\dim Z = \dim(\mathbb{Z}_2^{B_r})^S - \dim\ker \mathbf{H}^{\mathsf{T}}_{(V\setminus\mathtt{Vert}_r(H;S))\times E}. \tag{27}$$

So let us consider the ways in which a mod-2 sum of the rows of $\mathbf{H}$ can vanish on $V \setminus \mathtt{Vert}_r(H;S)$. First, each individual row of $\mathbf{H}$ that is indexed by an element of $\mathtt{Check}_r(H;S)$ vanishes outside $\mathtt{Vert}_r(H;v)$ for some single element $v \in S$. Since each of these sets is a bijective copy of B_r, these checks by themselves show that Z is actually a linear subspace of $X_{B_r}^S \subset (\mathbb{Z}_2^{B_r})^S$.

However, given any other mod-2 sum of rows of $\mathbf{H}$ which vanishes on $V \setminus \mathtt{Vert}_r(H;S)$, we may remove any summands indexed by $\mathtt{Check}_r(H;S)$ without losing that property. So now consider a mod-2 sum of rows of $\mathbf{H}$ that is supported in $\mathtt{Vert}_r(H;S)$ and uses no rows indexed by $\mathtt{Check}_r(H;S)$. It might use only rows indexed by $\mathtt{Check}_r(H;R) \setminus \mathtt{Check}_r(H;S)$. Or it might include at least one summand indexed by an element of $E \setminus \mathtt{Check}_r(H;R)$, in which case those summands by themselves define a nonzero vector in

$$D := \ker \mathbf{H}^{\mathsf{T}}_{(V\setminus\mathtt{Vert}_r(H;R))\times(E\setminus\mathtt{Check}_r(H;R))}$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

because the rows of $\mathbf{H}$ indexed by $\mathtt{Check}_r(H;R)$ are supported in $\mathtt{Vert}_r(H;R)$, and so ignoring them cannot change the support in $V \setminus \mathtt{Vert}_r(H;R)$.

Combining these possibilities, (27) implies the lower bound

$$\dim Z \geq |S| \cdot \dim X_{B_r} - |\mathtt{Check}_r(H;R) \setminus \mathtt{Check}_r(H;S)| - \dim D. \tag{28}$$

By properties (ii) and (iii), $|S| \geq (1-2\epsilon)|R|$. Since $|\mathtt{Check}_r(H;R \setminus S)| \leq (dk)^{r-1}k|R \setminus S|$, this implies

$$|\mathtt{Check}_r(H;R) \setminus \mathtt{Check}_r(H;S)| \leq |\mathtt{Check}_r(H;R \setminus S)| \leq Kk\varepsilon|S|.$$

However, by property (i), if $\mathbf{y} \in D$ and we write $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_d)^{\mathsf{T}}$, then for each i, we have $\mathbf{y}_i \in \mathbb{Z}_2^{E_i \setminus \mathtt{Check}_r(H;R)}$ and

$$\text{either } |\mathbf{y}_i| < \varepsilon\delta\frac{n}{k} \quad \text{or} \quad |\mathbf{e}_{E_i \setminus \mathtt{Check}_r(H;R)} - \mathbf{y}_i| < \varepsilon\delta\frac{n}{k}.$$

Therefore, provided $\varepsilon\delta < 1/3$, Lemma 7.3 gives

$$\dim\{\mathbf{y}_i : \ \mathbf{y} \in D\} \leq 2\varepsilon\delta\frac{n}{k} + 1 \quad \text{for each } i,$$

and hence, $\dim D \leq 2\varepsilon\delta nd/k + d$.

Inserting these bounds into (28), we finally arrive at

$$\begin{aligned}\dim Z &\geq |S| \cdot \dim X_{B_r} - Kk\varepsilon|S| - 2\varepsilon\delta\frac{nd}{k} - d \\ &\geq \left(\dim X_{B_r} - Kk\varepsilon - \frac{2\varepsilon d}{k} - o(1)\right) \cdot |S|.\end{aligned}$$

This gives us a lower bound on the desired joint entropy:

$$\begin{aligned}\mathrm{H}\big((\mu_n)_{\sigma_n^{B_r}(R)}\big) &\geq \mathrm{H}\big((\mu_n)_{\sigma_n^{B_r}(S)}\big) \\ &= \log 2 \cdot \dim Z \\ &\geq \left(\log 2 \cdot \dim X_{B_r} - \varepsilon \log 2 \cdot \left(Kk - \frac{2d}{k} - o(1)\right)\right) \cdot (1-2\varepsilon)|R| \\ &\geq (\mathrm{H}(m_{B_r}) - \varepsilon C)\,|R|,\end{aligned}$$

where $C = \log 2 \cdot (Kk - \frac{2d}{k}) + 2\,\mathrm{H}(m_{B_r})$. Since C depends only on d,k and r, we could have taken at the beginning $\varepsilon = \eta/C$, so this proves (11). □

Remark. Proposition 7.1 shows that, with high conditional probability in the choice of $\mathbf{H}$, the only parity checks among the bits in W that are created by $\mathbf{H}$ are (i) those created by the rows in F, and possibly (ii) a few others that are generated by some vectors $\mathbf{y}$ whose support is either extremely small or extremely close to the whole of $E \setminus F$. Since the number of possible vectors of type (ii) is very small compared with those of type (i), this implies that there are few enough ‘spurious’ parity checks among the bits in W to give Theorem 6.1. However, we do not expect that there are *no* extra parity checks of type (ii): just by chance, one should typically find as many as a very small multiple of n of these.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Proof of Theorem 6.1 item (1) from Theorem 7.4. We have to prove there are subsets

$$\Omega'_n \subseteq \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V_n))$$

such that $\mathbb{P}_n(\Omega'_n) \to 1$ and if $\sigma_n \in \Omega'_n$ for each n, then $(\mu_n)_n$ has property M with respect to $\Sigma = (\sigma_n)_n$. This means that for every $\epsilon > 0$ and $0 < r < \infty$, there is a sequence of subsets $S_n \subset V_n$ such that

$$\liminf \frac{|S_n|}{|V_n|} > 0$$

and

$$\mathrm{H}((\mu_n)_{\sigma_n^{B_r}(S_n)}) \geq |S_n|(\mathrm{H}(m_{B_r}) - \epsilon) \tag{29}$$

for all n.

Let $r,\varepsilon > 0$. By Theorem 7.4, if $\delta > 0$ is sufficiently small and for each n, $S_n \subset [n]$ is an arbitrary set of size $\lceil \delta n \rceil$, then there is some sequence $(\Omega'_n)_n$ with $\mathbb{P}(\Omega'_n) \to 1$ such that the desired entropy inequality holds when $\sigma_n \in \Omega'_n$, and $\liminf \frac{|S_n|}{n} = \liminf \frac{\lceil \delta n \rceil}{n} = \delta > 0$.

Then, by diagonalizing over the countably many choices of $r \in \mathbb{N}$ and $\varepsilon \in \{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$, we can get a single sequence of sets $(\Omega'_n)_n$ with $\mathbb{P}(\Omega'_n) \to 1$ such that for any $r \in \mathbb{N}$ and $\varepsilon > 0$, there is a sequence $(S_n)_n$ with the desired properties. □

7.2. Proof that there are few additional checks

This subsection proves Proposition 7.1. Let the sets F_i, W_i and M and parameters K, ε, w_i and w be as in that statement. For $\mathbf{y} \in \mathbb{Z}_2^{E \setminus F}$, let

$$G_{\mathbf{y}} := \left\{(\sigma, H) : (\mathbf{H}^{\mathsf{T}}\mathbf{y})_{V \setminus W} = \mathbf{0}\right\} \subset \tilde{\Omega}_n.$$

Write $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_d)^{\mathsf{T}}$ with $\mathbf{y}_i \in \mathbb{Z}_2^{E_i \setminus F_i}$. We will focus on those vectors $\mathbf{y}$ which satisfy restrictions on the cardinalities $|\mathbf{y}_i|$ coming from Proposition 7.1. To be precise, let $\mathcal{R} = \mathcal{R}(\varepsilon, \delta)$ be the set of all nonnegative integer tuples $\ell = (\ell_1, \ldots, \ell_d)$ that satisfy

$$\ell_i \leq \frac{n - w_i}{k} \quad \text{for every } i \quad \text{and} \quad \min\left\{\ell_i, \frac{n - w_i}{k} - \ell_i\right\} \geq \varepsilon\delta \frac{n}{k} \quad \text{for some } i. \tag{30}$$

The main conclusion of Proposition 7.1 is equivalent to

$$\tilde{\mathbb{P}}^M\left(\bigcup_{\mathbf{y}:\ (|\mathbf{y}_1|, \ldots, |\mathbf{y}_d|) \in \mathcal{R}} G_{\mathbf{y}}\right) \to 0$$

as $n \to \infty$. Our proof uses a simple union bound over $\mathbf{y}$. We will derive estimates on $\tilde{\mathbb{P}}^M(G_{\mathbf{y}})$ that depend on ε and δ for $\mathbf{y} \in \mathcal{R}(\varepsilon, \delta)$ and then use these to conclude Proposition 7.1 provided δ is chosen correctly.

Fix a vector $\mathbf{y} \in \mathbb{Z}_2^{E \setminus F}$ with $(|\mathbf{y}_1|, \ldots, |\mathbf{y}_d|) \in \mathcal{R}$, and set $r_i := k|\mathbf{y}_i|$ and $\mathbf{r} := (r_1, \ldots, r_d)$. Let Y_i be the support of the random vector $\mathbf{H}^{\mathsf{T}}\mathbf{y}_i$ for each i. This means $\mathbf{H}^{\mathsf{T}}\mathbf{y}_i = \mathbf{e}_{Y_i}$. These random sets are independent by the independence in Lemma 7.2. Each Y_i is uniformly random among all subsets of $V \setminus W_i$ of size r_i.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Each Y_i may have nonempty intersection with $W \setminus W_i$. We bound $\tilde{\mathbb{P}}^M(G_{\mathbf{y}})$ by breaking into a few further cases depending on the sizes of these intersections. To this end, let $Z_i := Y_i \setminus W$. Under $\tilde{\mathbb{P}}^M$, the cardinality $|Z_i|$ is a random quantity obtained by sampling r_i points from $V \setminus W_i$ without replacement and counting how many of them land in $V \setminus W$. This random quantity has the hypergeometric distribution with parameters $n - w_i$, $n - w$ and r_i: see, for instance, [28, Section II.6]. Its possible values are the integers s_i that satisfy

$$\max(0, r_i - (w - w_i)) \leq s_i \leq \min(r_i, n - w), \tag{31}$$

and for such values, we have

$$\tilde{\mathbb{P}}^M(|Z_i| = s_i) = \frac{\binom{n-w}{s_i}\binom{w-w_i}{r_i - s_i}}{\binom{n-w_i}{r_i}}.$$

By the standard exponential estimate for binomial coefficients (see, for instance, [23, Example 11.1.3]), this ratio is at most

$$\exp\left(\mathrm{H}\left(\frac{s_i}{n-w}\right)(n-w) + \mathrm{H}\left(\frac{r_i - s_i}{w - w_i}\right)(w - w_i) - \mathrm{H}\left(\frac{r_i}{n - w_i}\right)(n - w_i) + o(n)\right), \tag{32}$$

where quality of the error term does not depend on any other parameters.

Towards Proposition 7.1, we estimate the probability of $G_{\mathbf{y}}$ after further conditioning on the tuple of cardinalities $|Z_i|$, and then combine this estimate with (32) using the law of total probability. That refined conditional probability estimate depends on the following lemma.

Let $\mathtt{Even}(d)$ be the subset of all strings in $\{0,1\}^d$ that have even weight.

Lemma 7.5. *Let q_1, ..., q_d be probability distributions on $\{0,1\}$, and assume that q is a coupling of q_1, ..., q_d that is supported on* $\mathtt{Even}(d)$. *Then*

$$\mathrm{H}(q) \leq \left(1 - \frac{1}{d}\right)\big(\mathrm{H}(q_1) + \cdots + \mathrm{H}(q_d)\big).$$

Proof. By permuting indices, we may assume without loss of generality that

$$\mathrm{H}(q_1) \geq \cdots \geq \mathrm{H}(q_d).$$

Let $\mathbf{z} = (z_1, \ldots, z_d)$ be the identity map on $\{0,1\}^d$, and regard it as a random binary string with distribution q. Then $\mathbf{z}$ has even weight almost surely, and hence, the coordinates z_2, ..., z_d determine z_1 almost surely. Therefore,

$$\mathrm{H}(q) = \mathrm{H}_q(\mathbf{z}) = \mathrm{H}_q(z_2, \ldots, z_d) \leq \mathrm{H}_q(z_2) + \cdots + \mathrm{H}_q(z_d) = \mathrm{H}(q_2) + \cdots + \mathrm{H}(q_d).$$

Because of our ordering of the indices, this is at most the desired upper bound. □

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Now let $\mathcal{S}(\mathbf{r})$ be the set of all integer tuples $\mathbf{s} = (s_1,\dots,s_d)$ that satisfy (31) for every i.

Lemma 7.6. *Fix a vector* $\mathbf{y} \in \mathbb{Z}_2^{E\setminus F}$ *with* $(|\mathbf{y}_1|,\dots,|\mathbf{y}_d|) \in \mathcal{R}$ *as above. If* $\mathbf{s} \in \mathcal{S}(\mathbf{r})$*, then*

$$\tilde{\mathbb{P}}^M\left(G_{\mathbf{y}} \;\middle|\; |Z_i| = s_i \textit{ for each } i\right) \le \exp\left(-\frac{1}{d}\sum_{i=1}^{d} \mathrm{H}\left(\frac{s_i}{n-w}\right)\cdot(n-w) + o_d(n)\right).$$

Proof. To lighten notation, within this proof, let

$$\tilde{\mathbb{P}}^{M,\mathbf{s}} := \tilde{\mathbb{P}}^M\big(\,\cdot \;\big|\; |Z_i| = s_i \text{ for each } i\big).$$

Record the random sets Z_1, …, Z_d into the random vector

$$\boldsymbol{\eta} = (\eta_v)_{v\in V\setminus W} \in (\{0,1\}^d)^{V\setminus W} \quad \text{where } \eta_v := (1_{Z_1}(v),\dots,1_{Z_d}(v)).$$

Let $P_{\boldsymbol{\eta}} = \frac{1}{|V\setminus W|}\sum_{v\in V\setminus W}\delta_{\eta_v}$ be the empirical distribution of $\boldsymbol{\eta}$. This is a probability distribution on $\{0,1\}^d$, and the event $G_{\mathbf{y}}$ occurs if and only if this probability distribution is supported on the subset $\mathtt{Even}(d)$. The marginals of $P_{\boldsymbol{\eta}}$ are $(q_i, 1-q_i)$ for $i = 1,2,\dots,d$, where $q_i := s_i/(n-w)$. Moreover, $P_{\boldsymbol{\eta}}$ must take values that are multiples of $1/(n-w)$, and the total number of such possible distributions is at most

$$(n-w+1)^{2^d} \le n^{2^d} = e^{o_d(n)}$$

(here and in some subsequent steps, we generally loosen $o_d(n-w)$ to $o_d(n)$). Therefore,

$$\tilde{\mathbb{P}}^{M,\mathbf{s}}(G_{\mathbf{y}}) \le e^{o_d(n)} \max\big\{\tilde{\mathbb{P}}^{M,\mathbf{s}}(P_{\boldsymbol{\eta}} = q) : \ q \text{ a coupling of } q_1,\dots,q_d \text{ such that } q(\mathtt{Even}(d)) = 1\big\}. \tag{33}$$

For a distribution q as above, the set of vectors $\boldsymbol{\eta}$ that give $P_{\boldsymbol{\eta}} = q$ are the 'type class' of q, and their number is simply bounded using the entropy of q:

$$|\{\boldsymbol{\eta} \in (\{0,1\}^d)^{V\setminus W} : \ P_{\boldsymbol{\eta}} = q\}| \le e^{\mathrm{H}(q)\cdot(n-w)}$$

(see, for instance, [23, Theorem 11.1.3], except note that Cover and Thomas use $\log_2$ rather than natural logarithms to define H). By Lemma 7.5, this upper bound is always at most

$$\exp\left(\left(1-\frac{1}{d}\right)\cdot\big(\mathrm{H}(q_1)+\cdots+\mathrm{H}(q_d)\big)\cdot(n-w)\right). \tag{34}$$

However, under the conditional probability measure $\tilde{\mathbb{P}}^{M,\mathbf{s}}$, the set Z_i is a uniform random subset of $V\setminus W$ of size s_i, and these random sets are still independent. Therefore, the probability of any particular d-tuple of sets of these sizes occurring is

$$\prod_{i=1}^{d}\binom{n-w}{s_i}^{-1},$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

and by another use of standard exponential estimates on binomial coefficients [23, Example 11.1.3], this is at most

$$\exp\left(-\left(\mathrm{H}(q_1)+\cdots+\mathrm{H}(q_d)\right)\cdot(n-w)+o_d(n)\right).$$

Multiplying by the cardinality upper bound (34), we obtain

$$\tilde{\mathbb{P}}^{M,\mathbf{s}}(P_{\boldsymbol{\eta}}=q)\le\exp\left(-\frac{1}{d}\left(\mathrm{H}(q_1)+\cdots+\mathrm{H}(q_d)\right)\cdot(n-w)+o_d(n)\right)$$

for any such coupling q. Since this upper bound is independent of the particular coupling q, and the extra factor in (33) is sub-exponential, this gives the result. □

Lemma 7.7. *Fix k and d as before, and define the function*

$$f(t,\alpha',\alpha''):=(1-k^{-1})\,\mathrm{H}\left((1-t)\alpha'+t\alpha''\right)-(1-t)(1-d^{-1})\,\mathrm{H}(\alpha')-t\,\mathrm{H}(\alpha'')$$

for $0\le\alpha',\alpha'',t\le 1$. Then

$$\tilde{\mathbb{P}}^{M}\left(\bigcup_{\mathbf{y}:\ (|\mathbf{y}_1|,\ldots,|\mathbf{y}_d|)\in\mathcal{R}}G_{\mathbf{y}}\right)\le\sum_{\mathbf{r}\in k\mathcal{R},\ \mathbf{s}\in\mathcal{S}(\mathbf{r})}\exp\left(-\sum_{i=1}^{d}f_i(r_i,s_i)\cdot(n-w_i)+o_d(n)\right),$$

where

$$f_i(r_i,s_i)=f\left(\frac{w-w_i}{n-w_i},\frac{s_i}{n-w},\frac{r_i-s_i}{w-w_i}\right).$$

Proof. For each $\mathbf{y}$, we let $r_i:=k|\mathbf{y}_i|$ and bound $\tilde{\mathbb{P}}^{M}(G_{\mathbf{y}})$ from above using (32), Lemma 7.6 and the law of total probability. The resulting upper bound is

$$\begin{aligned}\tilde{\mathbb{P}}^{M}(G_{\mathbf{y}})\le\sum_{\mathbf{s}\in\mathcal{S}(\mathbf{r})}\exp\Bigg(&-\frac{1}{d}\sum_{i=1}^{d}\mathrm{H}\left(\frac{s_i}{n-w}\right)(n-w)+\sum_{i=1}^{d}\mathrm{H}\left(\frac{s_i}{n-w}\right)(n-w)\\&+\sum_{i=1}^{d}\mathrm{H}\left(\frac{r_i-s_i}{w-w_i}\right)(w-w_i)-\sum_{i=1}^{d}\mathrm{H}\left(\frac{r_i}{n-w_i}\right)(n-w_i)+o_d(n)\Bigg).\end{aligned}$$

However, the number of vectors $\mathbf{y}\in\mathbb{Z}_2^{E\setminus F}$ with given weights $|\mathbf{y}_i|=r_i/k$ is at most

$$\exp\left(\sum_{i=1}^{d}\mathrm{H}\left(\frac{r_i}{n-w_i}\right)\frac{n-w_i}{k}\right).$$

Therefore, the sum of $\tilde{\mathbb{P}}^{M}(G_{\mathbf{y}})$ over all $\mathbf{y}$ satisfying $(|\mathbf{y}_1|,\ldots,|\mathbf{y}_d|)\in\mathcal{R}$ is at most

$$\sum_{\mathbf{r}\in k\mathcal{R},\ \mathbf{s}\in\mathcal{S}(\mathbf{r})}\exp\left(-\sum_{i=1}^{d}f_i(r_i,s_i)\cdot(n-w_i)+o_d(n)\right),\qquad(35)$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

where

$$\begin{aligned}
f_i(r_i,s_i) &:= -\frac{1}{k}\mathrm{H}\left(\frac{r_i}{n-w_i}\right)+\frac{1}{d}\cdot\frac{n-w}{n-w_i}\cdot\mathrm{H}\left(\frac{s_i}{n-w}\right)\\
&\qquad -\frac{n-w}{n-w_i}\cdot\mathrm{H}\left(\frac{s_i}{n-w}\right)-\frac{w-w_i}{n-w_i}\cdot\mathrm{H}\left(\frac{r_i-s_i}{w-w_i}\right)+\mathrm{H}\left(\frac{r_i}{n-w_i}\right)\\
&= \left(1-\frac{1}{k}\right)\cdot\mathrm{H}\left(\frac{r_i}{n-w_i}\right)\\
&\qquad -\left(1-\frac{1}{d}\right)\cdot\frac{n-w}{n-w_i}\cdot\mathrm{H}\left(\frac{s_i}{n-w}\right)-\frac{w-w_i}{n-w_i}\cdot\mathrm{H}\left(\frac{r_i-s_i}{w-w_i}\right)\\
&= f\left(\frac{w-w_i}{n-w_i},\frac{s_i}{n-w},\frac{r_i-s_i}{w-w_i}\right).
\end{aligned}$$

□

We are nearly ready to prove Proposition 7.1. For that proof, we must combine Lemma 7.7 with an elementary but rather fiddly estimate. That estimate refers to the functions

$$\gamma_1(t):=\frac{1}{(\log(1/t))^{1/3}}\quad\text{and}\quad\gamma_2(t):=\frac{1}{(\log(1/t))^{2/3}},$$

both for $0<t<1$. The exponents $1/3$ and $2/3$ are not particularly special here: all we really need is the ordering $0<1/3<2/3<1$. The next lemma gives a collection of simple bounds on the quantity $f(t,\alpha',\alpha'')$ for different ranges of the arguments. Each part requires that t is sufficiently small in terms of d and k. The quantities $t_0^{(\mathrm{a})},t_0^{(\mathrm{b})}$, $t_0^{(\mathrm{c})}$ and $t_0^{(\mathrm{d})}$ are unspecified positive numbers that are sufficiently small in terms of only d and k. Recall we assume $k>d\geq 3$.

Lemma 7.8. *Write $\alpha:=(1-t)\alpha'+t\alpha''$. The function f from Lemma 7.7 satisfies the following.*

a. *If $t<t_0^{(\mathrm{a})}$ and either $\alpha''\leq\alpha'\leq 1/2$ or $\alpha''\geq\alpha'>1/2$, then*

$$f(t,\alpha',\alpha'')\gtrsim_{d,k}\mathrm{H}(\alpha).$$

(This includes the assertion that the left-hand side is nonnegative; see Section 1.6.)

b. *If $t<t_0^{(\mathrm{b})}$ and $t\gamma_p(t)\leq\alpha'\leq 1-t\gamma_p(t)$, then*

$$f(t,\alpha',\alpha'')\gtrsim_{d,k}t\cdot(\log(1/t))^{1-p/3}.$$

c. *If $t<t_0^{(\mathrm{c})}$ and $\gamma_p(t)\leq\alpha''\leq 1-\gamma_p(t)$, then*

$$f(t,\alpha',\alpha'')\gtrsim_{d,k}t\cdot(\log(1/t))^{1-p/3}.$$

d. *If $t<t_0^{(\mathrm{d})}$ and*

$$[\ \alpha'<t\gamma_p(t)\ \text{or}\ \alpha'>1-t\gamma_p(t)]\ \text{and}\ [\ \alpha''<\gamma_p(t)\ \text{or}\ \alpha''>1-\gamma_p(t)\ ],$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

then

$$\max\{0,-f(t,\alpha',\alpha'')\}\lesssim_{d,k} t\cdot(\log(1/t))^{1-p/3}.$$

(The maximum is used here to maintain the nonnegativity convention for $\lesssim$*.)*

Observe that at least one of the parts (b), (c) or (d) must hold whenever $t<\min\{t_0^{(\mathrm{b})},t_0^{(\mathrm{c})},t_0^{(\mathrm{d})}\}$.

Proof. Each part of this lemma is symmetric under replacing (α',α'') with $(1-\alpha',1-\alpha'')$, and so is the function f. We therefore assume that $\alpha'\le 1/2$ throughout the proof.

Part (a). By the concavity of H, we have

$$\begin{aligned}&(1-t)(1-d^{-1})\,\mathrm{H}(\alpha')+t\,\mathrm{H}(\alpha'')\\&\le\big((1-t)(1-d^{-1})+t\big)\,\mathrm{H}\left(\frac{(1-t)(1-d^{-1})}{(1-t)(1-d^{-1})+t}\alpha'+\frac{t}{(1-t)(1-d^{-1})+t}\alpha''\right).\end{aligned}$$

Since $1-d^{-1}<1$, the convex combination inside the argument of H here skews more towards α'' than does the convex combination that gives α. Therefore, since $\alpha''\le\alpha'\le 1/2$ and H is increasing on $[0,1/2]$, the right-hand side above is bounded above by

$$\big((1-t)(1-d^{-1})+t\big)\,\mathrm{H}(\alpha)=(1-d^{-1}+td^{-1})\,\mathrm{H}(\alpha).$$

Therefore,

$$f(t,\alpha',\alpha'')\ge\big((1-k^{-1})-(1-d^{-1}+td^{-1})\big)\,\mathrm{H}(\alpha)=(d^{-1}-k^{-1}-td^{-1})\,\mathrm{H}(\alpha),$$

which is $\gtrsim_{d,k}\mathrm{H}(\alpha)$ provided $t<t_0^{(\mathrm{a})}$.

Part (b). For this part, our assumptions are now $\alpha'\le 1/2$ and $\alpha'\ge t\gamma_p(t)$. Let $c_1:=(d^{-1}-k^{-1})/2>0$. Since $\alpha'\le 1/2$ and H is continuous on $[0,1]$ and increasing on $[0,1/2]$, we have that

$$\mathrm{H}((1-t)\alpha'+t\alpha'')\ge\frac{1-d^{-1}+c_1}{1-k^{-1}}\,\mathrm{H}(\alpha')\quad\text{whenever } t<t_0^{(\mathrm{b})}. \tag{36}$$

If $\alpha'\ge t\gamma_p(t)$ and $t<t_0^{(\mathrm{b})}$, then

$$\begin{aligned}\mathrm{H}(\alpha')&\ge\mathrm{H}(t\gamma_p(t))\ge t\cdot\gamma_p(t)\cdot\big(\log(1/\gamma_p(t))+\log(1/t)\big)\\&\ge t\cdot\gamma_p(t)\cdot\log(1/t)=t\cdot(\log(1/t))^{1-p/3}.\end{aligned} \tag{37}$$

Combining (36) and (37), we obtain

$$\begin{aligned}f(t,\alpha',\alpha'')&\ge(1-d^{-1}+c_1)\,\mathrm{H}(\alpha')-(1-t)(1-d^{-1})\,\mathrm{H}(\alpha')-t\,\mathrm{H}(\alpha'')\\&\ge c_1\,\mathrm{H}(\alpha')-t\,\mathrm{H}(\alpha'')\\&\ge c_1\cdot t\cdot(\log(1/t))^{1-p/3}-\log 2\cdot t\\&\gtrsim_{d,k} t\cdot(\log(1/t))^{1-p/3}\qquad\qquad \text{if } t<t_0^{(\mathrm{b})}.\end{aligned}$$

Part (c). For this part, our assumptions are now $\alpha'\le 1/2$ and $\alpha''\ge\gamma_p(t)$. We may also assume that $\alpha'<t\gamma_p(t)$, for otherwise, part (b) already gives the desired bound.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

If, in addition, we have $\alpha' \geq \alpha''$, then part (a) gives

$$f(t,\alpha',\alpha'') \gtrsim_{d,k} \mathrm{H}(\alpha),$$

and this in turn satisfies

$$\mathrm{H}(\alpha) \geq \mathrm{H}(t\alpha'') \geq t \cdot \alpha'' \cdot \log(1/t) \geq t \cdot \gamma_p(t) \cdot \log(1/t) = t \cdot (\log(1/t))^{1-p/3}. \tag{38}$$

So for the rest of this part, assume in addition that $\alpha' \leq \alpha''$. Then for sufficiently small t, we must have the ordering $\alpha' \leq \alpha < 1/2$, and so

$$\begin{aligned}(1-t)(1-d^{-1})\mathrm{H}(\alpha') + t\mathrm{H}(\alpha'') &\leq (1-d^{-1})\mathrm{H}(\alpha') + t\mathrm{H}(\alpha'') \\ &\leq (1-d^{-1})\mathrm{H}(\alpha) + t\mathrm{H}(\alpha'').\end{aligned}$$

Therefore, in this case, it suffices to show that

$$(d^{-1} - k^{-1})\mathrm{H}(\alpha) - t\mathrm{H}(\alpha'') \gtrsim_{d,k} t \cdot (\log(1/t))^{1-p/3}.$$

Since the second left-hand term here is $O(t)$, this follows by another use of (38).

Part (d). For this case, we simply neglect the positive term in f entirely. If

$$t < t_0^{(\mathrm{d})}, \quad \alpha' < t \cdot \gamma_p(t) \quad \text{and} \quad \alpha'' < \gamma_p(t),$$

then

$$\mathrm{H}(\alpha') \lesssim_{d,k} t \cdot \gamma_p(t) \cdot \log\frac{1}{t} = t \cdot (\log(1/t))^{1-p/3}$$

and

$$t\mathrm{H}(\alpha'') = O(t) \lesssim_{d,k} t \cdot (\log(1/t))^{1-p/3}.$$

In the cases where $\alpha' > 1 - t \cdot \gamma_p(t)$ or $\alpha'' > 1 - \gamma_p(t)$, the same estimates hold, by the symmetry $\mathrm{H}(x) = \mathrm{H}(1-x)$. Adding these estimates gives the conclusion. □

Corollary 7.9. *Fix $K \geq 1$ and $\delta > 0$, and let the other notation be as for Lemma 7.8. If δ is sufficiently small in terms of d, k and K, and if*

$$t \leq K\delta \quad \text{and} \quad \delta \cdot \gamma_1(\delta) \leq \alpha \leq 1 - \delta \cdot \gamma_1(\delta),$$

then

$$f(t,\alpha',\alpha'') \gtrsim_{d,k,K} \delta \cdot (\log(1/\delta))^{2/3}$$

(irrespective of any further bounds on α' and α'').

Proof. By the same symmetry as for Lemma 7.8, we may assume that $\alpha' \leq 1/2$.

Having done so, suppose first that $\alpha'' \leq \alpha'$. Then part (a) of Lemma 7.8 gives

$$f(t,\alpha',\alpha'') \gtrsim_{d,k} \mathrm{H}(\alpha),$$

and our assumed range for α gives

$$\mathrm{H}(\alpha) \geq \mathrm{H}(\delta \cdot \gamma_1(\delta)) \geq \delta \cdot (\log(1/\delta))^{2/3}, \tag{39}$$

giving a lower bound of the desired form.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

So now suppose that $\alpha' \leq 1/2$ and $\alpha'' \geq \alpha'$. Since $t \leq K\delta$, it follows that

$$\alpha' \leq \alpha \leq (1-K\delta)\alpha' + K\delta.$$

Provided δ is sufficiently small in terms of d and k, this range of possible values for α' and α implies that

$$(1-d^{-1})\,\mathrm{H}(\alpha') \leq \left(1 - \frac{d^{-1}+k^{-1}}{2}\right)\mathrm{H}(\alpha)$$

(noting that the constant in front of the entropy on the left is slightly smaller than the constant in front of the entropy on the right). Rearranging, this implies that

$$(1-k^{-1})\,\mathrm{H}(\alpha) - (1-d^{-1})\,\mathrm{H}(\alpha') \geq \frac{d^{-1}-k^{-1}}{2}\,\mathrm{H}(\alpha) \geq \frac{d^{-1}-k^{-1}}{2}\cdot\delta\cdot(\log(1/\delta))^{2/3},$$

using again the lower bound (39). This now gives

$$\begin{aligned} f(t,\alpha',\alpha'') &\geq \frac{d^{-1}-k^{-1}}{2}\cdot\delta\cdot(\log(1/\delta))^{2/3} - t\big(\mathrm{H}(\alpha'') - (1-d^{-1})\,\mathrm{H}(\alpha')\big) \\ &\geq \frac{d^{-1}-k^{-1}}{2}\cdot\delta\cdot(\log(1/\delta))^{2/3} - \log 2\cdot K\cdot\delta. \end{aligned}$$

This implies the desired lower bound on f for all sufficiently small δ. □

Proof of Proposition 7.1. Fix K and ε and also F_i, W_i, w_i and M as in the statement of Proposition 7.1, and suppose that (25) holds. We prove the convergence to zero of the required probabilities provided that δ is small enough in terms of d, k, ε and K.

First, assume δ is small enough that

$$\gamma_1(\delta) \leq \varepsilon, \tag{40}$$

and also small enough in terms of d, k and K that Corollary 7.9 applies.

By Lemma 7.7, and since a small choice of δ ensures that $n - w_i \geq n/2$ for each i, it suffices to show that, if δ is sufficiently small, then the negative exponent

$$\sum_{i=1}^{d} f\left(\frac{w-w_i}{n-w_i}, \frac{s_i}{n-w}, \frac{r_i - s_i}{w-w_i}\right) \tag{41}$$

is bounded below by a positive quantity that is independent of $\mathbf{r} \in k\mathcal{R}$ and $\mathbf{s} \in \mathcal{S}(\mathbf{r})$.

So fix $\mathbf{r}$ and $\mathbf{s}$, and let

$$(t_i, \alpha_i', \alpha_i'') := \left(\frac{w-w_i}{n-w_i}, \frac{s_i}{n-w}, \frac{r_i-s_i}{w-w_i}\right)$$

and

$$\alpha_i = (1-t_i)\alpha_i' + t_i\alpha_i'' = \frac{r_i}{n-w_i}.$$

For each i, this implies that

$$t_i \leq \frac{w}{n} \leq K\delta.$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Assume δ is also small enough that $K\delta < \min\{t_0^{(\mathrm{a})}, t_0^{(\mathrm{b})}, t_0^{(\mathrm{c})}, t_0^{(\mathrm{d})}\}$, so each t_i is also less than this minimum.

Classify the indices $i \in [d]$ into two subsets:

$$\begin{aligned} I_1 &= \{i \in [d] : \ \varepsilon\delta \le r_i/(n-w_i) \le 1-\varepsilon\delta\} \\ I_2 &= [d] \setminus I_1. \end{aligned}$$

The definition of $\mathcal{R}$ implies that $I_1 \neq \emptyset$. Consider the terms in (41) for indices in these subsets:

- If $i \in I_1$, then, by our choice of δ in (40) and since $t_i \le K\delta$, Corollary 7.9 gives
$$f\left(\frac{w-w_i}{n-w_i}, \frac{s_i}{n-w}, \frac{r_i - s_i}{w-w_i}\right) \ge C_1 \cdot \delta \cdot (\log(1/\delta))^{2/3}$$
for some positive constant C_1 depending only on d, k and K.
- Next, for $i \in I_2$, we use that at least one of parts (b), (c), and (d) of Lemma 7.8 must hold, which gives that
$$\begin{aligned} \max\left\{0, -f\left(\frac{w-w_i}{n-w_i}, \frac{s_i}{n-w}, \frac{r_i - s_i}{w-w_i}\right)\right\} &\lesssim_{d,k} t_i \cdot (\log(1/t_i))^{1/3} \\ &\lesssim_K \delta \cdot (\log(1/\delta))^{1/3} \end{aligned}$$
provided δ is sufficiently small, and so the left-hand side is bounded above by $C_2 \cdot \delta \cdot (\log(1/\delta))^{1/3}$ for some positive C_2 depending only on d, k and K.

We bound (41) from below by adding these estimates. At least one term has $i \in I_1$, and there are at most $d-1$ terms with $i \in I_2$. This leaves the lower bound

$$C_1 \cdot \delta \cdot (\log(1/\delta))^{2/3} - (d-1) \cdot C_2 \cdot \delta \cdot (\log(1/\delta))^{1/3}.$$

This is positive for all sufficiently small δ, uniformly over different choices of $\mathbf{r}$ or $\mathbf{s}$, as required. □

8. Proof of local and empirical convergence

In this section, we prove Theorem 6.1(2): we show that if $(\Omega_n')_n$ is the sequence given by Theorem 6.1(1) and $\sigma_n \in \Omega_n'$ for each n, then the measures μ_n converge locally and empirically in probability to the Haar measure m.

Recall $X_n = X_{\sigma_n}$ is the set of parity check codewords over $\sigma_n \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(V_n))$. For $v \in V_n$, let

$$X_{n,v} = \{\Pi_v^{\sigma_n}\mathbf{x} : \mathbf{x} \in X_n\}$$

be the set of pullback names at v of all the codewords in X_n. As above, let

$$\mathrm{Loc}(\mu_n, v) = (\Pi_v^{\sigma_n})_* \mu_n \in \mathrm{Prob}(X).$$

Let $X_{n,v,r}$ be the projection of $X_{n,v}$ onto $\mathbb{Z}_2^{B_r}$. Call a vertex $v \in V_n$ r**-proper** if $X_{n,v,r} = X_{B_r}$. Otherwise, call it r**-improper**.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

The following lemma is a (stronger) version of [61, Lemma 3.47] for our random factor graph model.

Lemma 8.1. *For any $r,\varepsilon > 0$*

$$\mathbb{P}\left(\frac{1}{|V_n|}|\{v \in V_n : v \textit{ is } r\textit{-improper}\}| \geq \varepsilon\right) \to 0$$

as $n \to \infty$.

Proof. In general, $X_{n,v,r}$ is a vector subspace of X_{B_r}. If v is r-improper, then it is a subspace of strictly smaller dimension, so

$$\mathrm{H}(\mathrm{Loc}(\mu_n,v)_{B_r}) \leq \log|X_{n,v,r}| = (\dim X_{n,v,r})\log 2 \leq (\dim X_{B_r} - 1)\log 2 = \mathrm{H}(m_{B_r}) - \log 2. \tag{42}$$

With ε,r as given, pick $\eta = \frac{\varepsilon \log 2}{4}$. By Theorem 7.4, for small enough $\delta > 0$, if for each n we pick a subset R_n of V_n of size $\lceil \delta n \rceil$, then with high probability, σ_n satisfies

$$\mathrm{H}((\mu_n)_{\sigma_n^{B_r} \cdot R_n}) \geq (\mathrm{H}(m_{B_r}) - \eta)\,|R_n|.$$

Now for the sake of contradiction, suppose that

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{|V_n|}|\{v \in V_n : v \text{ is } r\text{-improper}\}| \geq \varepsilon\right) > 0.$$

Then, by symmetry of the law of σ_n, and using that $|R_n| \geq \delta |V_n|$, the probability that the fraction of r-improper vertices within R_n is at least $\frac{\varepsilon}{2}$ is uniformly bounded below for infinitely many n. But if an $\frac{\varepsilon}{2}$ fraction of vertices of R_n are r-improper, then by (42) and subadditivity of Shannon entropy,

$$\mathrm{H}((\mu_n)_{\sigma_n^{B_r} \cdot R_n}) \leq \left(\mathrm{H}(m_{B_r}) - \tfrac{\varepsilon}{2}\log 2\right)|R_n|.$$

Combining with the above, this implies that with nonvanishing probability,

$$(\mathrm{H}(m_{B_r}) - \eta)\,|R_n| \leq \left(\mathrm{H}(m_{B_r}) - \tfrac{\varepsilon}{2}\log 2\right)|R_n|.$$

But this is false by choice of η, so it must be that

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{|V_n|}|\{v \in V_n : v \text{ is } r\text{-improper}\}| \geq \varepsilon\right) = 0,$$

as desired. □

Lemma 8.2. *Assume $k \geq 3$. Then the action of Γ on (X,m_X) is mixing and hence ergodic.*

Proof. We prove this using Fourier analysis on the compact Abelian group X.

Step 1. For any $g \in \Gamma$, let $|g|$ be its word length in the generating set $\{s_i^j : i = 1,\ldots,d,\ j = 1,\ldots,k-1\}$. The characters of X form an orthonormal basis for $L^2_{\mathbb{C}}(m_X)$, and

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

all non-identity characters have mean zero. It therefore suffices to prove that, for any two non-identity characters χ_1 and χ_2, we have

$$\langle \chi_1, \chi_2 \circ T^g \rangle = 0$$

whenever $|g|$ is sufficiently large.

Step 2. Every character of X is the restriction of a character of $\mathbb{Z}_2^\Gamma$, and all of these have the form

$$\chi_U(\mathbf{x}) := (-1)^{\sum_{g\in U} x_g} \qquad (\mathbf{x} \in \mathbb{Z}_2^\Gamma)$$

for some finite subset U of Γ. In addition, the action T is by group automorphisms and satisfies $\chi_U \circ T^g = \chi_{gU}$. Therefore, $\chi_U \cdot (\chi_W \circ T^g) = \chi_{U \triangle gW}$ for any U, W and g, where $\triangle$ denotes symmetric difference.

Let E be the set of all k-cycles of the form $\{g, gs_i, \ldots, gs_i^{k-1}\}$ in the Cayley graph of Γ, where $g \in \Gamma$ and $i \in \{1, \ldots, d\}$. As in the Introduction, we can regard E as the set of k-hyper-edges of a hypergraph on Γ, and the corresponding characters χ_e for $e \in E$ give the parity checks that define the LDPC subgroup X. Therefore, by Pontrjagin duality, a character χ_U restricts to the identity character on X if and only if there is a finite subfamily F of E such that

$$1_U = \sum_{e\in F} 1_E \mod 2.$$

Let us write $X^\perp$ for the collection of finite subsets U that have this property. Regarded as a subspace of $\mathbb{Z}_2^{\oplus\Gamma}$, this $X^\perp$ is the linear span of the set $\{1_e : \ e \in E\}$.

In these terms, we must show that if U and W are finite subsets of Γ and neither of them lies in $X^\perp$, then $U \triangle gW$ also does not lie in $X^\perp$ whenever $|g|$ is sufficiently large. Fix such U and W for the rest of the proof.

Step 3. Let B_r and B_s be closed balls around the identity in the right Cayley graph that contain U and W, respectively. More specifically, in this last step of the proof, we assume that $g \in \Gamma$ satisfies both (i) $|g| > r + s + 2$ and also (ii) $U \triangle gW \in X^\perp$, and derive a contradiction from these. Assumption (i) implies U and gW are disjoint, and therefore, $U \triangle gW = U \cup gW$. Having made assumption (ii), let F be a finite subfamily of E such that

$$1_{U\triangle gW} = \sum_{e\in F} 1_e \mod 2. \tag{43}$$

To work with elements of $X^\perp$, it helps to introduce the dual graph $(E, \tilde{E})$ of the hypergraph (Γ, E). The vertices of the dual graph are the hyperedges in E, and two hyperedges e_1 and e_2 are joined in $\tilde{E}$ if and only if $e_1 \cap e_2 \neq \emptyset$.

In the dual graph, F is the union of its connected components. At least one of these must meet both U and gW. Indeed, otherwise we could let G be the union of those connected components of F that meet U and would then find that 1_U agrees with $\sum_{e\in G} 1_e$ mod 2, contradicting our assumption that $U \notin X^\perp$.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Next, because the hypergraph (Γ,E) is a hyper-tree and we have $|g| > r+s+2$, there is a single hyperedge e_0 lying 'between' B_r and gB_s in the following sense: removing this hyperedge e_0 disconnects that hypergraph into k components, one of which contains the whole of B_r (and hence U), and a different one of which contains the whole of gB_s (and hence gW).

As a result, if F_0 is a connected component of F in the dual graph which meets both U and gW, then F_0 must contain e_0. Now consider again the k connected components of the hypergraph $(\Gamma, E\setminus e_0)$. Since $k\geq 3$, at least one of them does not meet either U or gW; let $V\subset\Gamma$ be one such component. Then V meets e_0 in a single vertex (that is, group element) h.

Finally, let h' be a vertex of $V\cap\bigcup_{e\in F} e$ that lies at maximal distance from h. This h' may be equal to h, or it may lie 'deeper inside' the component V. However, by that distance maximality and the hyper-tree structure of (Γ,E), h' can be contained in only one member of F, and because $h'\in V$, it cannot lie in either U or gW. So at h', the left-hand side of (43) must equal zero, while the right-hand side must equal 1. This is the desired contradiction with that equation. □

Proof of Theorem 6.1(2). Let $\mathrm{Loc}(\mu_n,v)_{B_r}$ denote the marginal of $\mathrm{Loc}(\mu_n,v)$ on $X_{n,v,r}$.

Since $\mathrm{Loc}(\mu_n,v)_{B_r}$ is the uniform distribution on $X_{n,v,r}$ and m_{B_r} is the uniform distribution on X_{B_r}, local convergence in probability to μ is implied by

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{1}{|V_n|}|\{v\in V_n : X_{n,v,r}\neq X_{B_r}\}|>\varepsilon\right)=0,$$

which is true by Lemma 8.1.

Since $(\mu_n)_n$ converges locally in probability to m_X, which is ergodic by Lemma 8.2, Lemma 4.1 completes the proof. □

9. Proof of Theorem 1.3

Next, we prove the upper bound in Theorem 1.3, part (a). After this, part (b) of the Theorem 1.3 is an immediate consequence of Theorems 6.1 (items 1,2) and 4.3.

9.1. Proof of the sofic entropy value

Proof of Theorem 1.3, part (a). Let $\sigma_n\sim\mathbb{P}_n$ and let $\Omega_n'\subseteq\Omega_n^{\mathrm{sofic}}$ be as in Proposition 3.2. Let $\Sigma=\{\sigma_n\}_{n=1}^{\infty}$ satisfy $\sigma_n\in\Omega_{i_n}'$ for some increasing sequence $(i_n)_n$ with $k\mid i_n$ for all n. It suffices to prove

$$\mathrm{h}_\Sigma(X,m_X,T)=(1-d/k)\log 2.$$

By definition,

$$\mathrm{H_K}(m_X)=(1-d)\,\mathrm{H}(W_{m_X}(\cdot))+\frac{1}{k}\sum_{i\in[d]}\mathrm{H}(W_{m_X}(\cdot;i)).$$

The weight W_{m_X} is uniform on $\mathbb{Z}_2$. So $\mathrm{H}(W_{m_X}(\cdot)) = \log(2)$. For each $i \in [d]$, the measure $W_{m_X}(\cdot;i)$ is supported on the subspace of $\mathbb{Z}_2^k$ which is the kernel of the homomorphism $(x_1,\ldots,x_k) \mapsto \sum_i x_i \in \mathbb{Z}_2$. This subspace has cardinality 2^{k-1}. So

$$\mathrm{H}(W_{m_X}(\cdot;i)) \le (k-1)\log(2).$$

Combined with the previous formula, we obtain

$$\mathrm{H}_{\mathrm{K}}(m_X) \le (1-d/k)\log(2).$$

The conclusion of Proposition 3.2 now implies the upper bound

$$\mathrm{h}_\Sigma(X,m_X,T) \le (1-d/k)\log 2.$$

By Theorem 6.1(2), μ_n converges empirically to m_X. Therefore, if $\mathcal{O}$ is any open neighborhood of m_X, then

$$1 = \lim_{n\to\infty} \mu_n(\Omega(\sigma_n,\mathcal{O})).$$

Since μ_n is the uniform measure on X_{σ_n}, this implies

$$\mathrm{h}_\Sigma(X,m_X,T) \ge \lim_{n\to\infty} \frac{1}{i_n}\log|X_{\sigma_n}|.$$

We may estimate $|X_{\sigma_n}|$ by simple dimension-counting. In fact, X_{σ_n} is the kernel of a homomorphism from $\mathbb{Z}_2^{i_n}$ to $\mathbb{Z}_2^{di_n/k}$. So

$$|X_{\sigma_n}| \ge 2^{(1-d/k)i_n}.$$

This gives the lower bound

$$\mathrm{h}_\Sigma(X,m_X,T) \ge (1-d/k)\log 2. \qquad \square$$

10. Proof of Theorems B and C

Theorem B is an immediate consequence of Theorem 6.1(3) and Corollary 5.5(2).

10.1. Expected number of low-density codewords

In this section, we derive an asymptotic formula for the expected number of low-density codewords. Versions of this appear in [49, Formula (11.10)] and [61, Lemma 3.163]. The authors of [49] and [61] use a configuration model instead of a permutation model. As discussed above, these models are not contiguous (see Appendix A). Moreover, we are able to show a stronger result than contiguity alone would imply.

For $\eta > 0$ and $\sigma \in \mathrm{Hom}_{\mathrm{unif}}(\Gamma, \mathrm{Sym}(kn))$, let X_σ^η denote the set of $\mathbf{x} \in (\mathbb{Z}_2)^{kn}$ for which the sum around all but at most a fraction η of each of the d hyper-edges types is even. More formally,

$$X_\sigma^\eta = \bigcap_{i\in[d]} \left\{ \mathbf{x} \in (\mathbb{Z}_2)^{kn} \,:\, \left|\left\{ v \in [kn] \,:\, \sum_{j=0}^{k-1} \mathbf{x}(\sigma_i^j v) = 0 \pmod 2 \right\}\right| > kn(1-\eta) \right\}.$$

We can think of these as ‘approximate codewords’.

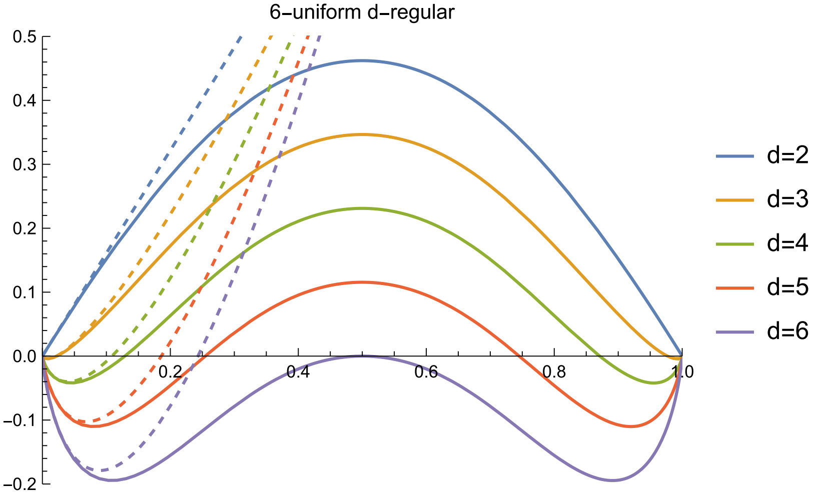

Figure 2. Comparison of $G_{\mathrm{cw}}(t)$ (solid lines) with asymptotic in Proposition 10.1 (dashed lines) for $k = 6$ and several choices of d.

For $t \in [0,1]$, we define the upper exponential growth rate of the expected number of approximate codewords of density t by

$$G_{\mathrm{cw}}(t) = \inf_{\varepsilon,\eta>0} \limsup_{n\to\infty} \frac{1}{kn} \log \mathbb{E}_{\sigma\sim\mathbb{P}_{kn}} |\{\mathbf{x} \in X^{\eta}_{\sigma} : \tfrac{1}{kn}|\mathbf{x}| \in (t-\varepsilon, t+\varepsilon)\}|.$$

Proposition 10.1. *For any d,k,*

$$G_{\mathrm{cw}}(t) = \frac{1}{2} t\big(d\log(k-1) - d + 2 + (d-2)\log(t)\big) + O(t^2).$$

In particular, for $k \geq 2$, this is negative for small $t > 0$ if and only if $d > 2$.

The '$O(t^2)$' term here is a power series convergent on some neighborhood of $t = 0$ with lowest-order term t^2.

Figure 2 compares exact plots of $G_{\mathrm{cw}}(t)$ (created using parametric plots in the parameter s of Lemma 10.2) with plots of this approximation.

The proof of Proposition 10.1 is based on the following lemma.

Lemma 10.2. *For any k,d and any $t \in [0,1]$,*

$$G_{\mathrm{cw}}(t) = (1-d)\,\mathrm{H}(t) + \frac{d}{k}(-kt\log s + \log Z),$$

where s and Z are related to t via

$$t = s\frac{(1+s)^{k-1} - (1-s)^{k-1}}{(1+s)^k + (1-s)^k} \quad \textit{and} \quad Z = \frac{1}{2}\left((1+s)^k + (1-s)^k\right).$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Proof of Lemma 10.2. Here, let $\mathtt{A} = \mathbb{Z}_2$. We also let $\mathcal{W}$ denote the set of weights which have edge weights that are cyclically invariant and supported on configurations with even parity. The terminology of weights was defined in Section 3.2.

We first show that, for every $t \in [0,1]$,

$$G_{\mathrm{cw}}(t) = \sup_{W \in \mathcal{W} : W(1) = t} \mathrm{H}_{\mathrm{K}}(W). \tag{44}$$

First, suppose $W \in \mathcal{W}$ has $W(1) = t$. For every $\varepsilon, \eta > 0$, if $\delta > 0$ is small enough, then we always have

$$\{\mathbf{x} \in (\mathbb{Z}_2)^{kn} : \|W_{\sigma,\mathbf{x}} - W\| < \delta\} \subseteq \{\mathbf{x} \in X_\sigma^\eta : \tfrac{1}{kn}|\mathbf{x}| \in (t-\varepsilon, t+\varepsilon)\}.$$

By Proposition 3.1, this implies that $\mathrm{H}_{\mathrm{K}}(W) \le G_{\mathrm{cw}}(t)$, which gives one half of Equation 44.

For the converse inequality: Given $\varepsilon, \eta > 0$, let $\mathcal{W}_{\varepsilon,\eta}$ be the set of cyclically-invariant weights with $|W(1) - t| < \varepsilon$ and each $W(\cdot; i)$ giving mass at least $1 - \eta$ to even-parity configurations.

Given $\delta > 0$, by Proposition 3.1 for each weight W, there is some $r_W > 0$ which satisfies

$$\limsup_{n\to\infty} \frac{1}{kn} \log \mathbb{E}_\sigma |\{\mathbf{x} \in (\mathbb{Z}_2)^{kn} : \|W_{\sigma,\mathbf{x}} - W\| < r_W\}| \le \mathrm{H}_{\mathrm{K}}(W) + \delta.$$

By compactness, there is a finite set $\mathcal{S} \subset \overline{\mathcal{W}_{\varepsilon,\eta}}$ such that the balls centered at $W \in \mathcal{S}$ of radius r_W cover $\mathcal{W}_{\varepsilon,\eta}$. Then we have

$$\{\mathbf{x} \in X_\sigma^\eta : \tfrac{1}{kn}|\mathbf{x}| \in (t-\varepsilon, t+\varepsilon)\} \subseteq \bigcup_{W \in \mathcal{S}} \{\mathbf{x} \in (\mathbb{Z}_2)^{kn} : \|W_{\sigma,\mathbf{x}} - W\| < r_W\}.$$

Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{kn} \log \mathbb{E}_\sigma |\{\mathbf{x} \in X_\sigma^\eta : \tfrac{1}{kn}|\mathbf{x}| \in (t-\varepsilon, t+\varepsilon)\}| \\
&\quad \le \max_{W \in \overline{\mathcal{W}_{\varepsilon,\eta}}} \limsup_{n\to\infty} \frac{1}{kn} \log \mathbb{E}_\sigma |\{\mathbf{x} \in (\mathbb{Z}_2)^{kn} : \|W_{\sigma,\mathbf{x}} - W\| < r_W\}| \\
&\quad \le \max_{W \in \overline{\mathcal{W}_{\varepsilon,\eta}}} \mathrm{H}_{\mathrm{K}}(W) + \delta.
\end{aligned}$$

Taking ε, η and then δ to 0 gives the other half of Equation 44.

Now any weight achieving the supremum in 44 must have all edge weights equal: to maximize H_{K} under the constraint $W(1) = t$, whichever $W(\cdot; i)$ maximizes the edge term in the definition of $\mathrm{H}_{\mathrm{K}}(W)$ should be used for all $i \in [d]$.

We can specify such a weight by a single probability vector $\mathbf{p} \in \mathrm{Prob}(\mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k)$ recording the edge weight. We use a probability measure on $\mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k$ rather than $\mathtt{A}^{\mathbb{Z}_k}$ because the edge weights must be invariant under cyclic permutations. Let $W_{\mathbf{p}}$ be the weight with edge weights specified by $\mathbf{p}$, and write $\alpha(\mathbf{p}) = W_{\mathbf{p}}(1)$. The Kikuchi entropy of $W_{\mathbf{p}}$ is

$$\mathrm{H_K}(W_{\mathbf{p}}) = (1-d)\,\mathrm{H}(\alpha(\mathbf{p})) + \frac{d}{k}\left(\mathrm{H}(\mathbf{p}) + \sum_{[\mathbf{a}]\in \mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k} \mathbf{p}([\mathbf{a}])\log|[\mathbf{a}]|\right),$$

where $|[\mathbf{a}]|$ is the number of elements of the equivalence class $[\mathbf{a}] \in \mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k$.

Now, we are interested in estimating

$$G_{\mathrm{cw}}(t) = \max_{\mathbf{p}\,:\,\alpha(\mathbf{p})=t} \mathrm{H_K}(W_{\mathbf{p}}),$$

where the maximum is also constrained to $\mathbf{p}$ supported on equivalence classes of even-parity configurations. Since $\alpha(\mathbf{p})$ is fixed to be t, we really just need to maximize

$$\mathrm{H}(\mathbf{p}) + \sum_{[\mathbf{a}]\in \mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k} \mathbf{p}([\mathbf{a}])\log|[\mathbf{a}]|$$

subject to the constraints

$$\sum_{[\mathbf{a}]} \mathbf{p}([\mathbf{a}])\alpha([\mathbf{a}]) = kt \quad \text{and} \quad \sum_{[\mathbf{a}]} \mathbf{p}([\mathbf{a}]) = 1,$$

where $\alpha([\mathbf{a}])$ is the number of 1's in any representative of $[\mathbf{a}]$. So we get two Lagrange multipliers λ_1, λ_2 so that a maximizer on the interior of the constraint region is given by

$$\mathbf{p}([\mathbf{a}]) = |[\mathbf{a}]| \cdot e^{\lambda_1 \alpha([\mathbf{a}])+\lambda_2-1}$$

for $[\mathbf{a}] \in \mathtt{A}^{\mathbb{Z}_k}/\mathbb{Z}_k$ with even parity and 0 otherwise. We rewrite this as

$$\mathbf{p}([\mathbf{a}]) = |[\mathbf{a}]| \cdot \frac{s^{\alpha([\mathbf{a}])}}{Z},$$

where Z is determined by the normalization constraint and s is determined by the density-t constraint. Since the objective function is strictly concave, the critical point given by these Z, s is in fact the unique maximum. The corresponding $\mathrm{H_K}$-maximizing weight W therefore satisfies

$$W(\mathbf{a}; i) = \frac{s^{\alpha(\mathbf{a})}}{Z}$$

if $\alpha(\mathbf{a})$ (the number of 1's in $\mathbf{a}$) is even and $W(\mathbf{a}; i) = 0$ otherwise.

Note that

$$Z = \frac{1}{2}\left((1+s)^k + (1-s)^k\right) \tag{45}$$

since

$$Z = \sum_{\substack{\mathbf{a}\in \mathtt{A}^k \\ \alpha(\mathbf{a})\text{ even}}} s^{\alpha(\mathbf{a})} = \sum_{\substack{m=0 \\ m\text{ even}}}^{k} s^m \binom{k}{m} = \frac{1}{2}\left(\sum_{m=0}^{k} s^m \binom{k}{m} + \sum_{m=0}^{k} (-s)^m \binom{k}{m}\right).$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

The constraint $\sum_{[\mathbf{a}]} \mathbf{p}([\mathbf{a}])\,\alpha([\mathbf{a}]) = kt$ gives the relation between t and s:

$$kt = \sum_{[\mathbf{a}]\text{ even}} |[\mathbf{a}]| \frac{s^{\alpha([\mathbf{a}])}}{Z} \cdot \alpha([\mathbf{a}]) = s\frac{\frac{dZ}{ds}}{Z} = sk\frac{(1+s)^{k-1}-(1-s)^{k-1}}{(1+s)^k+(1-s)^k}.$$

Finally, we can calculate

$$\mathrm{H}(W(\cdot;i)) = -\sum_{\mathbf{a}\text{ even}} \frac{s^{\alpha(\mathbf{a})}}{Z}\alpha(\mathbf{a})\log s + \log Z = -kt\log s + \log Z,$$

which gives the claimed formula. □

This lemma does not give an explicit formula for $G_{\mathrm{cw}}(t)$: we have t as a function of s, but we do not have an explicit formula for the inverse function. Still, we can use it to prove Proposition 10.1.

Proof of Proposition 10.1. Expanding the formula for t as a power series centered at $s = 0$, we get $t = (k-1)s^2 + O(s^4)$ so, by Lagrange inversion [75, §7.32], there is a power series for s^2 on some interval around 0 of the form

$$s^2 = \frac{t}{k-1} + O(t^2).$$

Hence,

$$t\log s = \frac{1}{2}t\log\frac{t}{k-1} + O(t^2)$$

and

$$\log Z = \frac{1}{2}k(k-1)s^2 + O(s^4) = \frac{1}{2}kt + O(t^2).$$

The estimate for Shannon entropy

$$\mathrm{H}(t) = -t\log t + t + O(t^2)$$

completes the proof. □

10.2. Proof of Theorem 6.1(3)

Recall that μ has totally shattered microstate spaces along $(\sigma_n)_n$ if there exists a $\delta > 0$ for which the following holds: For every $\varepsilon > 0$ there exist a weak* neighbourhood U of μ and a positive integer n_0 such that for any $n \geq n_0$ and any two microstates $\mathbf{x},\mathbf{y} \in \Omega(\sigma_n, U)$, we have either $\mathrm{d}^{(V_n)}(\mathbf{x},\mathbf{y}) \geq \delta$ or $\mathrm{d}^{(V_n)}(\mathbf{x},\mathbf{y}) < \varepsilon$.

Proof of Theorem 6.1(3). Given $k \geq 2$ and $d > 2$, by Proposition 10.1, there is some $\delta > 0$ such that $G_{\mathrm{cw}}(t) < 0$ for $t \in (0,\delta)$. We will use this δ to establish totally shattered microstates.

For any fixed $\varepsilon > 0$,

$$\inf_{\eta>0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}|\{\mathbf{x}\in X_n^\eta : \tfrac{1}{n}|\mathbf{x}| \in [\varepsilon,\delta]\}| = \sup\{G_{\mathrm{cw}}(t) : t \in (\varepsilon,\delta)\} < 0.$$

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

Pick some η such that the expression in the infimum is negative, and let $U^\eta \subset \mathrm{Prob}(\mathbb{Z}_2^\Gamma)$ be the set of all probability measures whose marginal on every hyper-edge gives probability greater than $1-\eta$ to labelings with even parity. Then U^η is a weak* neighborhood of m_X and

$$X_n^\eta = \Omega(\sigma_n, U^\eta),$$

so

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}|\{\mathbf{x} \in \Omega(\sigma_n, U^\eta) : \tfrac{1}{n}|\mathbf{x}| \in [\varepsilon, \delta]\}| < 0,$$

and there are subsets $\Omega_n^\varepsilon \subset \Omega_n^{\mathrm{sofic}}$ with $\mathbb{P}_n(\Omega_n^\varepsilon) \to 1$ and such that for all large enough n, if $\sigma_n \in \Omega_n^\varepsilon$, then $\{\mathbf{x} \in \Omega(\sigma_n, U^\eta) : \frac{1}{n}|\mathbf{x}| \in [\varepsilon, \delta]\} = \varnothing$. Now if $\mathbf{x}, \mathbf{y} \in \Omega(\sigma_n, U^{\eta/2})$, then $\mathbf{x} + \mathbf{y} \in \Omega(\sigma_n, U^\eta)$, so

$$\mathrm{d}^{(V_n)}(\mathbf{x}, \mathbf{y}) = \tfrac{1}{n}|\mathbf{x} + \mathbf{y}| \notin [\varepsilon, \delta].$$

We can then get a single sequence Ω_n' that works for every ε by picking one for each $\varepsilon = \frac{1}{2}, \frac{1}{3}, \ldots$ and then using a diagonal argument. □

10.3. Proof of Theorem C

Theorem C is an immediate consequence of Corollary 5.5(3) and Theorem 6.1(3).

11. Directions for further study

A probability measure-preserving action $\Gamma \curvearrowright (X, \mu)$ is **anti-Pinkser** if it has positive entropy but does not have any nontrivial direct Bernoulli factors. We are being deliberately vague here by not specifying whether 'positive entropy' refers to sofic, Rokhlin or some other notion of entropy.

11.1. Possible anti-Pinsker actions of other groups

1. Are there explicit anti-Pinsker actions of a free group? One candidate is the frozen model associated to independent sets [25].
2. Do all non-amenable groups admit anti-Pinsker actions? Here, it might be necessary to use Rokhlin rather than sofic entropy.
3. Given a positive number h and a non-amenable group Γ, does there exist an uncountable family of pairwise non-measurably conjugate ergodic pmp actions of Γ which are anti-Pinsker, have completely positive sofic entropy and have sofic entropy h (with respect to some fixed sofic approximation)? Starting from the example of the present paper, one place to look might be among its 'typical' compact extensions.

11.2. Open problems for the parity-check subshift

Let (X, m_X, T) be the system in Theorem A.

1. Does there exist a sofic approximation Σ to Γ with respect to which (X, m_X, T) does not have completely positive sofic entropy?

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

2. Is (X,m_X,T) finitely determined? This would mean that if $(\{0,1\}^\Gamma,\mu,T)$ is another system for which μ is close to m_X both in sofic entropy and in the weak* topology, then there is a joining of these two systems under which the identity coordinates agree with high probability. This property characterizes those processes isomorphic to Bernoulli shifts over amenable groups [58], but little is known about it for non-amenable groups.
3. Formally, the family of equations that defines our LDPC shift X can be used to define a system of algebraic origin inside A^Γ for another compact Abelian group A, such as a finite cyclic group or the continuous circle $\mathbb{R}/\mathbb{Z}$. For which such A (and which values of d and k) is the resulting system still anti-Pinsker? If there are such examples with $A = \mathbb{R}/\mathbb{Z}$, do these have infinite sofic entropy?

A. A failure of contiguity

Let V be a vertex set of size n divisible by k, let E be a set of size dn/k, and let $\tilde{\mathbb{P}}_n$ be the measure on k-uniform d-regular factor graphs on (V,E) that is constructed in Section 6.3. In addition, let $\mathbb{P}_n^{\mathrm{unif}}$ be the uniform distribution on all such k-uniform d-regular factor graphs on (V,E) produced without respect to a partition $E = \bigsqcup_{i\in[d]} E_i$.

Proposition A.1. *Let $k > d \geq 3$, and let*

$$U_n := \{H \subset V \times E :\ \exists E' \subseteq E \textit{ with every } v \in V \textit{ adj. to exactly two elements of } E'\}.$$

Then $\tilde{\mathbb{P}}_n(U_n) = 1$ for all n, but $\mathbb{P}_n^{\mathrm{unif}}(U_n) \to 0$ as $n \to \infty$. As a result, the models $\tilde{\mathbb{P}}_n$ and $\mathbb{P}_n^{\mathrm{unif}}$ are not contiguous.

Note that the definition of U_n ensures that the contiguity also fails for the associated multi-hyper-graph models with unlabeled hyper-edges.

Proof. If H arises from $\tilde{\mathbb{P}}_n$, then let $i,j \in [d]$ be distinct and let $E' = E_i \cup E_j$. Then every vertex is adjacent to exactly one check node in each of E_i, E_j, and in particular to exactly two check nodes in E'.

However, the very precise calculations in [51] show that, with high probability according to $\mathbb{P}_n^{\mathrm{unif}}$, the transposed parity-check matrix associated to H has kernel that is either trivial (if d is odd) or one-dimensional (if d is even, in which case, the all 1's vector is in the kernel). In either case, there can be no E' as promised by the event U_n, since it would give an additional nontrivial element of the kernel. □

Acknowledgements. The authors thank Brandon Seward for helpful discussions and an anonymous referee for pointing out several corrections. The authors (ordered alphabetically by last name) were supported by NSF grants DMS-1855694, DMS-2154680 and DMS-1937215, respectively.

Competing interests. The authors have no competing interests to declare.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

References

[1] Achlioptas D and Ricci-Tersenghi F (2006) On the solution-space geometry of random constraint satisfaction problems. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*. New York: ACM, 130–139.

[2] Airey D, Bowen L and Lin Y (2022) A topological dynamical system with two different positive sofic entropies. *Trans. Amer. Math. Soc. Ser. B* **9**(2), 35–98.

[3] Austin T (2016) Additivity properties of sofic entropy and measures on model spaces. *Forum Math. Sigma* 4, 79p.

[4] Austin T (2016) The geometry of model spaces for probability-preserving actions of sofic groups. *Anal. Geom. Metr. Spaces* **4**(1), 160–186.

[5] Austin T (2018) Measure concentration and the weak Pinsker property. *Publ. Math. Inst. Hautes Études Sci.* **128**, 1–119.

[6] Austin T (2020) Multi-variate correlation and mixtures of product measures. *Kybernetika* **56**(3), 459–499.

[7] Austin T and Burton P (2019) Uniform mixing and completely positive sofic entropy. *J. Anal. Math.* **138**(2), 597–612.

[8] Bethe HA (1935) Statistical theory of superlattices. *Proc. Roy. Soc. London Ser. A* **152**, 552–575.

[9] Bowen L (2010) The ergodic theory of free group actions: Entropy and the f-invariant. *Groups Geom. Dyn.* **4**(3), 419–432.

[10] Bowen L (2010) A measure-conjugacy invariant for free group actions. *Ann. of Math. (2)* **171**(2), 1387–1400.

[11] Bowen L (2010) Measure conjugacy invariants for actions of countable sofic groups. *J. Amer. Math. Soc.* **23**(1), 217–245.

[12] Bowen L (2011) Entropy for expansive algebraic actions of residually finite groups. *Ergodic Theory Dynam. Systems* **31**(3), 703–718.

[13] Bowen L (2012) Every countably infinite group is almost Ornstein. In *Dynamical Systems and Group Actions*, vol. 567. Contemp. Math. Providence, RI: Amer. Math. Soc., 67–78.

[14] Bowen L (2018) A brief introduction to sofic entropy theory. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited Lectures.* Hackensack, NJ: World Sci. Publ., 1847–1866.

[15] Bowen L (2019) Finitary random interlacements and the Gaboriau-Lyons problem. *Geom. Funct. Anal.* **29**(3), 659–689.

[16] Bowen L (2020) Examples in the entropy theory of countable group actions. *Ergodic Theory Dynam. Systems* **40**(10), 2593–2680.

[17] Bowen L (2022) Sofic homological invariants and the Weak Pinsker Property. *Amer. J. Math.* **144**(1), 169–226.

[18] Bowen L and Tucker-Drob RD (2022) Superrigidity, measure equivalence, and weak Pinsker entropy. *Groups Geom. Dyn.* **16**(1), 247–286.

[19] Burton PJ and Kechris AS (2020) Weak containment of measure-preserving group actions. *Ergodic Theory Dynam. Systems* **40**(10), 2681–2733.

[20] Cameron PJ and van Lint JH (1991) *Designs, Graphs, Codes and Their Links*, vol. 22. London Mathematical Society Student Texts. Cambridge: Cambridge University Press.

[21] Capraro V and Lupini M (2015) *Introduction to Sofic and Hyperlinear Groups and Connes' Embedding Conjecture*, vol. 2136. Lecture Notes in Mathematics. Cham: Springer. With an appendix by Vladimir Pestov.

[22] Coja-Oghlan A and Efthymiou C (2015) On independent sets in random graphs. *Random Structures Algorithms* **47**(3), 436–486.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

[23] Cover TM and Thomas JA (2006) *Elements of Information Theory*, 2nd edn. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].

[24] Dembo A, Montanari A, Sly A and Sun N (2014) The replica symmetric solution for Potts models on d-regular graphs. *Comm. Math. Phys.* **327**(2), 551–575.

[25] Ding J, Sly A and Sun N (2016) Maximum independent sets on random regular graphs. *Acta Math.* **217**(2), 263–340.

[26] Domb C and Green MS (eds) (1972) *Phase Transitions and Critical Phenomena*, vol. 2. London-New York: Academic Press.

[27] Elek G and Szabó E (2004) Sofic groups and direct finiteness. *J. Algebra* **280**(2), 426–434.

[28] Feller W (1968) *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd edn. New York-London-Sydney: John Wiley & Sons, Inc.

[29] Fu Y and Anderson PW (1986) Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *J. Phys. A* **19**(9), 1605–1620.

[30] Gallager RG (1963) Low-Density Parity-Check Codes. PhD thesis, MIT. http://www.inference.org.uk/mackay/gallager/papers/ldpc.pdf.

[31] Gallager RG (1962) Low-density parity-check codes. *IRE Trans.* **IT-8**, 21–28.

[32] Gamarnik D (2021) The overlap gap property: A topological barrier to optimizing over random structures. *Proc. Natl. Acad. Sci.* **118**(41). Available online at arXiv.org:2109.14409.

[33] Gamarnik D and Sudan M (2017) Limits of local algorithms over sparse random graphs. *Ann. Probab.* **45**(4), 2353–2376.

[34] Greenhill C, Janson S, Kim JH and Wormald NC (2002) Permutation pseudographs and contiguity. *Combin. Probab. Comput.* **11**, 273–298.

[35] Gromov M (1999) Endomorphisms of symbolic algebraic varieties. *J. Eur. Math. Soc. (JEMS)* **1**(2), 109–197.

[36] Hayes B (2016) Fuglede-Kadison determinants and sofic entropy. *Geom. Funct. Anal.* **26**(2), 520–606.

[37] Hayes B (2017) Mixing and spectral gap relative to Pinsker factors for sofic groups. In *Proceedings of the 2014 Maui and 2015 Qinhuangdao Conferences in Honour of Vaughan F. R. Jones' 60th Birthday.* Canberra: Australian National University, Centre for Mathematics and Its Applications, 193–221.

[38] Hayes B (2021) Relative entropy and the Pinsker product formula for sofic groups. *Groups Geom. Dyn.* **15**(2), 413–463.

[39] Kerr D (2013) Sofic measure entropy via finite partitions. *Groups Geom. Dyn.* **7**(3), 617–632.

[40] Kerr D and Li H (2016) *Ergodic Theory.* Springer Monographs in Mathematics. Cham: Springer. Independence and dichotomies.

[41] Kieffer JC (1975) A generalized Shannon-McMillan theorem for the action of an amenable group on a probability space. *Ann. Probab.* **3**(6), 1031–1037.

[42] Kikuchi R (1951) A theory of cooperative phenomena. *Phys. Rev.* **81**, 988–1003.

[43] Kolmogorov AN (1958) A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk SSSR (N.S.)* **119**, 861–864.

[44] Kolmogorov AN (1959) Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk SSSR* **124**, 754–755.

[45] Krzaķała F, Montanari A, Ricci-Tersenghi F, Semerjian G and Zdeborová L (2007) Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA* **104**(25), 10318–10323.

[46] Lyons R (2017) Factors of IID on trees. *Combin. Probab. Comput.* **26**(2), 285–300.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

[47] MacKay DJC (1999) Good error-correcting codes based on very sparse matrices. *IEEE Trans. Inform. Theory* **45**(2), 399–431.

[48] MacKay DJC (2003) *Information Theory, Inference and Learning Algorithms*. New York: Cambridge University Press.

[49] Mézard M and Montanari A (2009) *Information, Physics, and Computation*. Oxford: Oxford University Press.

[50] Mézard M, Parisi G and Virasoro MA (1987) *Spin Glass Theory and Beyond*, vol. 9. World Scientific Lecture Notes in Physics. Teaneck, NJ: World Scientific Publishing Co., Inc.

[51] Miller G and Cohen G (2003) The rate of regular LDPC codes. *IEEE Trans. Inform. Theory* **49**(11), 2989–2992.

[52] Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.* **4**, 337–352.

[53] Ornstein D (1970) Two Bernoulli shifts with infinite entropy are isomorphic. *Adv. Math.* **5**, 339–348.

[54] Ornstein DS (1973) An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Adv. Math.* **10**, 49–62.

[55] Ornstein DS (1973) A K automorphism with no square root and Pinsker's conjecture. *Adv. Math.* **10**, 89–102.

[56] Ornstein DS (1973) A mixing transformation for which Pinsker's conjecture fails. *Adv. Math.* **10**, 103–123.

[57] Ornstein DS and Shields PC (1973) An uncountable family of K-automorphisms. *Adv. Math.* **10**, 63–88.

[58] Ornstein DS and Weiss B (1987) Entropy and isomorphism theorems for actions of amenable groups. *J. Analyse Math.* **48**, 1–141.

[59] Pestov VG (2008) Hyperlinear and sofic groups: A brief guide. *Bull. Symb. Log.* **14**(4), 449–480.

[60] Popa S and Sasyk R (2007) On the cohomology of Bernoulli actions. *Ergodic Theory Dynam. Systems* **27**(1), 241–251.

[61] Richardson T and Urbanke R (2008) *Modern Coding Theory*, 1st edn. Cambridge: Cambridge University Press.

[62] Rudolph DJ and Weiss B (2000) Entropy and mixing for amenable group actions. *Ann. of Math. (2)* **151**(3), 1119–1150.

[63] Seward B (2019) Krieger's finite generator theorem for actions of countable groups I. *Invent. Math.* **215**(1), 265–310.

[64] Seward B (2019) Krieger's finite generator theorem for actions of countable groups II. *J. Mod. Dyn.* **15**, 1–39.

[65] Seward B (2020) Positive entropy actions of countable groups factor onto Bernoulli shifts. *J. Amer. Math. Soc.* **33**(1), 57–101.

[66] Seward B (2022) Bernoulli shifts with bases of equal entropy are isomorphic. *J. Mod. Dyn.* **18**, 345–362.

[67] Sinaĭ JG (1964) On a weak isomorphism of transformations with invariant measure. *Mat. Sb. (N.S.)* **63**(105), 23–42.

[68] Stepin AM (1975) Bernoulli shifts on groups. *Dokl. Akad. Nauk SSSR* **223**(2), 300–302.

[69] Thouvenot JP (1977) On the stability of the weak Pinsker property. *Israel J. Math.* **27**(2), 150–162.

[70] Voiculescu D (1996) The analogues of entropy and of Fisher's information measure in free probability theory. III. The absence of Cartan subalgebras. *Geom. Funct. Anal.* **6**(1), 172–199.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press

[71] Wainwright MJ and Jordan MI (2008) Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1**(1–2), 1–305.

[72] Watanabe S (1960) Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development* **4**(1), 66–82.

[73] Weiss B (2000) Sofic groups and dynamical systems. *Sankhyā Ser. A* **62**(3), 350–359. Ergodic theory and harmonic analysis (Mumbai, 1999).

[74] Weiss B (2003) Actions of amenable groups. In *Topics in Dynamics and Ergodic Theory*, vol. 310. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press, 226–262.

[75] Whittaker ET and Watson GN (2021) *A Course of Modern Analysis—An Introduction to the General Theory of Infinite Processes and of Analytic Functions with an Account of the Principal Transcendental Functions*, 5th edn. Cambridge: Cambridge University Press. Edited by VH Moll, with a foreword by SJ Patterson.

[76] Wormald NC (1999) Models of random regular graphs. In *Surveys in Combinatorics, 1999 (Canterbury)*, vol. 267. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press, 239–298.

[77] Yedidia J (2000) An idiosyncratic journey beyond mean field theory. Technical Report TR-2000-27, Mitsubishi Electric Res. Lab.

[78] Yedidia J, Freeman W and Weiss Y (2001) Bethe free energy, Kikuchi approximations, and belief propagation algorithms. Technical Report TR-2001-16, Mitsubishi Electric Res. Lab.

https://doi.org/10.1017/S147474802510100X Published online by Cambridge University Press