Class Degree and Relative Maximal Entropy

Mahsa Allahbakhshi
Anthony Quas

Abstract

Given a factor code π from a shift of finite type X onto an irreducible sofic shift Y, and a fully supported ergodic measure ν on Y, we give an explicit upper bound on the number of ergodic measures on X which project to ν and have maximal entropy among all measures in the fiber $\pi^{-1}\{\nu\}$. This bound is invariant under conjugacy. We relate this to an important construction for finite-to-one symbolic factor maps.

1 Introduction

It is a well-known result that a 1-dimensional irreducible shift of finite type on a finite alphabet has a unique measure of maximal entropy, the so-called Parry measure [18]. If $\tilde{X}$ is a shift of finite type conjugate to X under a conjugacy $\phi : X \to \tilde{X}$, then the image of the Parry measure on X under ϕ is also the Parry measure on $\tilde{X}$ of the same entropy. In contrast, we consider the relative case in which one is given a factor map $\pi : X \to Y$ from a shift of finite type X to a sofic shift Y, and a measure ν on Y. In this case measures on X in the fiber $\pi^{-1}\{\nu\}$ having maximal entropy in the fiber, so-called measures of relative maximal entropy, are not well understood.

Measures of relative maximal entropy appear frequently in different areas of Mathematics. One of the applications is providing some techniques to compute Hausdorff dimension. This reveals the connections of measures of relative maximal entropy with functions of Markov chains [1, 2, 3, 17], measures that maximize a weighted entropy functional [6, 25], the theory of pressure and equilibrium states [9, 10, 22], relative pressure and relative equilibrium states [14, 15, 27], and compensation functions [2, 27]. Other uses of such measures arise from their application in the mathematics of information transfer [19] and information-compressing channels [17].

The connection between measures of relative maximal entropy and computation of Hausdorff dimension is that rather than calculating the Hausdorff dimension of a set directly, one instead attempts to maximize the Hausdorff dimension of a measure supported on the set. Let $f : M \to M$ be an expanding C^2-diffeomorphism on a connected compact Riemannian manifold M. Extending earlier results of Ruelle [23], Furstenberg [5], Hofbauer [8], and Raith [21], Gatzouras and Peres show that if $f \in C^1$ is conformal then any compact f-invariant set $K \subseteq M$ on which f is expanding, supports an ergodic measure of the same Hausdorff dimension as K (measure of "full Hausdorff dimension" for K) [7].

The question on measures of full Hausdorff dimension where f is nonconformal is not, however, solved in the general case. Using the Ledrappier-Young formula [15], Gatzouras and Peres translated the problem on seeking measures of full Hausdorff dimension to a problem in symbolic dynamics [7]: Let $\pi : X \to Y$ be a factor code from a shift of finite type X to a sofic shift Y. Fix $\alpha > 0$. Is there a unique ergodic measure μ on X which maximizes the weighted entropy functional defined by $h(\mu) + \alpha h(\pi\mu)$?

Shin approaches this problem in [25] using what is known about images of Markov measures under factor codes (functions of Markov chains [1] or metrically sofic measures [17]). To understand when a Markov measure on Y has a Markov measure in its pre-images on X, Boyle and Tuncel introduced the idea of a compensation function [2], which is developed further by Walters [27]: Given a factor code $\pi : X \to Y$ from a shift space X to a shift space Y, $F \in C(X)$ is a compensation function for π if and only if $\sup\{h(\mu) + \int F \, d\mu \colon \mu \in \pi^{-1}\{\nu\}\} = h(\nu)$ for all $\nu \in M(Y)$, where $M(Y)$ stands for the set of invariant measures on Y.

For $F \in C(X)$ the topological pressure $P(F)$ is given by $P(F) = \sup\{h(\mu) + \int F \, d\mu \colon \mu \in M(X)\}$ and a measure $\mu \in M(X)$ is an equilibrium state for F if $P(F) = h(\mu) + \int F \, d\mu$. It is shown by Shin that if there is a saturated compensation function $G \circ \pi$, then for any $\alpha > 0$ the set of all invariant measures μ which maximize the weighted entropy functional is the set of equilibrium states for the function $(\alpha/(\alpha + 1))G \circ \pi$ [25].

Motivated by Shin's result, Yayama in [31] studies the uniqueness of an equilibrium state of a saturated compensation function to discuss the measures of full Hausdorff dimension for a compact invariant set K of an expanding nonconformal map given by a diagonal matrix. She proves the uniqueness of an equilibrium state of some saturated compensation functions when a factor code $\pi : X \to Y$ has a "singleton clump" (some symbol in the alphabet of Y has only one pre-image in the alphabet of X) using a theorem proved by Petersen, Quas, and Shin [20]: when X is a 1-step SFT, and $\pi : X \to Y$ is a 1-block factor map, then the minimum number of pre-images of a symbol b as b runs over the symbols in the alphabet of Y is an upper bound on the number of measures of relative maximal entropy. This number was the best known bound for the number of measures of relative maximal entropy. However, it suffers from not being invariant under conjugacy while the number of ergodic measures of relative maximal entropy is invariant under conjugacy. To avoid this issue one possibility is to take the minimum of this bound over all irreducible shifts of finite type $\tilde{X}$ which are conjugate to X. This obviously improves the original bound but is very hard to compute whether two shifts of finite type are conjugate, see [12, 13].

In this work we find a more satisfactory conjugacy-invariant upper bound defined intrinsically to the shift of finite type. We define an equivalence relation on the set of pre-images of a point $y \in Y$ and study the number, $N(y)$, of equivalence classes of pre-images of y. We show that for a.e. $y \in Y$ we have $N(y) = \min\{N(z) \colon z \in Y\}$ and the minimum number of such equivalence classes is an upper bound on the number of measures of relative maximal entropy.

2 Background

Throughout the paper, the triple (X, Y, π) is called a factor triple when $\pi : X \to Y$ is a factor code (onto and shift-commuting map) from a shift of finite type (SFT) X on a finite alphabet to a sofic shift Y. The alphabet of a shift space X is denoted by $\mathcal{A}(X)$ and the σ-algebra on X generated by measurable rectangles is denoted by $\mathscr{B}_X$. The set of all n-blocks that occur in points of X is denoted by $\mathcal{B}_n(X)$, and the language of X is the collection $\mathscr{L}(X) = \bigcup_{n=0}^{\infty} \mathcal{B}_n(X)$. Let $x \in X$ and $G \subseteq \mathbb{Z}$, then the configuration which occurs in x on G is denoted by x_G. If $G = \{i, \ldots, j\}$ is a connected subset of

$\mathbb{Z}$ we sometimes denote x_G by $x_{[i,j]}$. By recoding if necessary, we may assume that X is a 1-step SFT and π is a 1-block code, so that the triple (X, Y, π) can be described by a directed labeled graph. We say two factor triples (X, Y, π) and $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ are conjugate, and denote it by $(X, Y, \pi) \sim (\tilde{X}, \tilde{Y}, \tilde{\pi})$, if $\tilde{X}$ is conjugate to X under a conjugacy ϕ, $\tilde{Y}$ is conjugate to Y under a conjugacy ψ, and $\pi \circ \phi = \psi \circ \tilde{\pi}$. Let (X, Y, π) be a factor triple where π is a 1-block factor code induced by the map $\pi_{\mathrm{sb}} : \mathcal{A}(X) \to \mathcal{A}(Y)$ (sb stands for sliding block). The map π_{sb} naturally extends to blocks in $\mathcal{B}_n(X)$ for each $n \in \mathbb{N}$. Above every Y-block W of length n there is a set of X-blocks W' of length n which are sent to W by π_{sb}; i.e., $\pi_{\mathrm{sb}}(W') = W$. Given $0 \le i \le n-1$, set

$$d(W, i) = |\{a \in \mathcal{A}(X) \colon \exists W' \text{ with } \pi_{\mathrm{sb}}(W') = W,\, W'_i = a\}|,$$

and let

$$d^*_\pi = \min\{d(W, i) : W \in \mathscr{L}(Y),\, 0 \le i \le |W| - 1\}.$$

A magic block is a block W such that $d(W, i) = d^*_\pi$ for some $0 \le i \le |W| - 1$. Such an index i is called a magic coordinate of W. A factor code π has a magic symbol if there is a magic word of π of length 1.

Proposition 2.1. *Let (X, Y, π) be a factor triple. There is a factor triple $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ conjugate to (X, Y, π) such that $\tilde{X}$ is a 1-step SFT and $\tilde{\pi}$ is a 1-block code with a magic symbol.*

Proof. By recoding, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code. Let W be a magic block of π of length n with a magic coordinate t. Define two n-blocks U and V in X to be equivalent if $\pi(U) = \pi(V)$ and $U_t = V_t$. This obviously defines an equivalence relation on $\mathcal{B}_n(X)$. Denote the equivalence class containing U by $C(U)$. Let $\tilde{X}$ be the new 1-step SFT whose alphabet consists of equivalence classes of blocks in $\mathcal{B}_n(X)$, and the legal transitions between the equivalence classes be defined by saying $C(U)$ can be followed by $C(V)$ if and only if there are $U' \in C(U)$ and $V' \in C(V)$ such that $U'_{i+1} = V'_i$ for $i = 0, \ldots, n-2$. Let $\phi : \tilde{X} \to X$ be the 1-block code induced by the map ϕ_{sb} which takes $C(U)$ to U_t. Then ϕ does map $\tilde{X}$ into X, since whenever $C(U)C(V)$ is a 2-block in $\tilde{X}$, then U_tV_t is a 2-block in X. Moreover, ϕ is a conjugacy with the inverse n-block code induced by the map which takes U to $C(U)$. Therefore, $\tilde{X}$ is a 1-step SFT conjugate to X.

Let $\tilde{Y}$ be the nth higher block presentation of Y and let $\psi : \tilde{Y} \to Y$ be the n-block code induced by the map ψ_{sb} which takes $y_0 \dots y_{n-1}$ to y_t. Define

$$\tilde{\pi} = \psi^{-1} \circ \pi \circ \phi : \tilde{X} \to \tilde{Y}.$$

Observe that $\tilde{\pi}$ is a 1-block code induced by the well-defined map which projects a symbol $C(U)$ in $\mathcal{A}(\tilde{X})$ to $\pi_{\text{sb}}(U)$ (regarding $\pi_{\text{sb}}(U)$ as a symbol of $\tilde{Y}$).

Note that if B is a m-block of $\tilde{Y}$ then it is of the form $E^{(0)}E^{(1)} \dots E^{(m-1)}$ where $E^{(i)}$ is a n-block of Y. Then B corresponds to the $(n+m-1)$-block

$$E_1^{(0)} E_2^{(0)} \dots E_{n-1}^{(0)} E_{n-1}^{(1)} E_{n-1}^{(2)} \dots E_{n-1}^{(m-1)}$$

of Y. For each $0 \leq i \leq m-1$ we have

$$\begin{aligned} d_{\tilde{\pi}}(B, i) &= d_{\tilde{\pi}}(E^{(0)}E^{(1)} \dots E^{(m-1)},\, i) \\ &= |\{C(U) \in \mathcal{A}(\tilde{X})\colon \exists F \in \mathscr{L}(\tilde{X}) \text{ with } \tilde{\pi}_{sb}(F) = B,\, F_i = C(U)\}| \\ &= |\{a \in \mathcal{A}(X)\colon \exists K \in \mathscr{L}(X) \text{ with } \pi_{\text{sb}}(K) = B,\, K_{i+t} = a\}| \\ &= d_{\pi}(E_1^{(0)} E_2^{(0)} \dots E_{n-1}^{(0)} E_{n-1}^{(1)} E_{n-1}^{(2)} \dots E_{n-1}^{(m-1)},\, i+t). \end{aligned}$$

Hence $d^*_{\tilde{\pi}} \geq d^*_{\pi}$. Since $d_{\tilde{\pi}}(W, 0) = d_{\pi}(W, t)$ it follows that W is a magic symbol of $\tilde{\pi}$. □

Proposition 2.2. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y with a magic symbol. Let W be a n-block of Y which begins and ends with magic symbols, and let V be a n-block of X. Then $\pi_{sb}(V) = W$ only if for every $y \in Y$ with $y_{[0,n-1]} = W$ there exists $u \in \pi^{-1}(y)$ with $u_{[0,n-1]} = V$.*

Proof. Let W be an n-block of Y where W_0 and W_{n-1} are magic symbols. Let V be an n-block of X mapping to W under π_{sb}. Let y be a point in Y with $y_{[0,n-1]} = W$. Observe that there is a point $x \in \pi^{-1}(y)$ such that $x_0 = V_0$ since otherwise for some $k \geq 0$, $d(y_{[-k,k]}, 0)$ is less than the number of pre-images of W_0. This contradicts the assumption that W_0 is a magic symbol of π. Similarly there is $z \in \pi^{-1}(y)$ with $z_{n-1} = V_{n-1}$. Form the new point u where

$$u_i = \begin{cases} x_i & \text{if } -\infty < i < 0 \\ V_i & \text{if } 0 \leq i \leq n-1 \\ z_i & \text{if } n-1 < i < \infty. \end{cases}$$

The facts that $x_0 = V_0$, $z_{n-1} = V_{n-1}$ and X is a 1-step SFT guarantee that u is a point of X. Moreover, from the construction we have $\pi(u) = y$ which completes the proof. □

3 Uniform Conditional Distribution

It is a well-known result of Parry [18], generalizing an earlier result of Shannon [24], that in one dimension every irreducible shift of finite type on a finite alphabet has a unique measure of maximal entropy. Burton and Steif [4] give a counterexample to this statement in higher dimensions. However, they show that such measures all have the uniform conditional distribution property stated in Theorem 3.1. Given a finite set $G \subseteq \mathbb{Z}^d$, the boundary of the complement of G is $\partial G^c = \{i \in G^c \colon \exists j \in G \text{ with } \|i - j\| = 1\}$.

Theorem 3.1. *[4, Proposition 1.19] Let μ be a measure of maximal entropy for a SFT in d dimensions. Then the conditional distribution of μ on any finite set $G \subseteq \mathbb{Z}^d$ given the configuration on G^c is μ-a.s. uniform over all configurations on G which extend the configuration on ∂G^c.*

Given a factor triple (X, Y, π) and an ergodic measure ν on Y, there can exist more than one ergodic measure of relative maximal entropy over ν; i.e., there can be more than one ergodic measure on X which projects to ν and has maximal entropy among all measures in the fiber $\pi^{-1}\{\nu\}$, see [20, Example 3.3]. We use Lemma 3.2 and follow techniques developed by Burton and Steif in the proof of Theorem 3.1 to show the uniform conditional distribution property for measures of relative maximal entropy in Theorem 3.3.

Lemma 3.2. *[26, Theorem 4.7] Let $(X, \mathscr{B}, \mu)$ be a probability space. Let $\mathcal{A}$ be a finite sub-algebra of $\mathscr{B}$ and let $(\mathscr{F}_n)_{n=1}^{\infty}$ be an increasing sequence of sub-σ-algebras of $\mathscr{B}$ with $\bigvee_{n=1}^{\infty} \mathscr{F}_n = \mathscr{F}$. Then $H(\mathcal{A}|\mathscr{F}_n) \to H(\mathcal{A}|\mathscr{F})$.*

Theorem 3.3. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y, ν an invariant measure on Y, and μ an invariant measure of relative maximal entropy over ν. Then the conditional distribution of μ on any finite set $G \subseteq \mathbb{Z}$ given the configuration on G^c is μ-a.s. uniform over all configurations on G which extend the configuration on ∂G^c and map to the same configuration in Y under the factor code π.*

Proof. Let $G \subseteq \mathbb{Z}$ be finite. Let Δ be a configuration of Y on G. Pick a configuration η of X on ∂G^c such that $\mu(\eta \cap \pi^{-1}(\Delta)) > 0$. Starting from μ, we define a measure $\tilde{\gamma}$ on X by uniformizing over pre-images of Δ that have η on the boundary. We then show that if μ does not have the required uniform conditional distribution property then $\tilde{\gamma}$ has greater entropy than μ, but still is an element of the fiber $\pi^{-1}\{\nu\}$. This is a contradiction and will thus establish the required uniform conditional distribution property of μ.

Let $D = \{\alpha_1, \ldots, \alpha_L\}$ be the set of all configurations on G which extend η and map to Δ under the factor code π. Let $R = \{-m, \ldots, m-1\}$ be large enough so that $G \cup \partial G^c \subseteq R$ and $(R_n)_{n\in\mathbb{Z}}$ be the partition of $\mathbb{Z}$ by translates of R ($R_0 = R$), i.e. $R_n = \{(2n-1)m, \ldots, (2n+1)m-1\}$. Let G_n and ∂G_n^c be the corresponding translates of G and ∂G^c in R_n. If $S \subseteq \mathbb{Z}$ let $\mathcal{P}(X_S)$ be the partition of X generated by the configurations of X on S, and $\sigma(X_S)$ be the σ-algebra generated by $\mathcal{P}(X_S)$. When $S = \{a, \ldots, b\}$ is a connected subset of $\mathbb{Z}$ we sometimes denote X_S by X_a^b. Considering R above, $\sigma(X_{-m}^{m-1})$ is the finite σ-algebra generated by the partition

$$\mathcal{P}(X_{-m}^{m-1}) = \{{}_{-m}[x_{-m}x_{-m+1}\ldots x_{m-1}]_{m-1} : x_{-m}x_{-m+1}\ldots x_{m-1} \in \mathscr{L}(X)\},$$

and $\sigma(X_{-\infty}^{-m-1}) = \sigma(X_{-3m}^{-m-1}) \vee \sigma(X_{-5m}^{-3m-1}) \vee \ldots$.

Let γ be the measure obtained from μ and η as follows. Define $\Phi : X \times D^{\mathbb{Z}} \to X$ by

$$\Phi(x,\zeta)_{G_n} = \begin{cases} \zeta_n & \text{if } x_{\partial G_n^c} = \eta \text{ and } x_{G_n} \in D \\ x_{G_n} & \text{otherwise,} \end{cases}$$

and $\Phi(x,\zeta)_{G_n^c \cap R_n} = x_{G_n^c \cap R_n}$ for each $n \in \mathbb{Z}$. Since $\zeta_n \in D$ and each element of D extends η, the assumption that X is a 1-step SFT implies $\Phi(x,\zeta) \in X$. For each $\zeta \in D^{\mathbb{Z}}$ we have $\pi(x) = \pi(\Phi(x,\zeta))$ since $\Phi(x,\zeta)$ and x are the same except having alternative α_i's in the same positions ($\alpha_i \in \pi_{sb}^{-1}(\Delta)$). Let $C \in \mathscr{B}_X$. Define $\gamma(C) = (\mu \times \lambda)\Phi^{-1}(C)$ where λ is the Bernoulli $(1/L, \ldots, 1/L)$ measure on $D^{\mathbb{Z}}$. The measure γ is not necessarily invariant under T; however, for each $C \in \mathscr{B}_X$ we have $\gamma(C) = \gamma(T^{-2m}(C))$. So the new measure $\tilde{\gamma}$ on X defined by $\tilde{\gamma}(C) = \frac{1}{2m}(\gamma(C) + \cdots + \gamma(T^{-2m+1}(C)))$ is T-invariant. Since for each $E \in \mathscr{B}_Y$ we have $\gamma(\pi^{-1}(E)) = (\mu \times \lambda)(\pi^{-1}(E) \times D^{\mathbb{Z}}) = \nu(E)$, we deduce that both measures γ and $\tilde{\gamma}$ are in the fiber $\pi^{-1}\{\nu\}$.

Define an equivalence relation on X as follows; suppose $x, x' \in X$, say $x \sim_0 x'$ if either $x_R = x'_R$ or else $x_{R\cap G^c} = x'_{R\cap G^c}$, $x_{\partial G^c} = x'_{\partial G^c} = \eta$, and

$x_G, x'_G \in D$. Denote the equivalence class containing x by $C_0(x)$. Such equivalence classes form a sub-partition of $\mathcal{P}(X_{-m}^{m-1})$. Let $\mathcal{A}$ be the σ-algebra generated by these equivalence classes. We show $H_{\tilde{\gamma}}(T) \geq H_\mu(T)$ as follows, using a lemma which appears below;

$$
\begin{aligned}
H_{\tilde{\gamma}}(T) &= \frac{1}{2m} H_{\tilde{\gamma}}\left(T^{2m}\right) && (3.1)\\
&= \frac{1}{2m} H_{\gamma}\left(T^{2m}\right)\\
&= \frac{1}{2m} H_{\gamma}\left(\sigma(X_{-m}^{m-1})|\sigma(X_{-\infty}^{-m-1})\right)\\
&= \frac{1}{2m} H_{\gamma}\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) + \frac{1}{2m} H_{\gamma}\left(\sigma(X_{-m}^{m-1})|\sigma(X_{-\infty}^{-m-1}) \vee \mathcal{A}\right)\\
&= \frac{1}{2m} H_{\gamma}\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) + \frac{1}{2m} H_{\gamma}\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) && \text{by Lemma 3.4}(a)\\
&\geq \frac{1}{2m} H_{\gamma}\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) + \frac{1}{2m} H_{\mu}\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) && \text{by Lemma 3.4}(b)\\
&\geq \frac{1}{2m} H_{\mu}\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) + \frac{1}{2m} H_{\mu}\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) && \text{by Lemma 3.4}(c)\\
&\geq \frac{1}{2m} H_{\mu}\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) + \frac{1}{2m} H_{\mu}\left(\sigma(X_{-m}^{m-1})|\sigma(X_{-\infty}^{-m-1}) \vee \mathcal{A}\right)\\
&= \frac{1}{2m} H_{\mu}\left(\sigma(X_{-m}^{m-1})|\sigma(X_{-\infty}^{-m-1})\right)\\
&= \frac{1}{2m} H_{\mu}\left(T^{2m}\right)\\
&= H_\mu(T).
\end{aligned}
$$

Lemma 3.4. *Reusing previous notations, we have*

(a) $H_\gamma\left(\sigma(X_{-m}^{m-1})|\sigma(X_{-\infty}^{-m-1}) \vee \mathcal{A}\right) = H_\gamma\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right)$.

(b) $H_\gamma\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) \geq H_\mu\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right)$. *Equality occurs if and only if*

$$\frac{\mu(\alpha \cap \bar{A})}{\mu(\bar{A})} = 1/L,$$

where $\alpha \in D$ *and* $\bar{A} \in \mathcal{P}(\mathcal{A})$ *is an equivalence class in which for each* $x \in \bar{A}$ *we have* $x_{\partial G^c} = \eta$ *and* $x_G \in D$.

(c) $H_\gamma\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) \geq H_\mu\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right)$.

Proof. By definition, for $\rho \in \{\mu, \gamma\}$ we have

$$H_\rho\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) = -\sum_{i,k} \rho(O_i \cap A_k) \log \frac{\rho(O_i \cap A_k)}{\rho(A_k)}$$

where $O_i \in \mathcal{P}\left(X_{-m}^{m-1}\right)$ and $A_k \in \mathcal{P}(\mathcal{A})$ (If $\rho(O_i \cap A_k) = 0$ define $\rho(O_i \cap A_k) \log \frac{\rho(O_i \cap A_k)}{\rho(A_k)} = 0$). Let $x \in A_k$. If $x_{\partial G^c} \neq \eta$ or $x_G \notin D$ then for every $O_i \in \mathcal{P}\left(X_{-m}^{m-1}\right)$ we have either $O_i \cap A_k = \emptyset$ or $O_i \cap A_k = A_k$ which both imply $\rho(O_i \cap A_k) \log \frac{\rho(O_i \cap A_k)}{\rho(A_k)} = 0$. Let $\{\bar{A}_1, \ldots, \bar{A}_M\} \subseteq \mathcal{P}(\mathcal{A})$ be the set of equivalence classes in which for $x \in \bar{A}_k$ we have $x_{\partial G^c} = \eta$ and $x_G \in D$. Then

$$H_\rho(\sigma(X_{-m}^{m-1})|\mathcal{A}) = -\sum_{i,k} \rho(O_i \cap \bar{A}_k) \log \frac{\rho(O_i \cap \bar{A}_k)}{\rho(\bar{A}_k)}.$$

There are exactly L disjoint sets $O_i \in \mathcal{P}\left(X_{-m}^{m-1}\right)$ defined by blocks which agree everywhere except on G, and form a partition of $\bar{A}_k$. Let these sets be denoted by $O_{k,1}, \ldots, O_{k,L}$ where $O_{k,i} = \alpha_i \cap \bar{A}_k$ for each $1 \leq i \leq L$. It follows that

$$\begin{aligned} H_\rho(\sigma(X_{-m}^{m-1})|\mathcal{A}) &= -\sum_{k,i} \rho(O_{k,i} \cap \bar{A}_k) \log \frac{\rho(O_{k,i} \cap \bar{A}_k)}{\rho(\bar{A}_k)} \\ &= -\sum_{k,i} \rho(\alpha_i \cap \bar{A}_k) \log \frac{\rho(\alpha_i \cap \bar{A}_k)}{\rho(\bar{A}_k)}. \end{aligned} \tag{3.2}$$

Let $n \geq 1$. By definition of γ, for each $\alpha \in D$, $P \in \mathcal{P}\left(X_{-(2n+1)m}^{-m-1}\right)$, and $1 \leq k \leq M$ we have

$$\frac{\gamma(\alpha \cap P \cap \bar{A}_k)}{\gamma(P \cap \bar{A}_k)} = \frac{\gamma(\alpha \cap \bar{A}_k)}{\gamma(\bar{A}_k)} = \frac{1}{L}.$$

It follows that

$$\begin{aligned} H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\sigma\left(X_{-(2n+1)m}^{-m-1}\right) \vee \mathcal{A}\right) &= -\sum_{i,j,k} \gamma(\alpha_i \cap P_j \cap \bar{A}_k) \log \frac{\gamma(\alpha_i \cap P_j \cap \bar{A}_k)}{\gamma(P_j \cap \bar{A}_k)} \\ &= \log L \sum_{i,k} \gamma(\alpha_i \cap \bar{A}_k) \\ &= H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) \end{aligned}$$

(use the same argument we had before Equation (3.2) to get the first equality above). Then (a) follows from Lemma 3.2 and the fact that $\left(\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right)_{n=1}^{\infty}$ is an increasing sequence of σ-algebras with $\bigvee_{n=1}^{\infty}\sigma\left(X_{-(2n+1)m}^{-m-1}\right)=\sigma\left(X_{-\infty}^{-m-1}\right)$.

To show (b) note that

$$\begin{aligned} H_\mu\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) &= -\sum_{i,k}\mu(\alpha_i\cap\bar{A}_k)\log\frac{\mu(\alpha_i\cap\bar{A}_k)}{\mu(\bar{A}_k)} \\ &= -\sum_k \mu(\bar{A}_k)\sum_{i=1}^{L}\frac{\mu(\alpha_i\cap\bar{A}_k)}{\mu(\bar{A}_k)}\log\frac{\mu(\alpha_i\cap\bar{A}_k)}{\mu(\bar{A}_k)} \\ &= \sum_k \mu(\bar{A}_k)\sum_{i=1}^{L}\psi\left(\frac{\mu(\alpha_i\cap\bar{A}_k)}{\mu(\bar{A}_k)}\right) \end{aligned}$$

where $\psi(x)=-x\log x$ ($\psi(0)=0$) on the interval $[0,1]$. Since $\psi(x)$ is a strictly concave function it follows that $H_\mu(\sigma(X_{-m}^{m-1})|\mathcal{A})$ attains its maximum if and only if for each $1\le k\le M$ and $\alpha\in D$ we have

$$\frac{\mu(\alpha\cap\bar{A}_k)}{\mu(\bar{A}_k)}=1/L.$$

Therefore

$$\begin{aligned} H_\mu\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) &\le \log L\sum_{k,i}\mu(\alpha_i\cap\bar{A}_k) \\ &= \log L\sum_{k,i}\gamma(\alpha_i\cap\bar{A}_k) \\ &= H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right), \end{aligned}$$

with equality if and only if $\frac{\mu(\alpha\cap\bar{A}_k)}{\mu(\bar{A}_k)}=1/L$ for each $\alpha\in D$ and $1\le k\le M$.

We prove (c) by showing that

$$H_\gamma(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right))=H_\mu(\mathcal{A}|\mathcal{A}_{-\infty}^{-m-1}) \tag{3.3}$$

where $\mathcal{A}_{-\infty}^{-m-1}=\bigvee_{n=1}^{\infty}T^{-2nm}\mathcal{A}$ (the choice of notation is intended to remind the reader that $\mathcal{A}_{-\infty}^{-m-1}$ is a sub-σ-algebra of $\sigma(X_{-\infty}^{-m-1})$). The fact that $\mathcal{A}_{-\infty}^{-m-1}\subseteq\sigma(X_{-\infty}^{-m-1})$ implies

$$H_\mu(\mathcal{A}|\mathcal{A}_{-\infty}^{-m-1})\ge H_\mu(\mathcal{A}|\sigma(X_{-\infty}^{-m-1}))$$

which will complete the proof.

Define an equivalence relation $\sim_n$ on X as follows; say $x \sim_n x'$ provided that for all $-n \leq i \leq -1$ we either have $x_{R_i} = x'_{R_i}$ or else $x_{R_i \cap G_i^c} = x'_{R_i \cap G_i^c}$, $x_{\partial G_i^c} = x'_{\partial G_i^c} = \eta$, and $x_{G_i}, x'_{G_i} \in D$. Denote the equivalence class containing x by $C_n(x)$. Such equivalence classes form a sub-partition of $\mathcal{P}\left(X_{-(2n+1)m}^{-m-1}\right)$ and these equivalence classes generate the σ-algebra $\mathcal{A}_{-(2n+1)m}^{-m-1} = T^{-2m}\mathcal{A} \vee \cdots \vee T^{-2nm}\mathcal{A}$. Let $x \in X$, then $x \in E$ for some $E \in \mathcal{P}\left(\mathcal{A}_{-(2n+1)m}^{-m-1}\right)$. Set $K_x = \{-n \leq k \leq -1\colon x_{\partial G_k^c} = \eta,\ x_{G_k} \in D\}$. There are exactly $L^{|K_x|}$ disjoint blocks $P_1, \ldots, P_{L^{|K_x|}}$ of the partition $\mathcal{P}\left(X_{-(2n+1)m}^{-m-1}\right)$ which are the same except having alternative α_i's in the same positions and they form a partition of E. Let $A \in \mathcal{A}$. By definition of γ for each $1 \leq j \leq L^{|K_x|}$ we have

$$\frac{\gamma(A \cap P_j)}{\gamma(P_j)} = \frac{L^{-|K_x|}\gamma(A \cap E)}{L^{-|K_x|}\gamma(E)} = \frac{\gamma(A \cap E)}{\gamma(E)} = \frac{\mu(A \cap E)}{\mu(E)}.$$

It follows that

$$\mathbb{E}_\gamma\left(1_A|\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right) = \mathbb{E}_\mu\left(1_A|\mathcal{A}_{-(2n+1)m}^{-m-1}\right). \tag{3.4}$$

Let $x \in X$, $F \in \mathcal{A}_{-\infty}^{-m-1}$, and $\zeta \in D^{\mathbb{Z}}$. Then the two sided implication $\Phi(x,\zeta) \in F$ if and only if $x \in F$ implies that $1_F \circ \Phi(x,\zeta) = 1_F(x)$. Let g be a bounded $A_{-\infty}^{-m-1}$-measurable function. Since there is a sequence of $A_{-\infty}^{-m-1}$-measurable simple functions converging uniformly to g we deduce that $g \circ \Phi(x,\zeta) = g(x)$ and moreover,

$$\int_F g\ d\gamma = \int_F g\ d(\mu \times \lambda)(\Phi^{-1}) = \int_{F \times D^{\mathbb{Z}}} g \circ \Phi\ d(\mu \times \lambda) = \int_F g\ d\mu. \tag{3.5}$$

In particular if $g = \mathbb{E}_\mu\left(1_A|\mathcal{A}_{-\infty}^{-m-1}\right)$ for some $A \in \mathcal{P}(\mathcal{A})$ then we have

$$\begin{aligned}
H_\gamma\left(\mathcal{A}|\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right) &= \int_X \sum_{A \in \mathcal{P}(\mathcal{A})} \psi\left(\mathbb{E}_\gamma\left(1_A|\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right)\right)\ d\gamma \\
&= \int_X \sum_{A \in \mathcal{P}(\mathcal{A})} \psi\left(\mathbb{E}_\mu\left(1_A|\mathcal{A}_{-(2n+1)m}^{-m-1}\right)\right)\ d\gamma \qquad \text{by (3.4)} \\
&= \int_X \sum_{A \in \mathcal{P}(\mathcal{A})} \psi\left(\mathbb{E}_\mu\left(1_A|\mathcal{A}_{-(2n+1)m}^{-m-1}\right)\right)\ d\mu \qquad \text{by (3.5)} \\
&= H_\mu\left(\mathcal{A}|\mathcal{A}_{-(2n+1)m}^{-m-1}\right).
\end{aligned}$$

Then Lemma 3.4(c) follows from Lemma 3.2 and the fact that $\left(\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right)_{n=1}^{\infty}$ and $\left(\mathcal{A}_{-(2n+1)m}^{-m-1}\right)_{n=1}^{\infty}$ are both increasing sequences with $\bigvee_{n=1}^{\infty}\sigma\left(X_{-(2n+1)m}^{-m-1}\right) = \sigma\left(X_{-\infty}^{-m-1}\right)$ and $\bigvee_{n=1}^{\infty}\mathcal{A}_{-(2n+1)m}^{-m-1} = \mathcal{A}_{-\infty}^{-m-1}$. $\square$

Proof of Theorem 3.3 (continued). Since μ is an invariant measure of relative maximal entropy it follows that all of the inequalities in Equation (3.1) are forced to be equalities. In particular, we have $H_{\mu}\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) = H_{\gamma}\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right)$. Then Lemma 3.4(b) implies that

$$\frac{\mu\left(\alpha\cap\bar{A}\right)}{\mu(\bar{A})} = 1/L, \tag{3.6}$$

for each $\alpha \in D$ and $\bar{A} \in \mathcal{P}(\mathcal{A})$ in which for each $x \in \bar{A}$ we have $x_{\partial G^c} = \eta$ and $x_G \in D$. Note that given a finite set $G \subseteq \mathbb{Z}$, both configurations $\Delta \in \mathscr{B}_Y$ on G and $\eta \in \mathscr{B}_X$ on ∂G^c with $\mu(\eta \cap \pi^{-1}(\Delta)) > 0$ are chosen arbitrarily. By choosing different configurations and noting that $\bar{A} \in \mathcal{P}\left(\pi^{-1}(\sigma(Y_G))\bigvee\sigma(X_{R\cap G^c})\right) = \mathcal{P}\left(\pi^{-1}(\sigma(Y_R))\bigvee\sigma(X_{R\cap G^c})\right)$, Equation (3.6) implies that for any configuration α of X occurring at G we have

$$\mathbb{E}\left(1_{\alpha}|\pi^{-1}(\sigma(Y_R))\vee\sigma(X_{R\cap G^c})\right)(x) = \begin{cases}\frac{1}{L_{(x,G)}} & \text{if } \alpha \text{ extends } x_{\partial G^c},\ x_G \in \pi_{sb}^{-1}(\pi_{\mathrm{sb}}(\alpha))\\ 0 & \text{otherwise.}\end{cases}$$

where $L_{(x,G)}$ is the number of configurations of X occurring at G which extend $x_{\partial G^c}$ and project to $\pi_{\mathrm{sb}}(x_G)$.

Now for $t \in \mathbb{N}$ let $R^{(t)} = \{-(m+t),\dots,m+t-1\}$, and $(R_n^{(t)})_{n\in\mathbb{Z}}$ be the partition of $\mathbb{Z}$ by translates of $R^{(t)}$ where $R_0^{(t)} = R^{(t)}$, i.e. $R_n^{(t)} = \{(2n-1)(m+t),\dots,(2n+1)(m+t)-1\}$. Let $R^{(0)} = R$. Since $\left(\sigma\left(Y_{R^{(t)}}\right)\right)_{t=0}^{\infty}$ and $\left(\sigma\left(X_{R^{(t)}\cap G^c}\right)\right)_{t=0}^{\infty}$ are increasing sequences with $\bigvee_{t=0}^{\infty}\sigma(Y_{R^{(t)}}) = \mathscr{B}_Y$ and $\bigvee_{t=0}^{\infty}\sigma(X_{R^{(t)}\cap G^c}) = \sigma(X_{G^c})$ it follows from Levy's Upward Theorem that

$$\mathbb{E}\left(1_{\alpha}|\pi^{-1}(\mathscr{B}_Y)\vee\sigma(X_{G^c})\right)(x) = \begin{cases}\frac{1}{L_{(x,G)}} & \text{if } \alpha \text{ extends } x_{\partial G^c},\ x_G \in \pi_{sb}^{-1}(\pi_{\mathrm{sb}}(\alpha))\\ 0 & \text{otherwise.}\end{cases}$$

$\square$

4 Degree of an infinite-to-one factor code

When π is a finite-to-one factor code from a SFT X to a sofic shift Y, there is a uniform upper bound on the number of pre-images of points in Y, see [16, Theorem 8.1.16]. The minimal number of π-pre-images of points in Y is called the degree of the code and denoted by d_π. Theorem 4.1 states that when Y is irreducible then there are exactly d_π points in the pre-image of a typical point of Y.

Theorem 4.1. *[16, Theorem 9.1.11] Let $\pi : X \to Y$ be a finite-to-one factor code from a SFT X to an irreducible sofic shift Y. Then every doubly transitive point of Y has exactly d_π pre-images. If $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ is a factor triple conjugate to (X, Y, π) then $d_\pi = d_{\tilde{\pi}}$.*

Definition 4.2. *Blocks $E^{(1)}, \ldots, E^{(k)}$ of length n are mutually separated if, for each $0 \le i \le n-1$, $E_i^{(1)}, \ldots, E_i^{(k)}$ are all distinct.*

Proposition 4.3. *[16, Proposition 9.1.9] Let $\pi : X \to Y$ be a finite-to-one 1-block factor code from a 1-step SFT to an irreducible sofic shift Y with a magic symbol w, then we have $d_\pi = |\pi_{sb}^{-1}(w)|$. Moreover, if V is a block of Y which begins and ends with magic symbols then V has exactly d_π pre-images. These pre-images are mutually separated.*

We will find a quantity analogous to the degree when π is an infinite-to-one factor code. This will be done by developing the following equivalence relation on X. Figure 1 illustrates Definition 4.4.

Definition 4.4. *Suppose (X, Y, π) is a factor triple and x, $x' \in X$. We say there is a transition from x to x' and denote it by $x \to x'$ if, for each $n \in \mathbb{Z}$, there exists $v \in X$ so that*

1. *$\pi(v) = \pi(x) = \pi(x')$, and*
2. *$v_{-\infty}^{n} = x_{-\infty}^{n}, \quad v_i^{\infty} = x_i'^{\infty}$ for some $i \ge n$.*

Write $x \nrightarrow x'$ if the above conditions do not hold. We write $x \sim x'$, and say x and x' are in the same transition equivalence class if $x \to x'$ and $x' \to x$. It is left to the reader to check that $\sim$ is an equivalence relation. Denote the set of transition equivalence classes in X over $y \in Y$ by $\mathscr{C}_\pi(y)$. Sometimes we denote $\mathscr{C}_\pi(y)$ by only $\mathscr{C}(y)$ when there is no ambiguity in understanding

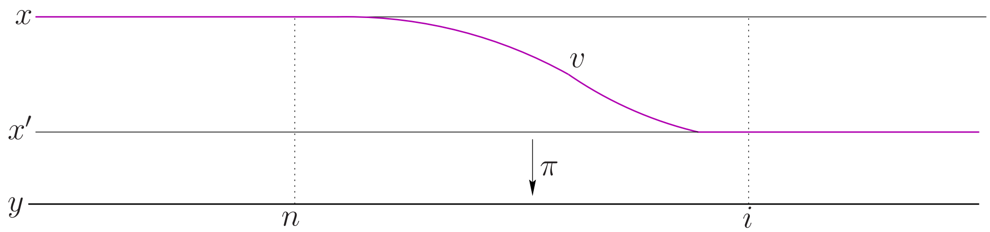

Figure 1: Transition from point x to x'

π. We say $[x] \to [x']$ if $x \to x'$ (well-defined); use the notation $[x] \not\to [x']$ otherwise.

The following fact is derived from Definition 4.4 immediately.

Fact 4.5. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y, and $y \in Y$. Let $x, x' \in \pi^{-1}(y)$ with $x_{a_i} = x'_{a_i}$ where $(a_i)_{i \in \mathbb{N}}$ is a strictly increasing sequence in $\mathbb{Z}$. Then we have $x \sim x'$.*

We mention that Fact 4.5 gives an obvious case when two points lie in the same equivalence class. Example 4.6 illustrates a more complicated case when two points are equivalent without having a common symbol at the same time.

Example 4.6. *Let $X_{\mathscr{F}} \subseteq \{a, b\}^{\mathbb{Z}}$ be a SFT with $\mathscr{F} = \{bb\}$. Let $\pi : X_{\mathscr{F}} \to \{0\}^{\mathbb{Z}}$, as shown by the labeled graph in Figure 2.*

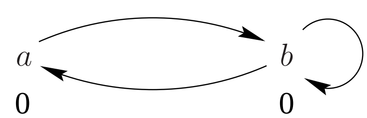

Figure 2: Graph for Example 4.6

Observe that points $x = \dots ab\overset{*}{a}baba \dots$ and $x' = \dots bab\overset{*}{a}bab \dots$ (where * indicates the 0-th position) have no common symbol at the same time but are equivalent. To see there is a transition from x to x', let $n \in \mathbb{Z}$ and consider the point v where

$$v_i = \begin{cases} x_i & i \le n \\ b & i = n+1 \\ x'_i & i > n+1. \end{cases}$$

It is clear that v holds the conditions in Definition 4.4. That $x' \to x$ is shown similarly.

Theorem 4.7 shows that every point of Y has a finite number of transition equivalence classes. Later in Theorem 4.20 we show that this number is constant over every right transitive point of Y.

Theorem 4.7. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Then $|\mathscr{C}(y)| < \infty$ for each $y \in Y$.*

Proof. Let $y \in Y$. Since there are only finitely many symbols in $\mathcal{A}(Y)$ there is a symbol $w \in \mathcal{A}(Y)$ and a strictly increasing sequence of integers $(a_j)_{j\in\mathbb{N}}$ such that $y_{a_j} = w$ for each $j \in \mathbb{N}$. We show that $|\mathscr{C}(y)| \leq |\pi_{sb}^{-1}(w)|$.

Let $|\pi_{sb}^{-1}(w)| = d$ and suppose $\pi^{-1}(y)$ contains $d+1$ distinct equivalence classes $C_1, \ldots, C_{d+1}$. Form a set $A = \{x_1, \ldots, x_{d+1}\}$ where x_i is an arbitrary point of C_i. Since $\pi_{sb}^{-1}(w)$ contains exactly d symbols, it follows that for each $j \in \mathbb{N}$ there are at least two points in A with the same a_jth coordinate. The Pigeonhole Principle implies that there is a subsequence $(b_k)_{k\in\mathbb{N}}$ of $(a_j)_{j\in\mathbb{N}}$ and at least two points x and x' in A with $x_{b_k} = x'_{b_k}$ for each $k \in \mathbb{N}$. Fact 4.5 implies that $x \sim x'$. This contradicts the assumption that x and x' are in different equivalence classes. □

Corollary 4.8 is derived directly from the proof of Theorem 4.7.

Corollary 4.8. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let ν be a measure on Y and w be a symbol of Y with $\nu([w]) > 0$. Let y be a right transitive point of Y. Then $|\mathscr{C}(y)| \leq |\pi_{sb}^{-1}(w)|$.*

Definition 4.9. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let $y \in Y$ and $n \in \mathbb{Z}$. The* **relative follower set** *of $i \in \pi_{sb}^{-1}(y_n)$, denoted by $\mathscr{F}(i, y_{n+1}^{\infty})$, is the set of equivalence classes in $\mathscr{C}(y)$ containing a point whose nth coordinate is i; i.e.,*

$$\mathscr{F}(i, y_{n+1}^{\infty}) = \{C \in \mathscr{C}(y)\colon \text{there is } x \in C \text{ with } x_n = i\}.$$

Say $i \in \mathcal{A}(X)$ **belongs** *to a transition equivalence class C* **at time** *n, and denote it by $i \in S_n(C)$, if $\mathscr{F}(i, y_{n+1}^{\infty}) = \{D \in \mathscr{C}(y) : C \to D\}$. Say i is* **transient at time** *n if there is no $C \in \mathscr{C}(y)$ for which $i \in S_n(C)$.*

Example 4.10 clarifies the definition above.

Example 4.10. *Consider the directed labeled graph in Figure 3 which presents a 1-block factor code* $\pi : X \to Y$ *where* $X \subseteq \{a, b, c, d, e, f, g\}^{\mathbb{Z}}$ *and* $Y \subseteq \{0, 1\}^{\mathbb{Z}}$. *Let* y *be the point* $\cdots 01\overset{*}{0}1010 \cdots$ *in* Y. *By Definition 4.4, there are 3 distinct classes in* $\mathscr{C}(y)$ *as follows:*

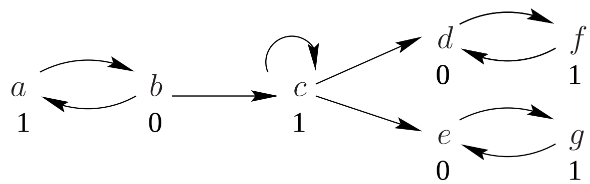

Figure 3: Graph for Example 4.10

1. *Class* $C_1 = [x_1] = \{x_1\}$ *where* $x_1 = \cdots ba\overset{*}{b}abab \cdots$.
2. *Class* $C_2 = [x_2]$ *where* $x_2 \in X$ *is a point in* $\pi^{-1}(y)$ *which does not contain symbols* e *or* g *but contains only* d*'s and* f*'s from some time onwards; for example:* $x_2 = \cdots ba\overset{*}{b}cdfd \cdots$.
3. *Class* $C_3 = [x_3]$ *where* $x_3 \in X$ *is a point in* $\pi^{-1}(y)$ *which does not contain symbols* d *or* f *but contains only* e*'s and* g*'s from some time onwards; for example:* $x_3 = \cdots eg\overset{*}{e}gege \cdots$.

Clearly $C_1 \to C_2$ *and* $C_1 \to C_3$, *but not vice versa.* $C_2 \nrightarrow C_3$, $C_3 \nrightarrow C_2$, *and symbol* c *is transient at any time.*

Definition 4.11. *Let* (X, Y, π) *be a factor triple. The minimal number of transition equivalence classes over points of* Y *is called the* **class degree** *of* π *and denoted by* c_π.

Theorem 4.12 shows that conjugate factor triples have the same class degree.

Theorem 4.12. *Let* (X, Y, π) *and* $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ *be conjugate factor triples. Then we have* $c_\pi = c_{\tilde{\pi}}$.

Proof. By Proposition 2.1, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code. Let $\phi : X \to \tilde{X}$ be a conjugacy induced by ϕ_{sb} so that $\phi(x)_i = \phi_{sb}(x_{[i-m,i+t]})$ for some $m, t \in \mathbb{N}$, and $\psi : Y \to \tilde{Y}$ be a conjugacy from Y to $\tilde{Y}$. Let $y \in Y$ and $u, v \in \pi^{-1}(y)$. Since $\tilde{\pi} \circ \phi = \psi \circ \pi$ then $\phi(u), \phi(v) \in \tilde{\pi}^{-1}(\psi(y))$. We show that there is a transition from u to v if and only if there is a transition from $\phi(u)$ to $\phi(v)$. Let $n \in \mathbb{Z}$, then $u \to v$ implies that there is $x \in \pi^{-1}(y)$ where

$$x_j = \begin{cases} u_j & \text{if } -\infty < j \leq n + t \\ v_j & \text{if } i \leq j < \infty \end{cases}$$

for some $i \geq t + n$. Clearly $\phi(x) \in \tilde{\pi}^{-1}(\psi(y))$. Moreover, we have

$$\phi(x)_j = \begin{cases} \phi(u)_j & \text{if } -\infty < j < n \\ \phi(v)_j & \text{if } i + m \leq j < \infty. \end{cases}$$

Having an arbitrary $n \in \mathbb{Z}$ implies that $\phi(u) \to \phi(v)$, and since ϕ is invertible the other implication follows similarly. This shows that C is a transition equivalence class in $\mathscr{C}_\pi(y)$ if and only if $\phi(C)$ is a transition equivalence class in $\mathscr{C}_{\tilde{\pi}}(\psi(y))$. It also shows that for every $C_1, C_2 \in \mathscr{C}_\pi(y)$ we have $C_1 \to C_2$ if and only if $\phi(C_1) \to \phi(C_2)$ which implies the equality of c_π and $c_{\tilde{\pi}}$. □

Theorem 4.13 shows that in the case of a finite-to-one factor code the degree of the code and the class degree of the code are the same.

Theorem 4.13. *Let $\pi : X \to Y$ be a finite-to-one factor code from a SFT to an irreducible sofic shift Y. Then $c_\pi = d_\pi$.*

Proof. Theorem 4.1 and Theorem 4.12 imply that the degree of a code and a class degree of a code are both invariant under conjugacy. So by using Proposition 2.1, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code with a magic symbol w. Let y be a right transitive point of Y. There is a strictly increasing sequence of integers $(a_i)_{i \in \mathbb{N}}$ such that $y_{a_i} = w$. Proposition 4.3 implies that for each $i \in \mathbb{N}$, $y_{[a_1, a_i]}$ has exactly d_π pre-images which are mutually separated. It follows that for each $x, x' \in \pi^{-1}(y)$ with $x_{a_1} = x'_{a_1}$ we must have $x_{[a_1,\infty)} = x'_{[a_1,\infty)}$ which implies $x \sim x'$. Since $y_{a_1} = w$ has exactly d_π pre-images we conclude that there are exactly d_π transition equivalence classes over y. □

We show when $\pi : X \to Y$ is a factor code from a SFT X to an irreducible sofic shift Y then there are exactly c_π transition equivalence classes over a typical point of Y. In order to show this, we introduce another quantity c^*_π in Definition 4.14, defined concretely in terms of blocks. Proposition 2.1 and Theorem 4.12 allow us to focus only on a triple (X, Y, π) where X is a 1-step SFT and π is a 1-block factor code with a magic symbol.

Definition 4.14. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y with a magic symbol. Let W be a Y-block of length $p+1$ which begins and ends with magic symbols. Let n be a positive integer less than p, and M be a subset of $\pi^{-1}_{sb}(W_n)$. We say $U \in \pi^{-1}_{sb}(W)$ **is routable through** $a \in M$ **at time** n if there is a block $U' \in \pi^{-1}_{sb}(W)$ with $U'_0 = U_0$, $U'_n = a$, and $U'_p = U_p$. A triple (W, n, M) is called a **transition block** of π if every block in $\pi^{-1}_{sb}(W)$ is routable through a symbol of M at time n. The cardinality of the set M is called the **depth** of the transition block (W, n, M).*

While Definition 4.14 seems complicated, Figure 4 illustrates it more clearly.

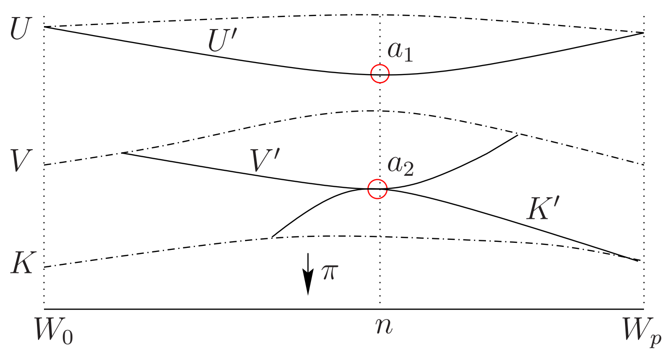

Figure 4: (W, n, M) is a transition block with $M = \{a_1, a_2\}$. The blocks U, V, $K \in \pi^{-1}_{sb}(W)$ are routable through members of M at time n via blocks U', V', $K' \in \pi^{-1}_{sb}(W)$.

Definition 4.15. *Reusing notations of Definition 4.14, define*

$$c^*_\pi = \min\{|M| \colon (W, n, M) \text{ is a transition block of the factor code } \pi\}.$$

*A **minimal transition block** of a 1-block factor code π from a 1-step SFT to a sofic shift is a transition block of depth c^*_π.*

Example 4.16. *In figure 5, we display an example of a labeled graph which defines an infinite-to-one 1-block factor code π. We see that $(1001,\ 2,\ \{b\})$ is a minimal transition block of π of depth 1. For example, observe that block*

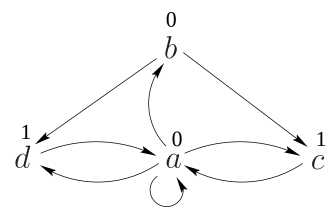

Figure 5: Graph for Example 4.16

$U = daac$ is routable through b at time 2 by considering $U' = dabc$.

We need to develop some lemmas to prove Theorem 4.20 below.

Lemma 4.17. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let $y \in Y$ and $x \in \pi^{-1}(y)$. There is an integer $m < \infty$ such that for each $n \geq m$ the symbol x_n is not transient. In fact, m can be found in such a way that for each $n \geq m$, x_n belongs to the class $[x]$ at time n, i.e., $x_n \in S_n([x])$.*

Proof. Consider $C \in \mathscr{C}(y)$ with $[x] \nrightarrow C$. It follows that there exists $i < \infty$ such that for $z \in \pi^{-1}(y)$ if $z_i = x_i$ then $z \notin C$. Denote the smallest such i by i^C and let $m = \max\{i^C : C \in \mathscr{C}(y),\, [x] \nrightarrow C\}$ ($m < \infty$ since $|\mathscr{C}(y)| < \infty$). Let $n \geq m$. The above shows that if $C \in \mathscr{C}(y)$ and $[x] \nrightarrow C$ then $C \notin \mathscr{F}(x_n, y_{[n+1,\infty)})$. On the other hand if $C \in \mathscr{C}(y)$ satisfies $[x] \to C$ let z be a point in C with $z_n = x_n$, therefore $C \in \mathscr{F}(x_n, y_{[n+1,\infty)})$. Hence we have $x_n \in S_n([x])$. We mention that m can be $-\infty$. ☐

Lemma 4.18. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let $y \in Y$. There is an integer $m < \infty$ such that for each $n \geq m$ and $C \in \mathscr{C}(y)$ there is a symbol $i \in \pi_{sb}^{-1}(y_n)$ with $i \in S_n(C)$.*

Proof. Let $\mathscr{C}(y) = \{C_1, \ldots, C_d\}$ for some $d < \infty$. Let $A = \{x^{(1)}, \ldots, x^{(d)}\}$ be a set containing an arbitrary point $x^{(i)} \in C_i$ for each $C_i \in \mathscr{C}(y)$. By Lemma 4.17 there is a finite $n_{C_i} \in \mathbb{Z}$ such that for each $n \geq n_{C_i}$ we have $x_n^{(i)} \in S_n(C_i)$. Let $m = \max\{n_{C_i} : 1 \leq i \leq d\}$. Then $m < \infty$ and for each $n \geq m$ and C_i we have $x_n^{(i)} \in S_n(C_i)$. ☐

Proposition 4.19. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let ν be an invariant measure on Y, then*

$$\nu\left(\{y \in Y : \forall C \in \mathscr{C}(y), \forall n \in \mathbb{Z} \text{ there is } i \in \pi_{sb}^{-1}(y_n) \text{ with } i \in S_n(C)\}\right) = 1.$$

Proof. Let ν be an invariant measure on Y and $y \in Y$. Let $m(y) < \infty$ be the infimum of the set of m's with the properties given in the statement of Lemma 4.18. We show $\nu(\{y \in Y : m(y) = -\infty\}) = 1$. Note that $C \in \mathscr{C}(y)$ if and only if $T(C) \in \mathscr{C}(T(y))$. This implies that $m(T(y)) = m(y) - 1$. For $k < \infty$ let $A_k = \{y \in Y : m(y) = k\}$ so that $T(A_k) = A_{k-1}$. Since ν is T-invariant it follows that $\nu(A_k) = 0$. Therefore $m(y) = -\infty$ for ν-a.e. $y \in Y$. □

Theorem 4.20. *Let $\pi : X \to Y$ be a factor code from a SFT X to an irreducible sofic shift Y. There are exactly c_π transition equivalence classes over every right transitive point of Y.*

Proof. First we show that $|\mathscr{C}(y)| \geq c^*_\pi$ for every $y \in Y$. This implies that $c_\pi = \inf\{|\mathscr{C}(y)| : y \in Y\} \geq c^*_\pi$. Then we prove $|\mathscr{C}(y)| \leq c^*_\pi$ when $y \in Y$ is right transitive. This implies $c_\pi \leq c^*_\pi$. Then it follows that for each right transitive $y \in Y$ we have $|\mathscr{C}(y)| = c^*_\pi = c_\pi$.

By Theorem 4.12, c_π is invariant under conjugacy. So by Proposition 2.1, without loss of generality, we may assume X is a 1-step SFT, and π is a 1-block factor code with a magic symbol.

We prove $|\mathscr{C}(y)| \geq c^*_\pi$ by showing that if $h = |\mathscr{C}(y)| < c^*_\pi$ then there is a transition block of depth h occurring in y. This gives a contradiction to the assumption that a minimal transition block of the code π is of depth c^*_π. We find such a transition block of y in the following 4 stages; suppose $\mathscr{C}(y) = \{C_1, \dots, C_h\}$. Choose a finite integer n_0 satisfying properties given in the statement of Lemma 4.18 such that y_{n_0} is a magic symbol.

Stage *1*. We claim there is $n_1 \in [n_0, \infty)$ such that for each $x \in \pi^{-1}(y)$, x_t is not transient for some $n_0 \leq t \leq n_1$. Suppose there is no such n_1. It follows that for each $j \geq n_0$ there is $x^{(j)} \in \pi^{-1}(y)$ such that $x^{(j)}_l$ is transient for all $n_0 \leq l \leq j$. Consider the sequence $(x^{(j)})_{j\in\mathbb{Z}}$, and let x be the limit of a convergent subsequence of it. Clearly $x \in \pi^{-1}(y)$. However, x_l is transient for each $l \geq n_0$; contradicting Lemma 4.17.

Stage *2*. We claim there is $n_2 \in [n_1, \infty)$ and a set of symbols $M' = \{a_1, \dots, a_h\}$ with $a_e \in S_{n_2}(C_e)$ such that for each $i \in S_{n_1}(C_e)$ there is a block

$U \in \pi_{sb}^{-1}(y_{n_1}y_{n_1+1}\dots y_{n_2-1}y_{n_2})$ which begins with i and ends with a_e. See Figure 6. Let $C_e \in \mathscr{C}(y)$ and $x^e \in C_e$. For each $i \in S_{n_1}(C_e)$ (non-empty by

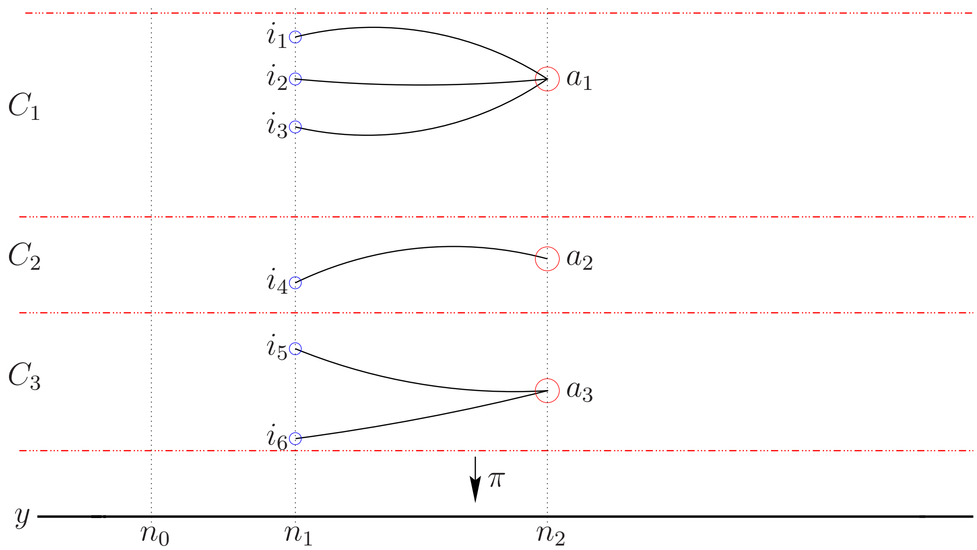

Figure 6: An example illustrating Stage 2. $\mathscr{C}(y) = \{C_1,\ C_2,\ C_3\}$, $i_1,\ i_2,\ i_3 \in S_{n_1}(C_1)$, $i_4 \in S_{n_1}(C_2)$, $i_5,\ i_6 \in S_{n_1}(C_3)$, and $M' = \{a_1,\ a_2,\ a_3\}$.

the choice of n_0) by the definition of a transition equivalence class, there is a point $z^i \in C_e$ with $z^i_{n_1} = i$, and z^i is matching x^e from some time $k^i \in [n_1, \infty)$. Let $k^e = \max\{k^i : i \in S_{n_1}(C_e)\} < \infty$ and $n_2 = \max\{k^e : C_e \in \mathscr{C}(y)\} < \infty$. Rename $x^e_{n_2}$ for each $C_e \in \mathscr{C}(y)$ as a_e, and let $M' = \{a_1, \dots, a_h\}$.

Stage *3*. We claim there is $n_3 \in [n_2, \infty)$ such that for each $x \in \pi^{-1}(y)$ there is $x' \in \pi^{-1}(y)$ so that $x'_r = x_r$ for every $r \in (-\infty, n_0] \cup [n_3, \infty)$, and $x'_{n_2} \in M'$. See Figure 7. Let $x \in \pi^{-1}(y)$. By stage 1 there is $n_0 \leq t \leq n_1$ such that x_t is not transient; i.e., $x_t \in S_t(C_e)$ for some $C_e \in \mathscr{C}(y)$. It follows that there is a point $u \in C_e$ with $u_t = x_t$. Then the new block $\dots x_{n_0} \dots x_t u_{t+1} \dots u_{n_1}$ belongs to $\mathscr{L}(X)$ (X is a 1-step SFT) and maps to $\dots y_t \dots y_{n_1}$. Moreover, by Stage 2 there is a path starting at $u_{n_1} \in S_{n_1}(C_e)$ ending at a_e, say $u_{n_1}v_{n_1+1}\dots v_{n_2-1}a_e$ mapping to $y_{n_1}\dots y_{n_2}$. Connect these two paths at u_{n_1} to get $\dots x_{n_0} \dots x_t u_{t+1} \dots u_{n_1} v_{n_1+1} \dots v_{n_2-1} a_e$ in $\mathscr{L}(X)$. Observe that having $x_t \in S_t(C_e)$ implies that x must belong to a class $C_f \in \mathscr{C}(y)$ with $C_e \to C_f$. On the other hand having $a_e \in S_{n_2}(C_e)$ implies that there is a point $b \in C_e$ with $b_{(-\infty, n_2]} = \dots x_{n_0} \dots x_t u_{t+1} \dots u_{n_1} v_{n_1+1} \dots v_{n_2-1} a_e$ which matches x from some time j onwards for some $n_2 \leq j < \infty$. Let

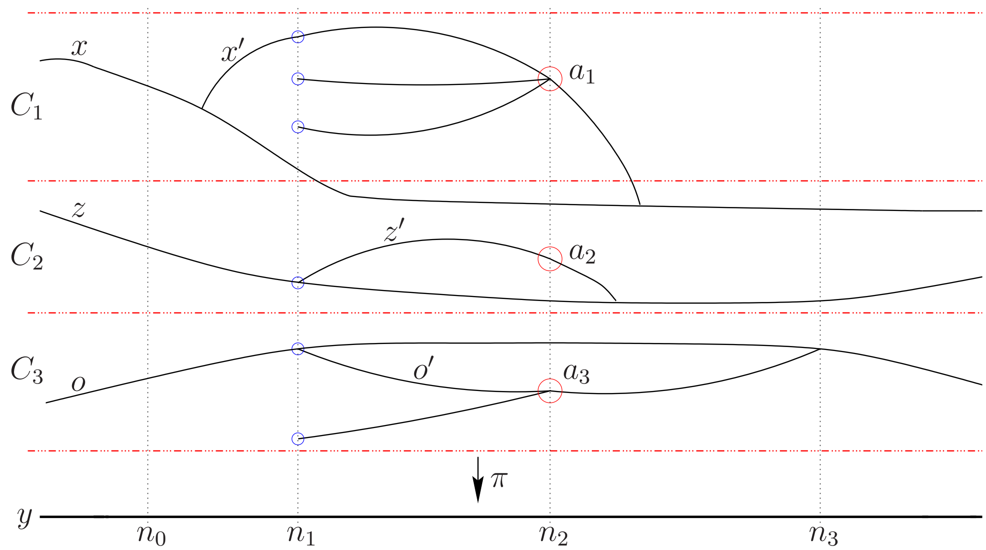

Figure 7: Graph for Stage 3. $M' = \{a_1,\, a_2,\, a_3\}$. $x, x' \in \pi^{-1}(y)$, $x'_r = x_r$ for each $r \in (-\infty, n_0] \cup [n_3, \infty)$, and $x'_{n_2} \in M'$. Same for z, z' and o, o'.

$n_x = \min\{n_2 \leq j < \infty : \exists x' \in \pi^{-1}(y)$ with $x'_r = x_r$ for all $r \in (-\infty, n_0] \cup [j, \infty),\ \ x'_{n_2} = a_e\} < \infty$. We claim that there is $n^e < \infty$ such that for each $x \in \pi^{-1}(y)$ with $x_t \in S_t(C_e)$ for some $n_0 \leq t \leq n_1$, there exists a point $x' \in \pi^{-1}(y)$ with $x'_r = x_r$ for each $r \in (\infty, n_0] \cup [n^e, \infty)$, and $x'_{n_2} = a_e$. Then letting $n_3 = \max\{n^e : C_e \in \mathscr{C}(y)\} < \infty$ will complete the proof of Stage 3.

Suppose, to derive a contradiction, that there does not exist such n^e. It follows that there is a sequence $x^{(l)}$ with $x^{(l)}_{t_l} \in S_{t_l}(C_e)$ for some $n_0 \leq t_l \leq n_1$, such that $\lim_{l\to\infty} n_{x^{(l)}} = \infty$. Let x^* be the limit of a convergent subsequence of $x^{(l)}$. Clearly $x^* \in \pi^{-1}(y)$. However, $n_{x^*} = \infty$ since if $n_{x^*} = g < \infty$ then for every point k in the given subsequence of $x^{(l)}$ with $d(x^*, k) < 1/2^g$ we have $n_k = g$ which contradicts the assumption that $\lim_{l\to\infty} n_{x^{(l)}} = \infty$.

Stage 4. Let $n_4 = \min\{i \geq n_3 : y_i$ is a magic symbol$\}$. Form the block $V = y_{n_0} \dots y_{n_4}$. We show that V is a transition block of depth h. Let $U \in \pi_{sb}^{-1}(V)$. By Proposition 2.2 there is $x \in \pi^{-1}(y)$ with $x_{[n_0,\, n_4]} = U$. Then by Stages 1, 2, and 3 there is a point $x' \in \pi^{-1}(y)$ with $x'_r = x_r$ for each $r \in (-\infty, n_0] \cup [n_3, \infty)$, and $x'_{n_2} \in M'$. Having $x'_{n_0} = x_{n_0}$, $x'_{n_2} \in M'$, $x'_{n_4} = x_{n_4}$ and $\pi_{\mathrm{sb}}(x'_{[n_0,\, n_4]}) = V$ simply means that $U = x_{[n_0,\, n_4]}$ is routable through a symbol of M' at time n_2. Since $U \in \pi_{sb}^{-1}(V)$ is arbitrary it follows that V is a minimal transition block.

Now we apply the same method we used in the proof of Theorem 4.7 to show that if y is a right transitive point of Y then $|\mathscr{C}(y)| \leq c^*_\pi$. Let (W, n, M) with $|W| = p + 1$ be a minimal transition block of the factor code π. Since y is a right transitive point there is a sequence of integers $(t_m)_{m\in\mathbb{N}}$ with $t_{m+1} > p + t_m$ such that $y_{t_m} \dots y_{p+t_m} = W$ for each $m \in \mathbb{N}$. Suppose $\mathscr{C}(y)$ contains more than c^*_π transition equivalence classes and A be a set containing one point from each of these equivalence classes. By the definition of minimal transition block, for each $u \in A$ there exists a point $x \in \pi^{-1}({}_{t_m}[W])$ with $x_{t_m} = u_{t_m},\ x_{n+t_m} \in M$, and $x_{p+t_m} = u_{p+t_m}$. Denote such a point by $x^{(u,m)}$. We construct a new point u' which agrees with u everywhere except at positions $t_m, \dots, s + t_m$ where it agrees with $x^{(u,m)}$ for each $m \in \mathbb{N}$. In other words we have

$$u'_i = \begin{cases} u_i & \text{if } -\infty < i < t_1,\ s + t_m < i < t_{m+1} \\ x_i^{(u,m)} & \text{if } t_m \leq i \leq s + t_m. \end{cases}$$

The assumption that X is a 1-step SFT guarantees that u' belongs to X, and Fact 4.5 implies that $u' \sim u$. For each element u of A construct such a point u' and collect them in a set denoted by A'. Since $|M| = c^*_\pi < |A'|$ then for each $m \in \mathbb{N}$ there must be at least two points in A' that agree in the $(n + t_m)$th position. It follows that there exists a subsequence $(k_l)_{l\in\mathbb{N}}$ of $(t_m)_{m\in\mathbb{N}}$ and two distinct points $u', v' \in A'$ such that for each $l \in \mathbb{N}$ we have $u'_{n+k_l} = v'_{n+k_l}$. Then by Fact 4.5 we have $u' \sim v'$ and consequently $u \sim v$, contrary to the fact that the points u and v are chosen from distinct transition equivalence classes. □

5 Bounding the number of ergodic measures of relative maximal entropy

In Section 3 we mentioned that although every 1-dimensional irreducible SFT has a unique ergodic measure (Parry measure) of maximal entropy, there can be more than one ergodic measure of relative maximal entropy over an ergodic measure; i.e., given a factor triple (X, Y, π) and a fully supported ergodic measure ν on Y, there can exist more than one ergodic measure on X that projects to ν under π and have maximal entropy among measures in the fiber $\pi^{-1}\{\nu\}$, see [20, Example 3.3]. In this section we show that the number of such measures can be no more than the class degree of π.

Since entropy is a conjugacy invariant the following observation follows immediately;

Observation 5.1. *Let* (X, Y, π) *and* $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ *be conjugate factor triples. Let* ν *be an ergodic measure on* Y *and* $\tilde{\nu}$ *be its corresponding ergodic measure on* $\tilde{Y}$*. The number of ergodic measures of relative maximal entropy over* ν *is the same as the number of ergodic measures of relative maximal entropy over* $\tilde{\nu}$*.*

In 2003, Petersen, Quas, and Shin [20] found an upper bound on the number of measures of relative maximal entropy;

Theorem 5.2. *[20, Corollary 1] Let* $\pi : X \to Y$ *be a 1-block code from a 1-step SFT* X *to a sofic shift* Y*. Let* ν *be a fully supported ergodic measure on* Y *and* $N(\pi) = \min\{\pi_{sb}^{-1}(b) \colon b \in \mathcal{A}(Y)\}$*. The number of ergodic measures of maximal entropy over* ν *is at most* $N(\pi)$*.*

This bound suffers from being invariant under conjugacy. For example, the full 2-shift and any higher block presentation of it give different bounds on the number of ergodic measures of maximal entropy which map to the trivial measure on the full 1-shift.

One possibility to avoid having a non-invariant upper bound is to take the minimum of the bound in [20] over all conjugate factor triples with 1-block factor codes. This even improves the original bound. However knowing that two SFTs are conjugate is not easy. Williams' Classification Theorem [30] gives an algebraic criterion for the conjugacy of shifts of finite type: two shifts of finite type X_A and X_B are conjugate if and only if their transition matrices A and B are strong shift equivalent. However, there is no known general algorithm for deciding whether two matrices are strong shift equivalent. Kim and Roush [11] showed some theoretical procedures that will decide whether or not two matrices are shift equivalent, but the question of whether shift equivalence implies strong shift equivalence, known as the "Shift Equivalence Problem" or "Williams' Conjecture" was open for more than twenty years. In 1999 Kim and Roush [13] solved the Shift Equivalence Problem in the negative by constructing two irreducible SFTs that are shift equivalent but not strong shift equivalent. Earlier they had found a reducible example [12].

Here we show that the class degree of a factor code is an upper bound on the number of measures of relative maximal entropy over a fully supported ergodic measure. Theorem 4.12 above verifies that this bound is invariant under conjugacy and Proposition 5.3 below shows that it beats the bound mentioned above obtained by minimizing the bound in [20] over conjugate factor triples.

Proposition 5.3. *Let $\pi : X \to Y$ be a 1-block factor code from a SFT X to a sofic shift Y. Then $c_\pi \leq \min\{N(\tilde{\pi})\colon (X,Y,\pi) \sim (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi}$ is 1-block$\}$. Equality holds if π is 1-block and finite-to-one.*

Proof. Let $d = \min\{N(\tilde{\pi})\colon (X,Y,\pi) \sim (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi}$ is 1-block$\}$ occur at a factor triple $(\tilde{X},\tilde{Y},\tilde{\pi})$ (i.e., $N(\tilde{\pi}) = d$) and $\min\{\tilde{\pi}_{sb}^{-1}(b)\colon b \in \mathcal{A}(\tilde{Y})\}$ occur at a symbol $b \in \mathcal{A}(\tilde{Y})$. Let ν be a fully supported ergodic measure on Y. Note that the ν's corresponding measure on $\tilde{Y}$ is also fully supported and ergodic, so with respect to this measure the block $[b]$ has a positive measure and almost every point of $\tilde{Y}$ is right transitive. Let y be a right transitive point of $\tilde{Y}$. There is a strictly increasing sequence of integers $(a_i)_{i\in\mathbb{N}}$ with $y_{a_i} = b$. Corollary 4.8 implies that $|\mathscr{C}(y)| \leq |\tilde{\pi}_{sb}^{-1}(b)| = d$ and therefore $c_{\tilde{\pi}} = |\mathscr{C}(y)| \leq d$. The first part of the Proposition follows from Theorem 4.12 which states that $c_\pi = c_{\tilde{\pi}}$.

Now suppose $\pi : X \to Y$ is 1-block and finite-to-one. Since the class degree is invariant under conjugacy, without loss of generality, we may assume π has a magic symbol denoted by w (Proposition 2.1). Since π is finite-to-one Proposition 4.3 implies that $|\pi_{sb}^{-1}(w)| = d_\pi$. Then by the definition of a magic symbol and the fact that d_π is invariant under conjugacy we have $\min\{N(\tilde{\pi})\colon (X,Y,\pi) \sim (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi}$ is 1-block$\} = |\pi_{sb}^{-1}(w)|$. Since by Theorem 4.13 $d_\pi = c_\pi$ it follows that $\min\{N(\tilde{\pi})\colon (X,Y,\pi) \sim (\tilde{X},\tilde{Y},\tilde{\pi})\} = c_\pi$. ☐

Note that the equality in Proposition 5.3 does not always hold. For example, consider the trivial factor code $\pi : \{0,1\}^{\mathbb{Z}} \to \{0\}^{\mathbb{Z}}$. Then $c_\pi = 1$; however, if $\left(\tilde{X}, \{0\}^{\mathbb{Z}}, \tilde{\pi}\right)$ is a factor triple conjugate to $\left(\{0,1\}^{\mathbb{Z}}, \{0\}^{\mathbb{Z}}, \pi\right)$ then $\mathcal{A}(\tilde{X})$ must be strictly greater than 1 and therefore $N(\tilde{\pi}) > 1$.

Definition 5.4. *Let $\pi : X \to Y$ be a 1-block factor code from a SFT X to a sofic shift Y and ν be an ergodic measure on Y. Let $\mu_1, \ldots, \mu_n$ be invariant measures in the fiber $\pi^{-1}\{\nu\}$. The* relatively independent joining *$\tilde{\mu} = \mu_1 \otimes \cdots \otimes_\nu \mu_n$ of $\mu_1, \ldots, \mu_n$ over ν is defined as follows: if $A_1, \ldots, A_n$*

are measurable subsets of X then

$$\hat{\mu}(A_1 \times \cdots \times A_n) = \int_Y \prod_{i=1}^{n} \mathbb{E}_{\mu_i}(1_{A_i} | \pi^{-1}\mathscr{B}_Y) \circ \pi^{-1} \, d\nu.$$

Writing p_i for the projection $X^n \to X$ onto the i'th coordinate, it follows from the definition that for $\hat{\mu}$-almost every $\hat{x} \in X^n$, $\pi(p_i(\hat{x}))$ is independent of i.

We will use Theorem 5.5 below which is the main theorem from [20] to prove a stronger theorem (Theorem 5.6).

Theorem 5.5. *[20, Theorem 1] Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let ν be an ergodic measure on Y, and two distinct ergodic measures μ_1 and μ_2 be measures of relative maximal entropy over ν. Then $(\mu_1 \otimes \mu_2)\{(u, v) \in X \times X \colon u_0 = v_0\} = 0$.*

Theorem 5.6. *Let (X, Y, π) be a factor triple. Let ν be an ergodic measure on Y, and two distinct measures μ_1 and μ_2 be ergodic measures of relative maximal entropy over ν. Then $(\mu_1 \otimes \mu_2)\{(u, v) \in X \times X \colon u \sim v\} = 0$.*

Proof. First we show that, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code with a magic symbol. Suppose $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ is a factor triple conjugate to (X, Y, π) and $\phi : X \to \tilde{X}$ is a conjugacy from X to $\tilde{X}$. By Theorem 4.12 we have $(\phi \times \phi)\{(u, v) \in X \times X \colon u \sim v\} = \{(\phi(u), \phi(v)) \in \tilde{X} \times \tilde{X} \colon \phi(u) \sim \phi(v)\}$. Moreover, the corresponding measure to $\mu_1 \otimes \mu_2$ under the conjugacy $\phi \times \phi : X \times X \to \tilde{X} \times \tilde{X}$ is $\tilde{\mu}_1 \otimes \tilde{\mu}_2$ where $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are corresponding measures to μ_1 and μ_2, correspondingly, under ϕ. It follows that $(\mu_1 \otimes \mu_2)\{(u, v) \in X \times X \colon u \sim v\} = (\tilde{\mu}_1 \otimes \tilde{\mu}_2)\{(\phi(u), \phi(v)) \in \tilde{X} \times \tilde{X} \colon \phi(u) \sim \phi(v)\}$. Therefore by Proposition 2.1 we may assume X is a 1-step SFT and π is a 1-block factor code with a magic symbol.

Let (W, n, M) be a minimal transition block of the factor code π with $|W| = s + 1$. Let $u \in X$, $i \in \mathbb{Z}$, and $a \in M$. We say u potentially passes through a at time i if the block $u_{[i-n,\, i+s-n]} \in \pi_{sb}^{-1}(W)$ and it is routable through the symbol a (see Definition 4.14). Let $v \in X$, and write $u \overset{a}{\approx}_i v$ if $u \sim v$ and both u and v potentially pass through a at time i (we mention that the relation $\overset{a}{\approx}_i$ is not in general an equivalence relation on X). We have

$$\{(u, v) : u \sim v\} = \bigcup_{a \in M} \bigcup_{i \in \mathbb{Z}} \{(u, v) \colon u \overset{a}{\approx}_i v\}.$$

Suppose $(\mu_1 \otimes \mu_2)\{(u,v) : u \sim v\} > 0$. There must be $i \in \mathbb{Z}$ and a symbol $a \in M$ such that

$$(\mu_1 \otimes \mu_2)\{(u,v) \colon u \overset{a}{\approx}_i v\} > 0.$$

By applying $(T \times T)^{-i}$, without loss of generality, we may assume $i = 0$. Considering only blocks of length $s+1$, there must be blocks A, $B \in \pi_{sb}^{-1}(W)$ such that

$$(\mu_1 \otimes \mu_2)\big\{(u,v) \colon u \overset{a}{\approx}_0 v,\ u_{[-n,\,s-n]} = A,\ v_{[-n,\,s-n]} = B\big\} > 0. \tag{5.1}$$

Both A and B are routable through a at time n; i.e., there are blocks A', $B' \in \pi_{sb}^{-1}(W)$ with $A'_0 = A_0$, $B'_0 = B_0$, $A'_n = B'_n = a$, $A'_s = A_s$, and $B'_s = B_s$. Let $G \subseteq \mathbb{Z}$ be the set $\{-n, \ldots, -n+s\}$. Basic properties of conditional expectation imply that

$$\begin{aligned}
\mathbb{E}_{\mu_1}\left(\mathbf{1}_{[A]}|\pi^{-1}(\mathscr{B}_Y)\right) &= \mathbb{E}\left(\mathbb{E}_{\mu_1}\left(\mathbf{1}_{\,{}_{-n}[A]}|\pi^{-1}(\mathscr{B}_Y) \vee \sigma(X_{G^c})\right)|\pi^{-1}(\mathscr{B}_Y)\right)\\
&= \mathbb{E}\left(\mathbb{E}_{\mu_1}\left(\mathbf{1}_{\,{}_{-n}[A']}|\pi^{-1}(\mathscr{B}_Y) \vee \sigma(X_{G^c})\right)|\pi^{-1}(\mathscr{B}_Y)\right) \quad (5.2)\\
&= \mathbb{E}_{\mu_1}\left(\mathbf{1}_{[A']}|\pi^{-1}(\mathscr{B}_Y)\right),
\end{aligned}$$

where the second equality follows from Theorem 3.3. Similarly we have

$$\mathbb{E}_{\mu_2}(\mathbf{1}_{[B]}|\pi^{-1}\mathscr{B}_Y) = \mathbb{E}_{\mu_2}(\mathbf{1}_{[B']}|\pi^{-1}\mathscr{B}_Y). \tag{5.3}$$

Let $D = \big\{(u,v) \colon u \overset{a}{\approx}_0 v,\ u_{[-n,\,s-n]} = A,\ v_{[-n,\,s-n]} = B\big\}$. Since $D \subseteq {}_{-n}[A] \times {}_{-n}[B]$ we have

$$\begin{aligned}
\mu(D) &\le (\mu_1 \otimes \mu_2)([A] \times [B])\\
&= \int_Y \mathbb{E}_{\mu_1}(\mathbf{1}_{[A]}|\pi^{-1}\mathscr{B}_Y)\mathbb{E}_{\mu_2}(\mathbf{1}_{[B]}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}\, d\nu\\
&= \int_Y \mathbb{E}_{\mu_1}(\mathbf{1}_{[A']}|\pi^{-1}\mathscr{B}_Y)\mathbb{E}_{\mu_2}(\mathbf{1}_{[B']}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}\, d\nu \quad \text{using(5.2) and (5.3)}\\
&= (\mu_1 \otimes \mu_2)([A'] \times [B'])\\
&= 0
\end{aligned}$$

where the last equality follows from Theorem 5.5 since $A'_n = B'_n = a$. This contradicts Equation (5.1). ☐

Theorem 5.7. *Let (X, Y, π) be a factor triple and ν be a fully supported ergodic measure on Y. The number of ergodic measures of relative maximal entropy over ν is at most c_π.*

Proof. By Observation 5.1, the number of ergodic measures of relative maximal entropy over ν is invariant under conjugacy. Therefore using Proposition 2.1, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code with a magic symbol. Suppose, for a contradiction, that there are $n > c_\pi$ ergodic measures $\mu_1, \dots, \mu_n$ on X of relative maximal entropy over ν. Form the relatively independent joining $\hat{\mu}$ on X^n of the measures $\mu_1, \dots, \mu_n$. Since ν is a fully supported ergodic measure on Y it follows that for $\hat{\mu}$-a.e. $\hat{x} \in X^n$, $\pi(p_i(\hat{x}))$ (which is independent of i), is a right transitive point of Y. Then the assumption $n > c_\pi$ implies that for $\hat{\mu}$-a.e. $\hat{x} = (x_1, \dots, x_n) \in X^n$ there are distinct i, j such that $p_i(\hat{x}) \sim p_j(\hat{x})$; i.e.,

$$\hat{\mu}(\bigcup_{1 \le i < j \le n} \{\hat{x} = (x_1, \dots, x_n) \colon x_i \sim x_j\}) = 1.$$

At least one of the set $S_{i,j} = \{(x_1, \dots, x_n) \colon x_i \sim x_j\}$ must have positive $\hat{\mu}$-measure. It follows that

$$\begin{aligned} 0 &< \hat{\mu}(S_{i,j}) \\ &= \hat{\mu}(\{(x_1, \dots, x_n),\ x_i \sim x_j\}) \\ &= (\mu_i \otimes \mu_j)\{(u, v) \colon u \sim v\}. \end{aligned}$$

This contradicts Theorem 5.6. ☐

6 Open Questions

Let (X, T) be a one-sided topologically mixing shift of finite type. An invariant measure μ on X is a Gibbs measure corresponding to $f \in C(X)$ if there are constants $C_1,\ C_2 > 0$ and $P > 0$ such that

$$C_1 \le \frac{\mu([x_0 x_1 \dots x_{n-1}])}{\exp(-Pn + (S_n f)(x))} \le C_2$$

for every $x \in X$ and $n \ge 1$, where $(S_n f)(x) = \sum_{k=0}^{n-1} f(T^k(x))$.

Walters [28, 29] introduces a class $\mathrm{Bow}(X, T)$ of functions that contains the functions with summable variation, all of which have unique equilibrium states. Let

$$\mathrm{var}_n(f) = \sup\{|f(x) - f(y)| \colon x,\ y \in X,\ x_i = y_i \text{ for all } 0 \le i \le n-1\}.$$

Then $\mathrm{Bow}(X, T) = \{f \in C(X) \colon \sup_{n \ge 1} \mathrm{var}_n(S_n f) < \infty\}$.

Theorem 6.1. *[28, Theorem 2.16] Let $f \in Bow(X,T)$. Then f has a unique equilibrium state μ which is a Gibbs measure.*

The relative version of this result is an open question. We make the following conjecture;

Conjecture 6.2. *Let (X,Y,π) be a factor triple and ν be a fully supported ergodic measure on Y. The number of ergodic measures of maximal pressure in the fiber $\pi^{-1}\{\nu\}$ is at most c_π.*

References

[1] D. Blackwell. The entropy of functions of finite-state Markov chains. *Transactions of the first Prague Conference on information theory, Statistical decision functions, random processes*, pages 13–20, 1956.

[2] M. Boyle and S. Tuncel. Infinite-to-one codes and Markov measures. *Trans. Amer. Math. Soc.*, 285:657–683, 1984.

[3] C. Burke and M. Rosenblatt. A Markovian function of a Markov chain. *Ann. Math. Statist.*, 29:1112–1122, 1958.

[4] R. Burton and J. Steif. Non-uniqueness of measures of maximal entropy for subshift of finite type. *Ergodic Theory Dynam. Systems.*, 14:213–235, 1994.

[5] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory*, 1:1–49, 1967.

[6] D. Gatzouras and Y. Peres. The variational principle for Hausdorff dimension. In *a survey. Ergodic Theory of Z^d-actions (Warwick, 1993–1994)*, volume 228, pages 113–125. London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, 1996.

[7] D. Gatzouras and Y. Peres. Invariant measures of full dimension for some expanding maps. *Ergodic Theory Dynam. Systems.*, 17(1):147–167, 1997.

[8] F. Hofbauer. Hausdorff dimension and pressure for piecewise monotonic maps of the interval. *J. London Math. Soc.*, 47(2):142–156, 1993.

[9] R. Israel. *Convexity in the theory of lattice gases*. Princeton University Press, Princeton, 1979.

[10] G. Keller. *Equilibrium states in ergodic theory*. Cambridge Univ. Press, Cambridge, 1998.

[11] K. Kim and F. Roush. Decidability of shift equivalence. In *Dynamical systems, Lecture Notes in Math.*, volume 1342, pages 374–424. Springer, Berlin, 1988.

[12] K. Kim and F. Roush. Williams's conjecture is false for reducible subshifts. *J. Amer. Math. Soc.*, 5:213–215, 1992.

[13] K. Kim and F. Roush. The Williams conjecture is false for irreducible subshifts. *Ann. of Math.*, 149:545–558, 1999.

[14] F. Ledrappier and P. Walters. A relativised variational principle for continuous transformations. *J. London Math. Soc*, 16:568–576, 1977.

[15] F. Ledrappier and L. Young. The metric entropy of diffeomorphisms ii. *Ann. of Math.*, 122:540–574, 1985.

[16] D. Lind and B. Marcus. *Symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

[17] B. Marcus, K. Petersen, and S. Williams. Transmission rates and factors of Markov chains. In *Conference in modern analysis and probability, Contemp. Math.*, volume 26, pages 279–293. Amer. Math. Soc., 1984.

[18] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.*, 112:55–66, 1964.

[19] K. Petersen. Information compression and retention in dynamical processes. Notes of lecture course at Workshop on Dynamics and Randomness, Santiago, Chile, 2000.

[20] K. Petersen, A. Quas, and S. Shin. Measures of maximal relative entropy. *Ergodic Theory Dynam. Systems.*, 23(1):207–223, 2003.

[21] P. Raith. Hausdorff dimension for piecewise monotonic maps. *Studia Math*, 94:17–33, 1989.

[22] D. Ruelle. Statistical mechanics on a compact set with F action satisfying expansiveness and specification. *Trans. Amer. Math. Soc.*, 185:237–251, 1973.

[23] D. Ruelle. Repellers for real analytic maps. *Ergodic Theory Dynam. Systems.*, 2:99–107, 1982.

[24] C. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[25] S. Shin. Measures that maximize weighted entropy for factor maps between subshifts of finite type. *Ergodic Theory Dynam. Systems.*, 21:1249–1272, 2001.

[26] P. Walters. *An introduction to ergodic theory.* Springer Verlag New York, 1982.

[27] P. Walters. Relative pressure, relative equilibrium, states, compensation functions and many-to-one codes between subshifts. *Trans. Amer. Math. Soc.*, 296:1–31, 1986.

[28] P. Walters. Convergence of the ruelle operator for a function satisfying bowen's condition. *Trans. Amer. Math. Soc.*, 353:327–347, 2001.

[29] P. Walters. Regularity conditions and Bernoulli properties of equilibrium states and g-measures. *J. London Math. Soc.*, 71(2):379–396, 2005.

[30] R. Williams. Classification of subshifts of finite type. *Ann. of Math.*, 98:120–153, 1973.

[31] Y. Yayama. Dimensions of compact invariant sets of some expanding maps. *Ergodic Theory Dynam. Systems.*, 29(1):281–315, 2009.