Ergod. Th. & Dynam. Sys. (2019), **39**, 865–888
doi:10.1017/etds.2017.50

Relative equilibrium states and class degree

MAHSA ALLAHBAKHSHI†, JOHN ANTONIOLI‡ and JISANG YOO§

† *Universidad de Santiago de Chile, Alameda 3363, Santiago, Chile*
(e-mail: mahnaz.allahbakhshi@usach.cl)
‡ *University of Denver, 2199 S University Boulevard, Denver, CO 80208, USA*
(e-mail: john.antonioli@du.edu)
§ *Seoul National University, 1 Gwanak-ro, Daehak-dong,*
Gwanak-gu, Seoul, South Korea
(e-mail: jisangy@kaist.ac.kr)

(*Received* 24 *October* 2016 *and accepted in revised form* 27 *April* 2017)

Abstract. Given a factor code π from a shift of finite type X onto a sofic shift Y, an ergodic measure ν on Y, and a function V on X with sufficient regularity, we prove an invariant upper bound on the number of ergodic measures on X which project to ν and maximize the measure pressure $h(\mu) + \int V\, d\mu$ among all measures in the fiber $\pi^{-1}(\nu)$. If ν is fully supported, this bound is the class degree of π. This generalizes a previous result for the special case of $V = 0$ and thus settles a conjecture raised by Allahbakhshi and Quas.

1. *Introduction*

It is a classical result that given a one-dimensional irreducible shift of finite type X there is a unique measure of maximal entropy which is an easily described Markov measure [**7**]. Equilibrium states, a concept more general than measures of maximal entropy, are probability measures on topological spaces that are characterized by variational principles. They maximize the sum of an entropy and an energy like quantity. This theory was developed by Sinai, Ruelle, and Bowen [**4**, **9**, **10**] for Hölder potentials on hyperbolic dynamical systems and since then has been applied to other systems as well. Given a real-valued function V defined on X with enough regularity, there is a unique measure on X, called the equilibrium state of V, that maximizes $h(\mu) + \int V\, d\mu$ (called the free energy of μ with respect to V or measure-theoretical pressure of V with respect to μ). So the equilibrium state of the potential function $V = 0$ is the measure of maximal entropy.

We consider the relative case where a factor code $\pi : X \to Y$ from a shift of finite type X to a sofic shift Y is fixed and an ergodic measure ν on Y is given. In the relative case, our attention is restricted to the measures in the fiber $\pi^{-1}(\nu)$. Even when X is irreducible, there can be more than one measure that maximizes entropy among measures in $\pi^{-1}(\nu)$. These measures are called measures of relative maximal entropy. Petersen, Quas, and Shin proved that the number of ergodic measures of relative maximal entropy is always finite and gave an explicit upper bound [**8**]. Allahbakhshi and Quas improved the upper bound to a conjugacy-invariant upper bound and introduced the notion of class degree [**2**]. In the special case of ν with full support, their upper bound is equal to the class degree of the factor code. In the same paper, they proposed the conjecture that the class degree may also be an upper bound for the number of ergodic relative equilibrium states. A proof of this conjecture is the main result of our paper. We state the main theorem as follows. (In this paper, an SFT means a two-sided subshift of finite type.)

THEOREM 1.1. *Let $\pi : X \to Y$ be a factor code from a* SFT *X to a sofic shift Y and ν an ergodic measure on Y. Let V be a function on X satisfying the Bowen condition. Then the number of ergodic relative equilibrium states of V over ν is at most the class degree of ν.*

The idea of relative pressure was introduced and studied by Ledrappier and Walters. Some of the interest of relative equilibrium states arises from its connections with compensation functions [**11**].

In order to motivate parts of the proof and explain the new techniques introduced, we outline briefly how previous results are shown. The finiteness of the number of ergodic measures of relative maximal entropy in [**8**] is proved in the following way. Considering the case (always achievable by recoding) that π is a one-block code from a one-step shift of finite type, suppose $\mu_1, \ldots, \mu_{d+1}$ are distinct ergodic measures of relative maximal entropy over ν where d is the number of letters for X that project to a fixed letter b for Y with $\nu(b) > 0$. Form a relatively independent joining of the $d + 1$ measures over ν. The pigeonhole principle then forces at least two of the $d + 1$ measures, say μ_1, μ_2, to have the property

$$\lambda(\{(x^{(1)}, x^{(2)}) : x_0^{(1)} = x_0^{(2)}\}) > 0,$$

where $\lambda = \mu_1 \otimes_\nu \mu_2$ is the relatively independent joining of the two measures over ν. Since for every $(x^{(1)}, x^{(2)})$ in some set of positive measure with respect to λ, there are infinitely many i for which $x_i^{(1)} = x_i^{(2)}$, one can construct a point $x^{(3)}$ which is the result of splicing together parts of $x^{(1)}$ or $x^{(2)}$ depending on the outcome of tossing a fair coin at every i for which $x_i^{(1)} = x_i^{(2)}$. The probability distribution of the new point $x^{(3)}$ is a measure μ_3 on X which projects to ν. The entropy of the new measure μ_3 is then shown, through the application of Jensen's inequality, to be strictly greater than the entropy of μ_1 or μ_2. This contradicts the initial assumption μ_1 and μ_2 are measures of relative maximal entropy. Therefore the number of ergodic measures of relative maximal entropy over ν cannot exceed the number of preimages of any symbol of Y with positive probability with respect to ν. This bound, however, suffers from not being invariant under conjugacy.

Given a finite-to-one factor code π, the *degree* of π is defined to be the minimal cardinality of the set of preimages of a point in Y. When π is allowed to be infinite-to-one, the class degree introduced in [**2**] is considered to be a natural generalization of the degree [**1**]. The *class degree* of a factor code is defined using an equivalence relation within fibers. Given a point y, the fiber $\pi^{-1}(y)$ is divided into finitely many components under the following equivalence relation: $x, x' \in \pi^{-1}(y)$ are equivalent if there is an x'' in the same fiber that agrees with x on $(-\infty, n]$ for a given arbitrary n and with x' on $[m, +\infty)$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

for some $m > n$ and *vice versa*. The equivalence classes here are called transition classes over y. The number of transition classes over any transitive point $y \in Y$ is finite and the same. This number is defined to be the class degree of the factor code. It is equal to the degree when π is finite-to-one, it is conjugacy invariant and inherits many important properties of the degree [**1**, **2**].

The result in [**2**] on bounding the number of ergodic measures of relative maximal measures by the class degree (when ν has full support) is proved in the following way. Suppose $\mu_1, \ldots, \mu_{d+1}$ are distinct ergodic measures of relative maximal entropy over ν where d is the class degree. As before, form a relatively independent joining of the $d + 1$ measures over ν and apply the pigeonhole principle to conclude that for at least two measures, say μ_1, μ_2, we have

$$\lambda(\{(x^{(1)}, x^{(2)}) : x^{(1)} \sim x^{(2)}\}) > 0,$$

where $x^{(1)} \sim x^{(2)}$ means the two points are in the same transition class and $\lambda = \mu_1 \otimes_\nu \mu_2$. Then the uniform conditional distribution property of measures of relative maximal entropy is used to show that this implies

$$\lambda(\{(x^{(1)}, x^{(2)}) : x_0^{(1)} = x_0^{(2)}\}) > 0,$$

and a contradiction follows.

The proof of our result starts similarly by supposing that there are $d + 1$ distinct ergodic relative equilibrium states of V. In our proof, we have to construct a new measure that satisfies a condition stronger than

$$h(\mu_3) > h(\mu_1) \quad \text{or} \quad h(\mu_3) > h(\mu_2),$$

namely

$$h(\mu_3) > h(\mu_1) + \left| \int V \, d\mu_1 - \int V \, d\mu_3 \right| \quad \text{or}$$
$$h(\mu_3) > h(\mu_2) + \left| \int V \, d\mu_2 - \int V \, d\mu_3 \right|.$$

In other words, we need to construct a measure with the increase in entropy big enough that it overcomes the difference in integrals. An ingredient in our proof is an observation made in the following result by Antonioli [**3**]. Given a relative equilibrium state μ of V with summable variation, he showed that if μ does not have full support and ν does, then one can construct a new measure μ' by routing parts of a point in X depending on the outcomes of tossing a coin. The new measure has strictly greater relative pressure than the original. If routing is done by using an X-block with zero measure, then it is known that the new measure has a bigger entropy (which is an earlier observation made by Yoo [**12**] and proves that any measure of maximal relative entropy over ν has full support). Antonioli's new observation is that if a biased coin is used, then as the probability of coming up tails approaches zero, the difference in entropy (between the new measure μ' and the old measure μ) dominates the difference in integral $\int V \, d\mu' - \int V \, d\mu$ (which proves that any relative equilibrium state of V over ν has full support). The observation relies on the restorability of the old point from the new point.

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

In our setting where we have a joining of two measures μ_1, μ_2, given two points $x^{(1)}, x^{(2)}$ (random points whose joint distribution is the joining), we have to form other points $x^{(3)}, x^{(4)}$ by alternating between parts of $x^{(1)}$ and $x^{(2)}$ in some way. The main difficulty in applying Antonioli's observation to our setting is that we cannot restore the old points $x^{(1)}, x^{(2)}$ from the new points $x^{(3)}, x^{(4)}$. But since μ_1, μ_2 are distinct ergodic measures, long blocks from $x^{(1)}$ are distinguishable from long blocks of $x^{(2)}$ with low probability of error. The rate of error goes to zero as the blocks become longer. The difficulty now is that we do not know enough about the speed of convergence of the error rate. To overcome this, a method of tossing a coin for every Nth occurrence of a fixed minimal transition block is introduced and N is chosen in response to the speed of convergence of the error rate. This allows us to construct two new points in such a way that the increase in entropy dominates the difference in integrals even if the speed of convergence of the error rate is slow. To enable this workaround, we prove some new results on the measure theoretic structure of infinite-to-one factor codes which are analogues of previous results on the topological structure.

As far as we know, this is also a new proof of the uniqueness of equilibrium of a sufficiently regular function on a mixing SFT (subshift of finite type), using the idea of routing. But one can argue that the idea of routing is already implicit in the classical proof relying on Ruelle's operator because the operator is a way of measuring the effect of re-routing an initial segment of a one-sided point.

2. *Background*

In this section, we introduce basic terminology and known results that will be used in our proof.

Throughout this paper, measures are always assumed to be probability measures. Shift spaces are assumed to be two-sided one-dimensional shift spaces.

A triple (X, Y, π) is called a *factor triple* if $\pi : X \to Y$ is a factor code from an SFT X to a sofic shift Y. If the factor triple is such that π is a 1-block factor code and X is a 1-step SFT, then it is called a *1-step 1-block factor triple*. Given a factor triple (X, Y, π), there is a 1-step 1-block factor triple (X', Y', π') that is topologically conjugate to (X, Y, π) [**6**].

Definition 2.1. Given a factor triple (X, Y, π) and an invariant measure ν on Y, an invariant measure μ on X is called a *measure of relative maximal entropy* over ν if it projects to ν and its entropy is the biggest among all invariant measures on X that project to ν.

There is always at least one measure of relative maximal entropy over ν. If ν is ergodic, then the ergodic decomposition of such μ decomposes it into ergodic measures of relative maximal entropy over ν.

In the non-relative setting, the Bowen class is one of the broadest classes of functions for which unique equilibrium can be established.

Let $C(X)$ be the set of all real-valued continuous functions on X, equipped with the sup metric. We write σ_X for the shift map of X. We write σ instead of σ_X when X is understood from the context.

Definition 2.2. For a given function $f \in C(X)$ and a finite interval of integers $I \subset \mathbb{Z}$ (i.e. $I = \{a, a+1, \ldots, b-1, b\}$ for some $a, b \in \mathbb{Z}$), let

$$\mathrm{var}_I(f) = \sup\{|f(x) - f(y)| : x_i = y_i, \text{ for all } i \in I\}$$

and define $S_I f = \sum_{i \in I}(f \circ \sigma^i)$. We also define $S_n f = S_{\{0,1,2,\ldots,n-1\}} f$ and

$$\mathrm{var}_n(f) = \mathrm{var}_{\{0,1,2,\ldots,n-1\}}(f).$$

For later convenience, we will write $[n]$ for $\{0, 1, 2, \ldots, n-1\}$.

Then the *Bowen class* of functions is defined by

$$\mathrm{Bow}(X) = \Big\{ f \in C(X) : \sup_{n \geq 1} \mathrm{var}_n(S_n f) < \infty \Big\}.$$

Equivalently, the Bowen class is the set of all $f \in C(X)$ with $\sup_I \mathrm{var}_I(S_I f) < \infty$.

The Bowen class is dense in $C(X)$, which is easily seen because it contains all locally constant functions. It also contains the Hölder continuous functions, functions with summable variation and many other functions with well known regularity conditions.

Definition 2.3. Given a factor triple (X, Y, π), an invariant measure ν on Y and a real-valued function $V \in \mathrm{Bow}(X)$, an invariant measure μ on X is called a *relative equilibrium state* of V over ν if it projects to ν and $h(\mu) + \int V\, d\mu$ is the biggest among all invariant measures on X that project to ν.

There is always at least one relative equilibrium state of V over ν since $\mu \mapsto h(\mu) + \int V\, d\mu$ is upper semi-continuous. If ν is ergodic, then the ergodic decomposition of such μ decomposes it into ergodic relative equilibrium states of V over ν.

We will use the notation $f(\mu)$ to denote the pushforward of μ by a measurable function f and $\mu(V)$ to denote the μ-integral of $V : X \to \mathbb{R}$.

For more on the general theory of relative equilibrium states, see [**11**].

We are not using any advanced probability theory, but in order to reduce the verbosity of our arguments, we will borrow the language of random variables. Random variables here are defined to be almost everywhere defined measurable functions from a fixed Lebesgue space to Polish spaces. The notion of functions of a random variable, joint random variable, and probability distribution of a random variable are adopted.

Since we already use capital letters such as X, Y, Z to denote topological spaces, we will use hat letters such as $\hat{x}$, $\hat{y}$, $\hat{z}$ to denote random variables. Other than that, we adopt the following notational practices that are standard in probability theory literature.

We adopt the usual shortcuts for denoting functions of random variables and joint variables described as follows. If some Lebesgue space $(\Omega, \mathcal{B}, \mu)$ is fixed and if $\hat{x} : \Omega \to X$, $\hat{y} : \Omega \to Y$, $\hat{z} : \Omega \to Z$ are random variables (where X, Y, Z are Polish spaces) and $f : X \to W$ is a measurable function, then $f(\hat{x})$, a function of the random variable $\hat{x}$, denotes the random variable whose value at each $\omega \in \Omega$ is $f(\hat{x}(\omega))$. The joint random variable denoted by $(\hat{x}, \hat{y})$ means the random variable whose value at each $\omega \in \Omega$ is $(\hat{x}(\omega), \hat{y}(\omega))$. If $g : X \times Y \to W$ is a measurable function, then $g(\hat{x}, \hat{y})$ also denotes the random variable defined in the obvious way and is a function of the joint random variable $(\hat{x}, \hat{y})$.

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

If X and Y happen to be SFTs, then the standard notations we use to denote the ith coordinate, subblocks and concatenation of words all carry over to random variables. For example, if x is a point in an SFT, we denote the ith coordinate of x as x_i and the restriction of x to the interval $[i, j-1] \cap \mathbb{Z}$ as $x_{[i,j-1]}$. If $\hat{x}$ is a random variable whose values are in an SFT, then similarly, $\hat{x}_i$ and $\hat{x}_{[i,j-1]}$ denote the random variables whose value for each $\omega \in \Omega$ is $\hat{x}(\omega)_i$ and $\hat{x}(\omega)_{[i,j-1]}$ respectively. They are functions of the random variable $\hat{x}$ since $x \mapsto x_i$ and $x \mapsto x_{[i,j-1]}$ are measurable functions. The restriction of $\hat{x}$ to the infinite interval $[i, \infty) \cap \mathbb{Z}$ is denoted as $\hat{x}_{[i,\infty)}$ or $\hat{x}_i^\infty$.

We say $\hat{x}$ determines $\hat{y}$ if the random variable $\hat{y}$ is a function of the random variable $\hat{x}$, or equivalently, if the sigma-algebra on Ω generated by $\hat{x}$ is finer than that of $\hat{y}$. We say $\hat{x}$ and $\hat{y}$ together determine $\hat{z}$ if $\hat{z}$ is a function of the random variable $(\hat{x}, \hat{y})$, or equivalently, if the join of the sigma-algebras generated by $\hat{x}$ and $\hat{y}$ is finer than that of $\hat{z}$.

We also adopt the usual notational shortcuts for events involving random variables. For example, if $\hat{x}$ and $\hat{y}$ are both real-valued random variables, then the notation $[\hat{x} < \hat{y}]$ denotes the event $\{\omega \in \Omega : \hat{x}(\omega) < \hat{y}(\omega)\}$. If $A \subset X$, then $[\hat{x} \in A]$ denotes the event $\{\omega \in \Omega : \hat{x}(\omega) \in A\}$. In particular, commas produce the intersection of events: the notation $[\hat{x} < \hat{y}, \hat{x} \in A]$ denotes $\{\omega : \hat{x}(\omega) < \hat{y}(\omega), \hat{x} \in A\}$ and therefore is just the intersection of the event $[\hat{x} < \hat{y}]$ and the event $[\hat{x} \in A]$.

The notation $\Pr(\cdot)$ will be used to denote the probability of events. For example $\Pr(\hat{x} < \hat{y}, \hat{x} \in A)$ denotes the probability of the event $[\hat{x} < \hat{y}, \hat{x} \in A]$ and is therefore just $\mu(\{\omega \in \Omega : \hat{x}(\omega) < \hat{y}(\omega), \hat{x}(\omega) \in A\})$.

If $\hat{x}$ is a random variable whose values are in X, then the (probability) distribution of the random variable $\hat{x}$ means the measure on X that is the pushforward of the measure μ on Ω under the map $\hat{x} : \Omega \to X$.

In particular, the joint distribution of $\hat{x}$ and $\hat{y}$ means the distribution of the joint random variable $(\hat{x}, \hat{y})$. It is easy to check that the distribution of $f(\hat{x})$ is the pushforward of the distribution of $\hat{x}$ under the map f.

3. *Lemmas for crossings*

Given a positive integer r and $f \in [0, 1]$, we will say a point $t''' \in \{0, 1, 2\}^{\mathbb{Z}}$ is a coding of crossings of length r and frequency f if the point t''' is a (bi-infinite) concatenation of blocks in $\{10^n 20^m : n, m > r\}$ and

$$\lim_n \frac{1}{n} |\{t_i''' = 1 : 0 \leq i < n\}| = f.$$

(The frequency f here is one-sided even though the point t''' is a two-sided sequence. In retrospect, using the two-sided frequency would have been more natural.) When we are given such a point t''', we can find a bi-infinite sequence of integers $\cdots < a_0 < b_0 < a_1 < b_1 < \cdots$ such that $\{i \in \mathbb{Z} : t_i''' = 1\} = \{a_k : k \in \mathbb{Z}\}$ and $\{i \in \mathbb{Z} : t_i''' = 2\} = \{b_k : k \in \mathbb{Z}\}$.

Given an SFT X and a positive integer r and $f \in [0, 1]$, we say $(z, z') \in X^2$ is obtained from $(x, x', t''') \in X \times X \times \{0, 1, 2\}^{\mathbb{Z}}$ with crossings of length r and frequency r if t''' is a coding of crossings of length r and frequency f and $(z, z')_i = (x, x')_i$ for all $a_k + r \leq i < b_k$ for all $k \in \mathbb{Z}$ and $(z, z')_i = (x', x)_i$ for all $b_k + r \leq i < a_k$ for all $k \in \mathbb{Z}$.

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

LEMMA 3.1. *Given a positive integer r and $V \in \mathrm{Bow}(X)$, there is a constant $C = C(r, V) > 0$ such that, for all (x, x', z, z', t''', f) for which $(z, z') \in X^2$ is obtained from $(x, x', t''') \in X \times X \times \{0, 1, 2\}^{\mathbb{Z}}$ with crossings of length r and frequency r, we have*

$$\limsup_{n\to\infty} \frac{1}{n}\left|\sum_{k=0}^{n-1} V \circ \sigma^k(z) + V \circ \sigma^k(z') - V \circ \sigma^k(x) - V \circ \sigma^k(x')\right| \le Cf.$$

Proof. We will proceed using a similar argument to the one in [**3**].

Define intervals of integers $A_k = [a_k + r, b_k - 1] \cap \mathbb{Z}$ and $B_k = [b_k + r, a_k - 1] \cap \mathbb{Z}$. Let $A = \bigcup_k A_k$ and $B = \bigcup_k B_k$. The remainder $R = \mathbb{Z} - (A \cup B)$ is a union of disjoint intervals of length r. The point z follows x along A and follows x' along B. The point z' follows x, x' the other way around.

The sum $|\sum_{k=0}^{n-1} V \circ \sigma^k(z) + V \circ \sigma^k(z') - V \circ \sigma^k(x) - V \circ \sigma^k(x')|$ is bounded by

$$D_n(A, z, x) + D_n(B, z', x) + D_n(A, z', x') + D_n(B, z, x') + 4 \cdot |R \cap [n]| \cdot \|V\|_\infty$$

where $D_n(A, z, x) = |S_{A\cap[n]} V(z) - S_{A\cap[n]} V(x)|$.

Let $C(V) := \sup_I \mathrm{var}_I(S_I V) < \infty$. The first term $D_n(A, z, x)$ is bounded by $C(V)$ times the number of consecutive blocks in $A \cap [n]$, which in turn is bounded by a constant times $|R \cap [n]|$. The same argument applies to the other three terms $D_n(\cdots)$. Therefore, the total sum is bounded by a constant times $|R \cap [n]|$. Finally, note that $|R \cap [n]|$ is bounded by a constant (depending on r) times the number of 1s in $t'''_{[0,n)}$. Now, taking the average over n and taking the limit lead to the desired inequality. □

LEMMA 3.2. *Let $r \in \mathbb{N}$ and $V \in \mathrm{Bow}(X)$. Let $\hat{x}, \hat{x}', \hat{z}, \hat{z}', \hat{t}'''$ and $\hat{f}$ be random variables such that the probability distributions of $\hat{x}, \hat{x}', \hat{z}, \hat{z}'$ are invariant measures $\mu_1, \mu_2, \mu_1', \mu_2'$ on X and the probability distribution of $\hat{t}'''$ is an invariant measure on $\{0, 1, 2\}^{\mathbb{Z}}$ and $(\hat{z}, \hat{z}') \in X^2$ is obtained from $(\hat{x}, \hat{x}', \hat{t}''') \in X \times X \times \{0, 1, 2\}^{\mathbb{Z}}$ with crossings of length r and frequency $\hat{r}$. Then*

$$|\mu_1'(V) + \mu_2'(V) - \mu_1(V) - \mu_2(V)| \le C \cdot \Pr[\hat{t}'''_0 = 1]$$

where C is the same constant from the previous lemma. In particular, C does not depend on the distributions of $\hat{x}, \hat{x}', \hat{z}, \hat{z}', \hat{t}''', \hat{f}$.

Proof. Recall that for each invariant measure μ on a shift space, $\mu(V) = \int \lim_n (1/n) S_n V x \, d\mu(x)$ which follows by the ergodic theorem and ergodic decomposition. Using $\mathbf{E}(\cdot)$ to denote the expectation value of real-valued random variables, we have

$$\begin{aligned}
&|\mu_1'(V) + \mu_2'(V) - \mu_1(V) - \mu_2(V)| \\
&\quad = |\mathbf{E}(V(\hat{z}) + V(\hat{z}') - V(\hat{x}) - V(\hat{x}'))| \\
&\quad = \left|\mathbf{E}\left(\lim_n \frac{1}{n} S_n V \hat{z} + \lim_n \frac{1}{n} S_n V \hat{z}' - \lim_n \frac{1}{n} S_n V \hat{x} - \lim_n \frac{1}{n} S_n V \hat{x}'\right)\right| \\
&\quad \le \mathbf{E}\left(\limsup_n \frac{1}{n} |S_n V \hat{z} + S_n V \hat{z}' - S_n V \hat{x} - S_n V \hat{x}'|\right) \\
&\quad \le C \cdot \mathbf{E}(\hat{f}) = C \cdot \mathbf{E}\left(\lim_n \frac{1}{n} |\{\hat{t}'''_i = 1 : 0 \le i < n\}|\right) = C \cdot \Pr[\hat{t}'''_0 = 1]. \qquad \square
\end{aligned}$$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

4. *Class degree of factor codes and ergodic measures*

In addition to the class degree of a factor code, we also use the definitions of transition blocks and minimal transition blocks as in [**2**].

Definition 4.1. Let (X, Y, π) be a one-block one-step factor triple. Let $w = w_{[0,p]}$ be a block in Y. Also let n be an integer in $(0, p)$ and M be a subset of $\pi^{-1}(w_n)$. We say a block $u \in \pi^{-1}(w)$ is *routable through* $a \in M$ *at time* n if there is a block $\bar{u} \in \pi^{-1}(w)$ with $\bar{u}_0 = u_0$, $\bar{u}_p = u_p$ and $\bar{u}_n = a$. A triple (w, n, M) is called a *transition block* of π if every block in $\pi^{-1}(w)$ 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). When there is no confusion, for example when $y \in Y$ and $w = y_{[i,i+p]}$ are fixed, we say the points $x, \bar{x} \in \pi^{-1}(y)$ are routable through $a \in M$ *at time* $i + n$ if $x_{[i,i+p]}$ and $\bar{x}_{[i,i+p]}$ are routable through *a at time* $i + n$.

Definition 4.2. Let

$$c^*_\pi = \min\{|M| : (w, n, M) \text{ is a transition block of } \pi\}.$$

A *minimal transition block* of π is a transition block of depth c^*_π.

A slightly modified definition of a minimal transition block, which is appropriate when dealing with measures without full support on Y, is also given.

Definition 4.3. Given a 1-step 1-block factor triple (X, Y, π) and an ergodic measure ν on Y we say (w, n, M) is a ν-*minimal transition block* if it is a transition block with $\nu(w) > 0$ and if it has the smallest depth among all such transition blocks. If ν has full support, then the ν-minimal transition blocks are exactly the minimal transition blocks.

Given a factor triple (X, Y, π) and an ergodic measure ν on Y, we define the class degree of ν to be a positive integer defined by the following result [**2**].

THEOREM 4.1. *Let (X, Y, π) be a factor triple and ν an ergodic measure on Y. Then ν-almost every point of Y has the same number of transition classes over it. We call this number the class degree of ν and denote it by $c_{\pi,\nu}$ or c_ν. If (X, Y, π) is a 1-step 1-block factor triple, then this number is equal to the depth of any ν-minimal transition block. If ν is fully supported, then this number is equal to the class degree of π, denoted by c_π.*

The proof of this theorem does not use the assumption that Y is irreducible.

If ν is not fully supported, the class degree of the measure and that of the factor map can be different and there are examples for both $c_\nu < c_\pi$ and $c_\nu > c_\pi$. An easy way to build such an example is to look in the class of finite-to-one factor maps and let ν be the measure concentrated on a periodic orbit. For example, if π is a degree-one factor map from an irreducible SFT X onto a sofic shift Y so that $c_\pi = 1$, and if Y has a fixed point $y^* \in Y$ and $\nu = \delta_{y^*}$, then all preimages of y^* are periodic points. In this case, $c_\nu = |\pi^{-1}(y^*)|$ holds. Since the condition $|\pi^{-1}(y^*)| > 1$ can be easily arranged, we can get $c_\nu > c_\pi$.

A stronger condition can also be arranged: the number of ergodic measures of maximal relative entropy over ν (same as the number of ergodic measures over ν since π is finite-to-one) can be bigger than c_π. A general way to arrange this condition with a non-atomic ν can be found in [**13**] (its last section).

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

5. *Measure theoretic properties of transition classes*

We establish a measure theoretic analogue of a result in [**1**]. We remark that the following theorems hold even when ν does not have full support.

THEOREM 5.1. *Let (X, Y, π) be a 1-step 1-block factor triple and ν an ergodic measure on Y. Let μ be an invariant measure on X that projects to ν. Let (w, n, M) be a ν-minimal transition block. Let u be an X-block with $\mu(u) > 0$ that projects to w. Then u is routable through a unique symbol in M at time n.*

Proof. Since w is a transition block, u is routable through at least one member of M at time n. To show that u is routable through at most one member of M at time n, we suppose to the contrary that u is routable through two distinct members $a^{(1)}$ and $a^{(2)}$ of $M = \{a^{(1)}, a^{(2)}, \ldots, a^{(d)}\}$ where $d \geq 2$ is the size of M. By Theorem 4.1, d is equal to the class degree of ν.

By Poincare's recurrence theorem, for μ-almost every point x in the cylinder $[u] \subset X$, the block u occurs infinitely many times to the right in x. And for μ-almost every point $x \in X$, the point $\pi(x)$ has exactly d transition classes over it. Therefore there exists a point $x \in X$ such that u occurs infinitely many times to the right in x and that $\pi(x)$ has exactly d transition classes over it. Fix such a point x. Fix $d-1$ points $x^{(1)}, x^{(2)}, \ldots, x^{(d-1)} \in X$ such that the d points $x, x^{(1)}, x^{(2)}, \ldots, x^{(d-1)}$ are in different transition classes over $\pi(x)$.

Since u occurs infinitely many times to the right in x, we can choose positions $\{[i_j, i_j + |w| - 1]\}_{j\geq 1}$ such that $i_{j+1} > i_j + |w|$ and $x_{[i_j, i_j+|w|-1]} = u$.

For each k and j, the block $x^{(k)}_{[i_j, i_j+|w|-1]}$ projects to w and hence is routable through a symbol in M at time n.

If there is $x^{(k)}$ such that $x^{(k)}_{[i_j, i_j+|w|-1]}$ is routable through $a^{(1)}$ or $a^{(2)}$ at time n for infinitely many j, then the point $x^{(k)}$ is in the same transition class as x, which gives a contradiction.

Therefore there is $J \geq 1$ such that for each $j \geq J$ and for each $x^{(k)}$, the block $x^{(k)}_{[i_j, i_j+|w|-1]}$ is routable through a symbol, say $a^{(q(k,j))}$, in $M \setminus \{a^{(1)}, a^{(2)}\}$ at time n. By the pigeonhole principle, for each $j \geq J$, there are distinct k'_j, k''_j such that $q(k'_j, j) = q(k''_j, j)$. By applying the pigeonhole principle again, there are two distinct points $x^{(k')}$, $x^{(k'')}$ among the $d-1$ points such that $k' = k'_j$ and $k'' = k''_j$ for infinitely many $j \geq J$. The two points have the property that for infinitely many j, the blocks $x^{(k')}_{[i_j, i_j+|w|-1]}$ and $x^{(k'')}_{[i_j, i_j+|w|-1]}$ are routable through a common symbol at time n. This forces the two points to be in the same transition class, which gives a contradiction. □

We now introduce the notion of relative joining, of which relatively independent joining is an example. Given a factor triple (X, Y, π), an invariant measure λ on the product X^2 is called a (2-fold) *relative joining* if for λ-almost every (x, x') we have $\pi(x) = \pi(x')$. Let $p_1 : X^2 \to X$ (respectively $p_2 : X^2 \to X$) be the projection onto the first coordinate (respectively the second coordinate). Given a relative joining λ on X^2, if μ_1 and μ_2 are invariant measures on X such that $\mu_1 = p_1(\lambda)$ and $\mu_2 = p_2(\lambda)$, then we say λ is a relative joining of μ_1 and μ_2. Given a relative joining λ on X^2, if ν is an invariant measure on Y such that $\nu = \pi \circ p_1(\lambda)$, then we say λ is a relative joining over ν.

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

If λ is a relative joining of μ_1 and μ_2 over ν, then $\pi(\mu_1) = \nu = \pi(\mu_2)$. Conversely, if μ_1, μ_2 are invariant measures on X and if ν is an invariant measure on Y such that $\pi(\mu_1) = \nu = \pi(\mu_2)$, then there exists at least one relative joining of μ_1 and μ_2 over ν, namely, the relatively independent joining.

We define and compare three subsets of X^2 given a 1-step 1-block factor triple (X, Y, π). Let D_1 be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that x, x' are in the same transition class over $\pi(x)$. We call this set the *class diagonal* of the factor triple (X, Y, π). Let D_2 be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that the points x, x' as bi-infinite words are routable through a common symbol at a common time in the sense that there exist $n \in \mathbb{Z}$ and points $\bar{x}, \bar{x}' \in \pi^{-1}\pi(x)$ such that $\bar{x}$ and $\bar{x}'$ are bi-asymptotic† to x and x' respectively and $\bar{x}_n = \bar{x}'_n$ holds. Let D_3 be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that there is a point $z \in \pi^{-1}\pi(x)$ that is left asymptotic to x and right asymptotic to x' and a point $z' \in \pi^{-1}\pi(x)$ that is left asymptotic to x' and right asymptotic to x. The three sets D_1, D_2, D_3 are invariant Borel-measurable subsets of X^2 and we have $D_1 \subset D_2 \subset D_3$.

THEOREM 5.2. *Given a* 1*-step* 1*-block factor triple* (X, Y, π) *and a relative joining* λ *on* X^2*, we have* $D_1 = D_2 = D_3 \pmod{\lambda}$.

Proof. It is enough to show that $D_3 \subset D_1 \pmod{\lambda}$.

Let C be the set of pairs (u, v) of X-blocks such that $\pi(u) = \pi(v)$ and that there is an X-block $w \in \pi^{-1}\pi(u)$ that starts with the same symbol as u and ends with the same symbol as v and a X-block $w' \in \pi^{-1}\pi(u)$ that starts with the same symbol as v and ends with the same symbol as u.

For each $(u, v) \in C$, let $D_{(u,v)}$ be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that the X^2-block (u, v) occurs in (x, x'). Then D_3 is the union of $D_{(u,v)}$.

For each $(u, v) \in C$, let $D'_{(u,v)}$ be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that the X^2-block (u, v) occurs infinitely many times to the right in (x, x'). By Poincare's recurrence theorem, $D_{(u,v)} = D'_{(u,v)} \pmod{\lambda}$.

It is easy to check that each $D'_{(u,v)}$ is a subset of D_1. □

A relative joining λ on X^2 is called a *class diagonal joining* if for λ-almost every (x, x') the two points x, x' are in the same transition class over the point $\pi(x) = \pi(x')$.

The following theorem is a measure theoretic analogue of another result in [**1**].

THEOREM 5.3. *Let* (X, Y, π) *be a* 1*-step* 1*-block factor triple and* ν *an ergodic measure on* Y*. Let* λ *be a class diagonal joining on* X^2 *over* ν*. Let* (w, n, M) *be a* ν*-minimal transition block. Let* u, v *be* X*-blocks that project to* w *such that* $\lambda([u] \times [v]) > 0$*. Then the two blocks* u, v *are routable through a common symbol in* M *at time* n*.*

Proof. Since w is a transition block, u is routable through a symbol in M, say a, at time n. Similarly, v is routable through a symbol in M, say b, at time n. It is enough to show that $a = b$. Suppose $a \neq b$.

Let C be the set as defined in the proof of Theorem 5.2.

† Left asymptotic and right asymptotic.

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

For each $(u'', v'') \in C$, let $D_{(u'',v'')}$ be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that the X^2-block (u'', v'') occurs in (x, x'). Then D_3 is the union of $D_{(u'',v'')}$ when (u'', v'') runs over the elements of C.

For each $(u'', v'') \in C$, let $D''_{(u'',v'')}$ be the set of $(x, x') \in X^2$ such that $\pi(x) = \pi(x')$ and that the X^2-block (u'', v'') occurs infinitely many times both in $(x, x')_{[0,\infty)}$ and in $(x, x')_{(-\infty,0]}$. By Poincare's recurrence theorem, $D_{(u'',v'')} = D''_{(u'',v'')}$ (mod λ).

Since $\lambda(D_1) = 1$, we have $\lambda(D_3) = 1$, but since $D_3 = \bigcup_{(u'',v'')\in C} D''_{(u'',v'')}$ (mod λ), there is $(u'', v'') \in C$ such that $\lambda([u] \times [v] \cap D''_{(u'',v'')}) > 0$. Fix such $(u'', v'') \in C$.

For each $(x, x') \in [u] \times [v] \cap D''_{(u'',v'')}$, the X^2-block (u'', v'') occurs in $(x, x')_{[|w|,\infty)}$ and in $(x, x')_{(-\infty,-1]}$ while (u, v) occurs between them. Therefore, there is $(\bar{u}, \bar{v})$ with $\lambda([\bar{u}] \times [\bar{v}]) > 0$ such that (u'', v'') occurs at the beginning and at the end of $(\bar{u}, \bar{v})$ and that (u, v) occurs at a position, say $[i, i + |w| - 1]$, between them in $(\bar{u}, \bar{v})$.

Since $\lambda([\bar{u}] \times [\bar{v}]) > 0$, we have $\pi(\bar{u}) = \pi(\bar{v})$ and $\mu(\bar{u}) > 0$ where $\mu = p_1(\lambda)$. Let $\bar{w} = \pi(\bar{u})$. Since $\bar{w}$ contains w and $\nu(\bar{w}) > 0$, we can conclude that $(\bar{w}, i + n, M)$ is another ν-minimal transition block.

The block $\bar{u}$ is routable through the symbol $a \in M$ at time $i + n$. Because $(u'', v'') \in C$ occurs at the beginning and at the end of $(\bar{u}, \bar{v})$, the block $\bar{u}$ is routable through also $b \in M$ at time $i + n$. This contradicts Theorem 5.1. □

We have the following pointwise statement.

COROLLARY 5.1. *Let (X, Y, π) be a 1-step 1-block factor triple and ν an ergodic measure on Y. Let λ be a class diagonal joining on X^2 over ν. Let (w, n, M) be a ν-minimal transition block. For λ-almost every (x, x'), we have that for each i with $\pi(x)_{[i,i+|w|-1]} = w$, the two blocks $x_{[i,i+|w|-1]}$ and $x'_{[i,i+|w|-1]}$ are routable through a common symbol in M at time n.*

Proof. For λ-almost every (x, x'), all X^2 blocks (u, v) occurring in (x, x') satisfy $\lambda([u] \times [v]) > 0$. □

The corollary enables us to route between two paths x and x' upon any observation of w in the image $\pi(x)$. In the proof of our main theorem, it turns out to be crucial to base our routing decisions on observations made in $\pi(x)$ rather than in x, x'. If we naively base routing decisions on observations in x, x' instead, then we do not know how to obtain good control on entropy rates.

6. *Relative entropy*

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a measurable finite partition $\mathcal{C}$, a sub-σ-algebra $\mathcal{D} \subset \mathcal{F}$ and an event $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, we denote by $H(\mathcal{C} \mid \mathcal{D} \mid A)$ the conditional entropy of $\mathcal{C}$ given $\mathcal{D}$ with respect to the conditional measure on A. Given a discrete random variable $\hat{x}$, a random variable $\hat{y}$ on Ω, and an event $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, we denote by $H(\hat{x} \mid \hat{y} \mid A)$ the conditional entropy of $\hat{x}$ given $\hat{y}$ with respect to the conditional measure on A.

We adopt some notational practices (from information theory literature) of interpreting commas in the $H(\cdot)$ entropy notation. If $\hat{x} : \Omega \to X$, $\hat{y} : \Omega \to Y$ are discrete random variables and $\hat{z} : \Omega \to Z$, $\hat{w} : \Omega \to W$ are random variables, then the notation $H(\hat{x}, \hat{y} \mid \hat{z}, \hat{w})$ is interpreted to be the conditional entropy of the joint random variable $(\hat{x}, \hat{y})$ given the joint random variable $(\hat{z}, \hat{w})$. The event part in the $H(\cdot)$ notation may refer to events without specifying square brackets. For example, if $\hat{z} : \Omega \to Z$ is a discrete random variable and z is some element of Z, then the notation $H(\hat{x} \mid \hat{y} \mid \hat{z} = z)$ means $H(\hat{x} \mid \hat{y} \mid [\hat{z} = z])$ where we recall that the event $[\hat{z} = z]$ is the set $\{w \in \Omega : \hat{z}(\omega) = z\}$. Similarly, if X' is some (measurable) subset of X, then $H(\hat{x} \mid \hat{y} \mid \hat{x} \in X', \hat{z} = z)$ means $H(\hat{x} \mid \hat{y} \mid [\hat{x} \in X', \hat{z} = z])$.

With three discrete random variables $\hat{x}$, $\hat{y}$, $\hat{z}$ and a positive event A, we have

$$H(\hat{x} \mid \hat{y}, \hat{z} \mid A) := H(\hat{x} \mid (\hat{y}, \hat{z}) \mid A) = \sum_z \mathbb{P}(\hat{z} = z \mid A) H(\hat{x} \mid \hat{y} \mid [\hat{z} = z] \cap A),$$

where z runs over values in the range of $\hat{z}$. (This follows easily by first considering the special case $A = \Omega$.) This equation holds even when $\hat{y} : \Omega \to Y$ is not discrete (i.e. if Y is infinite). The case where Y is infinite reduces to the discrete case because the σ-algebra generated by a non-discrete $\hat{y}$ can be written as the join of an increasing sequence of finite partitions on Ω.

If A is an event measurable with respect to $\hat{y}$, then

$$\begin{aligned} H(\hat{x} \mid \hat{y}) &= H(\hat{x} \mid \hat{y}, 1_A) \\ &= \mathbb{P}(A) H(\hat{x} \mid \hat{y} \mid A) + \mathbb{P}(A^c) H(\hat{x} \mid \hat{y} \mid A^c), \end{aligned}$$

where 1_A is the indicator function of A. If A is an event that is not measurable with respect to $\hat{y}$, then only the second equality from above is guaranteed.

The following two lemmas bound the entropy increase when we can determine the value of a discrete random variable from another random variable within a small probability of error.

LEMMA 6.1. *Let $\hat{x}$ be a discrete random variable and E be an event that is measurable with respect to a (not necessarily discrete) random variable $\hat{y}$. Suppose there are $K + 1$ Borel-measurable functions $f_0, \ldots, f_K$ such that $\hat{x} = f_0(\hat{y})$ holds a.s. on the event E^c, and that $\hat{x} \in \{f_1(\hat{y}), \ldots, f_K(\hat{y})\}$ holds a.s. on the event E. Then*

$$H(\hat{x} \mid \hat{y}) \leq \Pr(E) \log K.$$

Proof. Given that E is measurable with respect to $\hat{y}$ we have

$$\begin{aligned} H(\hat{x} \mid \hat{y}) &= H(\hat{x} \mid \hat{y}, 1_E) \\ &= \Pr(E) H(\hat{x} \mid \hat{y} \mid E) + \Pr(E^c) H(\hat{x} \mid \hat{y} \mid E^c) \\ &\leq \Pr(E) \log K + \Pr(E^c) \cdot 0. \end{aligned}$$ □

For each $0 \leq p \leq 1$, denote $H_p = -p \log p - (1 - p) \log(1 - p) \geq 0$ (where its value at $p = 0$ or $p = 1$ is defined to be 0). For small p, H_p is on the order of $-p \log p$: we have some constant $C > 1$ such that $C^{-1} \leq H_p/(-p \log p) \leq C$ holds for all $0 \leq p \leq \frac{1}{2}$.

LEMMA 6.2. *Let $\hat{x}$ be a discrete random variable and E be an event. Let $\hat{y}$ be a random variable. Suppose there are $K + 1$ Borel-measurable functions $f_0, \ldots, f_K$ such that*

$\hat{x} = f_0(\hat{y})$ holds a.s. on the event E^c and that $\hat{x} \in \{f_1(\hat{y}), \ldots, f_K(\hat{y})\}$ holds a.s. on the event E. Then

$$H(\hat{x} \mid \hat{y}) \leq \Pr(E) \log K + H_{\Pr(E)}.$$

Proof.

$$\begin{aligned} H(\hat{x} \mid \hat{y}) &\leq H(\hat{x}, 1_E \mid \hat{y}) \\ &= H(\hat{x} \mid \hat{y}, 1_E) + H(1_E \mid \hat{y}) \\ &\leq H(\hat{x} \mid \hat{y}, 1_E) + H(1_E). \end{aligned}$$

Now use Lemma 6.2 and the fact that $H(1_E) = H_{\Pr(E)}$. □

Lemma 6.1 can be used when we have knowledge of when errors in guessing are made, but Lemma 6.2 is available even in situations where we can't know when errors are made.

Given finite partitions α, β on a measure-theoretic dynamical system of finite entropy, the following quantities are all equal, where $\alpha_i^j = \bigvee_{k=i}^j T^{-k}\alpha$ (T is the measure preserving map associated with the dynamical system):

- $H(\alpha \mid \alpha_1^\infty \vee \beta_{-\infty}^\infty)$;
- $\lim_n H(\alpha \mid \alpha_1^n \vee \beta_{-\infty}^\infty)$;
- $\lim_n (1/n) H(\alpha_0^{n-1} \mid \beta_{-\infty}^\infty)$;
- the metric entropy of the factor system $\alpha_{-\infty}^\infty \vee \beta_{-\infty}^\infty$ minus the metric entropy of the factor system $\beta_{-\infty}^\infty$.

The equality between the first and the last is Abramov–Rokhlin formula (see for example [**5**]). This quantity is called the *relative entropy* of the system $\alpha_{-\infty}^\infty \vee \beta_{-\infty}^\infty$ relative to its factor $\beta_{-\infty}^\infty$.

We denote by $h(\alpha_{-\infty}^\infty \mid \beta_{-\infty}^\infty)$ this quantity. If a random variable $\hat{x}$ (respectively $\hat{y}$) generates the σ-algebra $\alpha_{-\infty}^\infty$ (respectively $\beta_{-\infty}^\infty$) for some finite partition α (respectively β), then we write $h(\hat{x} \mid \hat{y}) = h(\alpha_{-\infty}^\infty \mid \beta_{-\infty}^\infty)$.

We have the following subadditive property of relative entropy

$$\begin{aligned} h(\alpha_{-\infty}^\infty \mid \gamma_{-\infty}^\infty) &\leq h(\alpha_{-\infty}^\infty \vee \beta_{-\infty}^\infty \mid \gamma_{-\infty}^\infty) \\ &= h(\alpha_{-\infty}^\infty \mid \beta_{-\infty}^\infty \vee \gamma_{-\infty}^\infty) + h(\beta_{-\infty}^\infty \mid \gamma_{-\infty}^\infty). \end{aligned}$$

7. *Jump extension*

We introduce the notion of jump extension to formalize the idea of advancing a random process (such as the process of tossing coins) each time a specific word occurs in another random process.

Throughout this section, let μ be an invariant measure on a subshift X, and A a spanning subset of X with respect to μ. In other words,

$$\mu\left(\bigcup_{i \in \mathbb{Z}} \sigma^i(A)\right) = 1$$

and hence by Poincare's recurrence theorem

$$\mu\{x \in X : \sigma^i x \in A \text{ for bi-infinitely many } i\} = 1.$$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

If A is the cylinder set of a word, then we might be interested in tossing coins every time that word occurs from the process (X, μ).

Throughout this section, let η be an invariant measure on $C^{\mathbb{Z}}$ (where C is some finite set) and assume $0 \notin C$. (The symbol 0 will be used to fill in between jumps.) The measure η is the process to advance every time A occurs. Let D be the disjoint union of C and $\{0\}$. Then there is an extension $(X \times D^{\mathbb{Z}}, \bar{\mu}, \sigma)$ of the system (X, μ, σ) with the following properties.

- $\bar{\mu}$ is an invariant measure on $X \times D^{\mathbb{Z}}$ that projects to μ (this property is just another way of saying that $(X \times D^{\mathbb{Z}}, \bar{\mu}, \sigma)$ is an extension).
- For $\bar{\mu}$-almost every (x, t), for all $i \in \mathbb{Z}$, $\sigma^i x \in A$ if and only if $t_i \neq 0$.
- If q is a measurable function from X to $\mathbb{Z}$ such that $\sigma^{q(x)}(x) \in A$ holds for μ-almost every x, then $g_q(\bar{\mu}) = \mu \times \eta$ where g_q is the $\bar{\mu}$-almost everywhere defined measurable function from $X \times D^{\mathbb{Z}}$ to $X \times C^{\mathbb{Z}}$ given by $g_q(x, t) = (x, (t_{q_k(x)})_k)$, where
$$\cdots < q_{-1}(x) < q_0(x) = q(x) < q_1(x) < q_2(x) < \cdots$$
are all the coordinates i for which $\sigma^i(x) \in A$. (We do not require q to be the first visit time $q^*(x) := \min\{i \geq 0 : \sigma^i(x) \in A\}$, only that it is a visit time. Since $g_q(\bar{\mu}) = \mu \times \eta$ holds if and only if $g_{q^*}(\bar{\mu}) = \mu \times \eta$ holds†, the difference between using an arbitrary visit time q and using the first visit time q^* is only a technical difference that does not change the meaning of jump extensions.)

We call the extension $(X \times D^{\mathbb{Z}}, \bar{\mu}, \sigma)$ (or just the measure $\bar{\mu}$) the *jump extension* of (X, μ, σ) with respect to A and η.

THEOREM 7.1. *The entropy of the jump extension is*
$$h(\bar{\mu}) = h(\mu) + \mu(A)h(\eta).$$

Proof. We begin by constructing the jump extension, then prove that its entropy satisfies Theorem 7.1. Let (A, σ_A, μ_A) be the induced subsystem of first-returns to A. Form the product of this induced transformation with $C^{\mathbb{Z}}$. Now we perform a kind of skyscraper construction to 'uninduce' this product.

Let $X' = \{(x, s, k) \in A \times C^{\mathbb{Z}} \times \mathbb{Z}^+ : 0 \leq k \leq \tau_A(x)\}$ where $\tau_A(x)$ is the first return time to A and we identify $(x, s, \tau_A(x))$ with $(\sigma_A(x), \sigma(s), 0)$. The measure on this space will be the unique measure satisfying $\mu'(B \times E \times \{k\}) = (\mu \times \eta)(B \times E)$. The transformation will be given by
$$T(x, s, k) = \begin{cases} (x, s, k+1), & 0 \leq k < \tau_A(x), \\ (\sigma_A(x), \sigma(s), 1), & k = \tau_A(x). \end{cases}$$
Now we map X' onto the space $X \times D^{\mathbb{Z}}$ (where $D = C \cup \{0\}$ as described above). Let $\phi : X' \to X \times D^{\mathbb{Z}}$ be the shift-commuting map defined by
$$\phi(x, s, k)_0 = \begin{cases} (x_0, s_0), & k = 0, \\ (x_k, 0), & 0 < k < \tau_A(x), \\ (x_{\tau_A(x)}, s_1), & k = \tau_A(x). \end{cases}$$

† We do not use this equivalence in this paper, but it should be clear from the skew-product construction of $\bar{\mu}$ that they are equivalent.

We will show that $\bar{\mu} = \phi\mu'$ possesses the properties claimed above. Letting $\pi : X \times D^{\mathbb{Z}} \to X$ be projection onto the first coordinate, it is straightforward to see that $\bar{\mu}$ projects to μ. It is also clear by the construction that for $\bar{\mu}$-almost every $(x, t) \in X \times D^{\mathbb{Z}}$, $t_i \neq 0$ if and only if $\sigma^i x \in A$. Finally, by the definition of μ' and the fact that ϕ is a measure-conjugacy, we see that deleting the occurrences of 0 from t returns a point with distribution η.

Let $\hat{x}$ and $\hat{t}$ be random variables such that the probability distribution of $(\hat{x}, \hat{t})$ is $\bar{\mu}$. Note that $\pi(\hat{x}, \hat{t}) = \hat{x}$. In order to obtain the entropy of $\bar{\mu}$, we will use the Abramov–Rokhlin formula. Applying this to the factor π gives

$$\begin{aligned} h(\bar{\mu}) &= h(\mu) + H((\hat{x}, \hat{t})_0 \mid (\hat{x}, \hat{t})_1^{\infty}, \hat{x}) \\ &= h(\mu) + H(\hat{t}_0 \mid \hat{t}_1^{\infty}, \hat{x}). \end{aligned}$$

The last equality holds because $\hat{x}$ determines $\hat{x}_0$ and $\hat{x}_1^{\infty}$. Call the second term in this sum H_0.

Conditioning H_0 on the event $[\hat{x} \in A]$, whose probability is $\mu(A)$, we obtain

$$H_0 = \mu(A) H(\hat{t}_0 \mid \hat{t}_1^{\infty}, \hat{x} \mid \hat{x} \in A) + \mu(A^c) H(\hat{t}_0 \mid \hat{t}_1^{\infty}, \hat{x} \mid \hat{x} \in A^c).$$

The second term in this sum is zero because $\hat{t}_0$ is known to be 0 (hence deterministic) when $\hat{x} \in A^c$. It only remains to show that the first term of this sum is $\mu(A)h(\eta)$.

Define $\hat{s}$ to be the random variable defined by requiring $(\hat{x}, \hat{s}) = g_q(\hat{x}, \hat{t})$ where $q(x)$ is the first visit time of x into A, i.e. $q(x)$ is the smallest $i \geq 0$ such that $\sigma^i x \in A$. Since $\hat{s}_0 = \hat{t}_0$ on $[\hat{x} \in A]$ and $\hat{s}_1^{\infty}$ and $\hat{t}_1^{\infty}$ determine each other on the same event $[\hat{x} \in A]$, we have

$$H(\hat{t}_0 \mid \hat{t}_1^{\infty}, \hat{x} \mid \hat{x} \in A) = H(\hat{s}_0 \mid \hat{s}_1^{\infty}, \hat{x} \mid \hat{x} \in A).$$

The random variables $(\hat{s}_0, \hat{s}_1^{\infty})$ and $\hat{x}$ are independent because the probability distribution of $(\hat{x}, \hat{s})$ is the product measure $\mu \times \eta$. It is straightforward to see that this independence property implies the following equality (see [**12**, Lemma 2.3(2)]).

$$H(\hat{s}_0 \mid \hat{s}_1^{\infty}, \hat{x} \mid \hat{x} \in A) = H(\hat{s}_0 \mid \hat{s}_1^{\infty}).$$

But that is just $h(\eta)$. Therefore we have shown $H_0 = \mu(A)h(\eta)$ and the proof is done. □

LEMMA 7.1. *Let C' be a subset of C and let B be a measurable subset of X. Then*

$$\bar{\mu}\{(x, t) : x \in B, t_0 \in C'\} = \mu(B \cap A)\eta([C'])$$

where $[C']$ denotes the cylinder $\{z \in C^{\mathbb{Z}} : z_0 \in C'\}$.

Proof. For μ-almost every x, define $q(x)$ to be the smallest non-negative integer with $\sigma^{q(x)}(x) \in A$. Note that

$$\{(x, t) : t_0 \neq 0\} = \{(x, t) : x \in A\} = \{(x, t) : q(x) = 0\} \pmod{\bar{\mu}}.$$

So we can conclude

$$\{(x, t) : x \in B, t_0 \in C'\} = \{(x, t) : g_q(x, t) \in (B \cap A) \times [C']\} \pmod{\bar{\mu}}.$$ □

As a special case, we get the following lemma as a corollary.

LEMMA 7.2. *Let C' be a subset of C. Then*

$$\bar{\mu}\{(x, t) : t_0 \in C'\} = \mu(A)\eta([C']).$$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

8. *Proof of the main theorem*

LEMMA 8.1. *Let (X, Y, π) be a factor triple and ν an ergodic measure on Y. Let $V \in \mathrm{Bow}(X)$. Let λ be a class diagonal joining of distinct ergodic measures μ_1, μ_2 over ν. Then there is another relative joining λ' on X^2 over ν such that*

$$h(\lambda') + \mu_1'(V) + \mu_2'(V) > h(\lambda) + \mu_1(V) + \mu_2(V)$$

where $\mu_1' = p_1(\lambda')$ and $\mu_2' = p_2(\lambda')$.

Furthermore, the new measures μ_1', μ_2' are 'transformations' of the given measures μ_1, μ_2 through class diagonal joinings in the following sense. There exists a class diagonal joining of μ_1 and μ_1' over ν and the same is true for μ_2, μ_2'.

Proof. The first thing we construct is a process to advance every time a fixed minimal transition block occurs. The process corresponds to tossing a biased coin every Nth time.

For each $(N, p) \in \mathbb{N} \times (0, \frac{1}{2})$, we want to construct an invariant measure $\eta_{(N,p)}$ that corresponds to a stationary concatenated-block process by concatenating two blocks 13^{N-1} and 23^{N-1} independently with probabilities $1 - p$ and p and randomizing the start. First define a (non-invariant) measure $\eta^o = \eta^o_{(N,p)}$ on $\{1, 2, 3\}^{\mathbb{Z}}$: for each $i \in \mathbb{Z}$, $\eta^o([1]_{iN}) = 1 - p$, $\eta^o([2]_{iN}) = p$, $\eta^o([3]_{iN}) = 0$ and for each k not a multiple of N, $\eta^o([3]_k) = 1$ and the measure η^o makes each coordinate independent. Define the invariant measure $\eta = \eta_{(N,p)}$ on $\{1, 2, 3\}^{\mathbb{Z}}$ by

$$\eta = \frac{1}{N} \sum_{k=0}^{N-1} \sigma^k(\eta^o).$$

The invariant measure $\eta_{(N,p)}$ satisfies the following properties:

- η-almost every point is concatenation of blocks of length N that are either 13^{N-1} or 23^{N-1};
- its entropy is $h(\eta) = (1/N)H_p$ (we will only use $h(\eta) \geq (1/N)H_p$);
- $\eta(1) = (1 - p)/N$;
- $\eta(2) = p/N$ (hence, $\eta([\{1, 2\}]) \leq \eta(1) + \eta(2) \leq 1/N$);
- $\eta(13^{N-1}2) = (p(1 - p))/N = \eta(23^{N-1}1)$.

With this measure, we can construct the relevant jump extension, perform crossings on it, and obtain a joining λ' with desired properties. We proceed in several steps.

Step 1 (Construction of the jump extension $(\Omega, \bar{\lambda})$ and three random variables $\hat{x}, \hat{x}', \hat{t}$). We may assume (X, Y, π) is a 1-step 1-block factor triple. Let (w, n, M) be a ν-minimal transition block. Let $(N, p) \in \mathbb{N} \times (0, \frac{1}{2})$ be such that $N > |w|$. The exact value of the parameters (N, p) will be determined at the end of the proof after we obtain a good lower bound (a function of (N, p)) for the pressure difference.

Let $\bar{\lambda}$ be the jump extension of λ with respect to $(\pi \circ p_1)^{-1}[w]$ and $\eta_{(N,p)}$. We can form the jump extension because $(\pi \circ p_1)^{-1}[w]$ is spanning: in fact, λ-almost every point visits $(\pi \circ p_1)^{-1}[w]$ with frequency given by $\nu(w) > 0$ because ν is ergodic. Also, $\bar{\lambda}$ is an invariant measure on $\Omega = X^2 \times \{0, 1, 2, 3\}^{\mathbb{Z}}$. The measure-theoretic dynamical system $(\Omega, \bar{\lambda}, \sigma)$ is the ambient probability space on which we will build our random variables.

Let $\hat{x}$, $\hat{x}'$, $\hat{t}$ be random variables defined on $(\Omega, \bar{\lambda})$ by

$$\hat{x}(x, x', t) = x,$$
$$\hat{x}'(x, x', t) = x',$$
$$\hat{t}(x, x', t) = t.$$

Since the distribution of the joint random variable $(\hat{x}, \hat{x}')$ is the relative joining λ, we can define another random variable $\hat{y} = \pi(\hat{x}) = \pi(\hat{x}')$ which has distribution ν. The jump extension ensures that for each i, the event $\hat{t}_i > 0$ is the same as the event $\sigma^i(\hat{y}) \in [w]$. In other words, $\hat{t}$ is a sequence in which non-zero symbols occur exactly where the word w occurs in $\hat{y}$.

Step 2 (Construction of λ' via crossings). We have so far four random variables: $\hat{x}$, $\hat{x}'$, $\hat{t}$, $\hat{y}$. We now want to construct two more random variables $\hat{z}$, $\hat{z}'$ such that $\pi(\hat{z}) = \hat{y} = \pi(\hat{z}')$ and they will be formed by taking some segments from $\hat{x}$, $\hat{x}'$ in some way. We define $\hat{z}$ first. It will be defined in such a way that $\hat{z}$ is a function of $\hat{x}$, $\hat{x}'$, $\hat{t}$. Occurrence of the symbol 1 in $\hat{t}$ will mean: take from the first path, namely, $\hat{x}$. The symbol 2 will mean: take from the second path, namely, $\hat{x}'$. The other symbols 3 and 0 have no meaning.

The point $\hat{z}(x, x', t) \in X$ is defined for $\bar{\lambda}$-almost every (x, x', t) in the following way. Let $\cdots i_{-1} < i_0 < i_1 < \cdots$ be all the places where 1 or 2 occurs in t. (One can think of each i_j to be a integer-valued function defined almost everywhere on Ω if preferred.) Note that $i_{j+1} - i_j \geq N > |w|$ holds for each j (almost everywhere) because if we remove zeros from the block $t_{[i_j, i_{j+1}-1]}$ we would get either 13^{N-1} or 23^{N-1}. This means that we can divide the region $[i_j, i_{j+1} - 1]$ into two subregions $[i_j, i_j + |w| - 1]$ and $[i_j + |w|, i_{j+1} - 1]$.

We define $\hat{z}(x, x', t)$ for the latter type of subregions first. The value of $\hat{z}$ on those subregions are copied from x or x' depending on what t tells at i_j, in other words:

$$\hat{z}(x, x', t)_{[i_j+|w|, i_{j+1}-1]} = \begin{cases} x_{[i_j+|w|, i_{j+1}-1]} & \text{if } t_{i_j} = 1, \\ x'_{[i_j+|w|, i_{j+1}-1]} & \text{if } t_{i_j} = 2. \end{cases}$$

For the former type of subregions, note that for each of such subregion, the block w appears in $\hat{y}(x, x', t)$ at that subregion. Since λ is class diagonal, Corollary 5.1 ensures that for each of these subregions, the two blocks from x, x' at that subregion are routable through a common symbol. Theorem 5.3 ensures that for each X^2-block (u, v) that projects to w such that $\lambda([u] \times [v]) > 0$, one can choose an X-block $r^{12}(u, v)$ that projects to w and starts with the symbol u_0 and ends with the symbol $v_{|w|-1}$. We also choose $r^{21}(u, v)$ that projects to w and starts with the symbol v_0 and ends with the symbol $u_{|w|-1}$. We also define $r^{11}(u, v) = u$ and $r^{22}(u, v) = v$.

Now define $\hat{z}(x, x', t)$ for the former type of subregions by using the functions $r^{11}, r^{12}, r^{21}, r^{22}$ depending on what t is telling at i_{j-1} and i_j, in other words

$$\hat{z}(x, x', t)_{[i_j, i_j+|w|-1]} = r^{t_{i_{j-1}} t_{i_j}}(x_{[i_j, i_j+|w|-1]}, x'_{[i_j, i_j+|w|-1]}).$$

It is easy to check that for $\bar{\lambda}$-almost every (x, x', t), the point $\hat{z}(x, x', t)$ is well defined and is a point in X. As a random variable, one can also check that $\pi(\hat{z}) = \hat{y}$.

Define another random variable $\hat{z}'$ in much the same way as $\hat{z}$ except this time the meaning of the symbols 1 and 2 are swapped: the symbol 1 now means taking from the second path and 2 means taking from the first path. $\hat{z}'$ is in some sense dual to $\hat{z}$. It is easy to check that the joint random variable $(\hat{z}, \hat{z}') : \Omega \to X^2$ as a function is shift-commuting, therefore the distribution of $(\hat{z}, \hat{z}')$ is an invariant measure on X^2, which we denote by λ'. This measure λ' is a relative joining over ν because $\pi(\hat{z}) = \hat{y} = \pi(\hat{z}')$.

Step 3 (Entropy estimates). We have the following four equality or inequalities: the inequality holds because $h(\hat{t}, \hat{x}, \hat{x}' \mid \hat{z}, \hat{z}') = h(\hat{t}, \hat{x}, \hat{x}', \hat{z}, \hat{z}') - h(\hat{z}, \hat{z}')$ and the second-to-last equality holds because it is the entropy of the jump extension:

$$\begin{gathered} h(\lambda') = h(\hat{z}, \hat{z}'), \\ h(\hat{z}, \hat{z}') + h(\hat{t}, \hat{x}, \hat{x}' \mid \hat{z}, \hat{z}') \geq h(\hat{t}, \hat{x}, \hat{x}'), \\ h(\hat{t}, \hat{x}, \hat{x}') = h(\hat{x}, \hat{x}') + \Pr(\hat{t}_0 > 0) h(\eta), \\ h(\hat{x}, \hat{x}') = h(\lambda). \end{gathered}$$

So we can conclude

$$h(\lambda') - h(\lambda) \geq \Pr(\hat{t}_0 > 0) h(\eta) - h_0 = \nu(w) \cdot \frac{H_p}{N} - h_0 \tag{8.1}$$

where

$$h_0 := h(\hat{t}, \hat{x}, \hat{x}' \mid \hat{z}, \hat{z}').$$

We want to bound h_0 from above. We divide it into $h_0 = h_1 + h_2$ where

$$h_1 = h(\hat{t} \mid \hat{z}, \hat{z}')$$

and

$$h_2 = h(\hat{x}, \hat{x}' \mid \hat{t}, \hat{z}, \hat{z}').$$

Step 4 (Bound on h_1). We obtain an upper bound for h_1 first. To do that, we introduce two more random variables $\hat{t}'$ and $\hat{t}''$:

$$\hat{t}'_i = \begin{cases} \hat{t}_i & \text{when } \hat{t}_i = 0, 3, \\ 4 & \text{when } \hat{t}_i = 1, 2. \end{cases}$$

The random variable $\hat{t}'$ captures partial information of $\hat{t}$ by not distinguishing 1 and 2:

$$\hat{t}''_i = \begin{cases} 0 & \text{when } \hat{t}_i = 0, \\ 1 & \text{when } \hat{t}_i > 0. \end{cases}$$

The random variable $\hat{t}''$ captures partial information of $\hat{t}$ that corresponds to where zeros occur in $\hat{t}$ and where non-zeros occur. The following three events are equivalent mod $\bar{\lambda}$:

$$\begin{gathered} \hat{t}''_i = 1, \\ \hat{t}'_i > 0, \\ \sigma^i(\hat{y}) \in [w]. \end{gathered}$$

Note that $\hat{y}$ determines $\hat{t}''$. Also, $\hat{t}$ determines $\hat{t}'$ which in turn determines $\hat{t}''$.

We decompose h_1 into

$$h_1 \leq h(\hat{t}' \mid \hat{z}, \hat{z}') + h(\hat{t} \mid \hat{t}', \hat{z}, \hat{z}').$$

Since $\hat{z}$ determines $\hat{y}$ which in turn determines $\hat{t}''$, we have the following bound for the first term

$$h(\hat{t}' \mid \hat{z}, \hat{z}') \leq h(\hat{t}' \mid \hat{t}'')$$

but there are only N possible values for $\hat{t}'$ given the value of $\hat{t}''$, therefore $h(\hat{t}' \mid \hat{t}'') = 0$ and we have

$$h(\hat{t}' \mid \hat{z}, \hat{z}') \leq h(\hat{t}' \mid \hat{t}'') = 0$$

and so

$$h_1 \leq h(\hat{t} \mid \hat{t}', \hat{z}, \hat{z}').$$

Therefore

$$\begin{aligned}
h_1 &\leq H(\hat{t}_0 \mid \hat{t}_{[1,\infty)}, \hat{t}', \hat{z}, \hat{z}') \\
&\leq H(\hat{t}_0 \mid \hat{t}'_0, \hat{z}, \hat{z}') \\
&= \Pr(\hat{t}'_0 = 4) H(\hat{t}_0 \mid \hat{t}'_0, \hat{z}, \hat{z}' \mid \hat{t}'_0 = 4) \\
&\quad + \Pr(\hat{t}'_0 \neq 4) H(\hat{t}_0 \mid \hat{t}'_0, \hat{z}, \hat{z}' \mid \hat{t}'_0 \neq 4) \\
&\leq \Pr(\hat{t}'_0 = 4) H(\hat{t}_0 \mid \hat{z}, \hat{z}' \mid \hat{t}'_0 = 4) \\
&\quad + \Pr(\hat{t}'_0 \neq 4) H(\hat{t}_0 \mid \hat{t}'_0 \mid \hat{t}'_0 \neq 4) \\
&= \Pr(\hat{t}'_0 = 4) H(\hat{t}_0 \mid \hat{z}, \hat{z}' \mid \hat{t}'_0 = 4)
\end{aligned}$$

where the last equality holds because $H(\hat{t}_0 \mid \hat{t}'_0 \mid \hat{t}'_0 \neq 4) = 0$ which is because $\hat{t}'_0$ determines $\hat{t}_0$ given the event $\hat{t}'_0 \neq 4$.

So we have

$$h_1 \leq \Pr(\hat{t}'_0 = 4) H^*$$

where

$$H^* = H(\hat{t}_0 \mid \hat{z}, \hat{z}' \mid \hat{t}'_0 = 4).$$

We want to obtain an upper bound on H^* which approaches 0 as $N \to 0$ and does not depend on p.

For the convenience of further calculation, we let $J = [|w|, N - 1]$ which depends on N but not on p. This is an interval that will be used to collect long enough fixed-length segments of $\hat{x}, \hat{x}'$ right after occurrences of w. Note that given the event $\hat{t}'_0 = 4$, the value of $(\hat{z}, \hat{z}')_J$ is either $(\hat{x}, \hat{x}')_J$ or $(\hat{x}', \hat{x})_J$ depending on whether $\hat{t}_0$ is 1 or 2. Therefore, given the event $\hat{t}'_0 = 4$ and the event $(\hat{x}, \hat{x}')_J \in G_1 \times G_2$ where G_1 and G_2 are disjoint sets of blocks that we will define later, the value of $(\hat{z}, \hat{z}')_J$ determines the value of $\hat{t}_0$ (by just looking at which one of G_1 and G_2 the block $\hat{z}_J$ belongs to).

To define G_1, G_2, first choose a to be an X-block such that $\mu_1(a) \neq \mu_2(a)$ and let $d = |\mu_1(a) - \mu_2(a)| > 0$. Such a block exists because μ_1 and μ_2 are assumed to be distinct. Let G_1 be the set of all X-blocks b of length $|J| = N - |w|$ such that

$$|D(a \mid b) - \mu_1(a)| < \frac{d}{2}$$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

where $D(a \mid b)$ denotes the frequency of a in b. Similarly, let G_2 to be the set of all X-blocks b of length $|J|$ such that

$$|D(a \mid b) - \mu_2(a)| < \frac{d}{2}.$$

It is clear that the two sets G_1, G_2 are disjoint. By Lemma 6.2 we have

$$\begin{aligned} H^* &\leq H(\hat{t}_0 \mid (\hat{z}, \hat{z}')_J \mid \hat{t}'_0 = 4) \\ &\leq P^* \log 2 + H_{P^*} \end{aligned}$$

where P^* denotes the conditional probability given by

$$P^* = \Pr((\hat{x}, \hat{x}')_J \notin G_1 \times G_2 \mid \hat{t}'_0 = 4).$$

We want to show that P^* is a quantity that goes to 0 as $N \to \infty$ and does not depend on p.

Write

$$P^* = \frac{\Pr((\hat{x}, \hat{x}')_J \notin G_1 \times G_2, \hat{t}'_0 = 4)}{\Pr(\hat{t}'_0 = 4)}$$

and apply Lemmas 7.1 and 7.2 to the numerator and the denominator to get

$$P^* = \frac{\lambda(F_J)}{\nu(w)}$$

where $F_J \subset X^2$ denotes the set of (x, x') such that $(x, x')_J \notin G_1 \times G_2$ and $\pi(x) \in [w]$. The set F_J depends on J which in turn depends on N but the set does not depend on p. It is easy to show, using the mean ergodic theorem applied to ergodic μ_1 and μ_2, that $\lim_N \lambda(F_J) = 0$. Therefore P^* (and hence H^* too) is a quantity that does not depend on p and goes to 0 when $N \to \infty$. Denote H^* by $H^*(N)$ to express its dependency on the parameter N. We showed that

$$h_1 \leq \Pr(\hat{t}'_0 = 4) H^*(N)$$

where $H^*(N)$ is a quantity that does not depend on p and that $\lim_N H^*(N) = 0$.

Since $\Pr(\hat{t}'_0 = 4) = \Pr(\hat{t}_0 \in \{1, 2\}) = \nu(w) \cdot \eta(\{1, 2\})$ holds by Lemma 7.2 and $\eta(\{1, 2\}) \leq 1/N$, we have

$$h_1 \leq \frac{\nu(w)}{N} \cdot H^*(N). \tag{8.2}$$

Step 5 (Bound on h_2). Next we want to obtain an upper bound for

$$h_2 = h(\hat{x}, \hat{x}' \mid \hat{t}, \hat{z}, \hat{z}').$$

For λ-almost every (x, x'), let $q = q(x, x')$ be the smallest non-negative number such that $\sigma^q \pi(x) \in [w]$ and let

$$\cdots < q_{-1} < q_0 = q < q_1 < q_2 < \cdots$$

be all the coordinates i for which $\sigma^i \pi(x) \in [w]$.

Let $\hat{q}_k = q_k(\hat{x}, \hat{x}')$. Each $\hat{q}_k$ is an integer-valued random variable. Using them, define

$$\hat{u} = (\hat{t}_{\hat{q}_k})_{-N \leq k \leq 0}.$$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

The random variable $\hat{u}$ takes values in $\{1, 2, 3\}^{N+1}$ and the probability of the event $\hat{u} = u$ for each block u is given by $\Pr(\hat{u} = u) = \eta(u)$.

Define the two events

$$S_{12} = [\hat{y} \in [w], \hat{u} = 13^{N-1}2],$$
$$S'_{12} = \bigcup_{0 \le k < |w|} \sigma^k(S_{12}).$$

The event S'_{12} represents the event of the coordinate 0 falling to one of the subregions where we used the function r^{12}. Define S_{21} and S'_{21} similarly, with $23^{N-1}1$ in place of $13^{N-1}2$. Note that the four events we just defined are measurable with respect to $\hat{t}$. This allows us to use Lemma 6.1 to say

$$\begin{aligned} h_2 &\le H((\hat{x}, \hat{x}')_0 \mid \hat{t}, \hat{z}, \hat{z}') \\ &\le \Pr(S'_{12} \cup S'_{21}) \log(C_0^2) \end{aligned}$$

where C_0 is the number of letters used in the SFT X.

We want to estimate $\Pr(S'_{12} \cup S'_{21})$ now:

$$\begin{aligned} \Pr(S'_{12}) &\le |w| \Pr(S_{12}) \\ &= |w| \cdot \nu(w) \cdot \eta(13^{N-1}2) \\ &= |w| \cdot \nu(w) \cdot \frac{p(1-p)}{N}. \end{aligned}$$

So we have

$$\Pr(S'_{12} \cup S'_{21}) \le C_1 \cdot \frac{p}{N}$$

where C_1 is some constant depending on w but not on N or p.

We showed

$$h_2 \le C_1 \cdot \log(C_0^2) \cdot \frac{p}{N}. \tag{8.3}$$

Step 6 (Bound on the difference of integrals). It remains to estimate the difference of the integral terms, $|\mu'_1(V) + \mu'_2(V) - \mu_1(V) - \mu_2(V)|$. Let $\hat{t}'''$ be the random variable defined in a shift-commuting way by requiring that $\hat{t}'''_0$ takes value 1 on S_{21}, 2 on S_{12} and 0 otherwise. Then Lemma 3.2 applies to this random variable, with crossings of length $r := |w|$. Using our upper bound for $\Pr(t'''_0 = 1) = \Pr(S_{21})$, this implies that for some constant C_2 which may depend on $|w|$, V but not on p or N,

$$|\mu'_1(V) + \mu'_2(V) - \mu_1(V) - \mu_2(V)| \le C_2\left(\frac{p}{N}\right). \tag{8.4}$$

Step 7 (Proof of pressure increase). We obtained upper bounds for all relevant quantities to estimate

$$\Delta := (h(\lambda') + \mu'_1(V) + \mu'_2(V)) - (h(\lambda) + \mu_1(V) + \mu_2(V))$$

which is greater than or equal to

$$\begin{aligned} &(h(\lambda') - h(\lambda)) - |\mu'_1(V) + \mu'_2(V) - \mu_1(V) - \mu_2(V)| \\ &\ge \left(\nu(w) \cdot \frac{H_p}{N} - h_0\right) - C_2\frac{p}{N} \quad \text{by (8.1) and (8.4)} \\ &\ge \nu(w) \cdot \frac{H_p}{N} - h_1 - h_2 - C_2\frac{p}{N}. \end{aligned}$$

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

Using the upper bounds we obtained for the second and the third term in the last expression (i.e. the upper bounds (8.2) and (8.3) for h_1 and h_2), we gain the following lower bound for Δ:

$$\nu(w) \cdot \frac{H_p}{N} - \frac{\nu(w)}{N} \cdot H^*(N) - C_1 \cdot \log(C_0^2) \cdot \frac{p}{N} - C_2 \frac{p}{N}.$$

By choosing appropriate constants C_3, C_4, C_5 that do not depend on N or p, we have

$$\Delta \geq \frac{C_3 \cdot H_p - C_4 \cdot H^*(N) - C_5 \cdot p}{N}.$$

Now we determine (N, p). Choose p to be be small enough that

$$C_3 \cdot H_p - C_5 \cdot p > 0$$

and then choose N to be large enough that

$$C_4 \cdot H^*(N) < C_3 \cdot H_p - C_5 \cdot p.$$

We have now chosen (N, p) so that

$$\Delta > 0.$$

Step 8 (Relation to original measures). It remains to show that μ_1 and μ_1' are related by a class diagonal joining. The probability distribution of $(\hat{x}, \hat{z})$ is clearly a joining of μ_1 and μ_1'. By the construction of $\hat{z}$, the event of $(\hat{x}, \hat{z})$ falling to the class diagonal D_1 has probability one. Therefore, the distribution of $(\hat{x}, \hat{z})$ is a class diagonal joining. □

COROLLARY 8.2. *Let (X, Y, π) be a factor triple and ν an ergodic measure on Y. Let $V \in \mathrm{Bow}(X)$. Let λ be a relative joining of distinct ergodic measures μ_1, μ_2 over ν such that $\lambda(D_1) > 0$ where D_1 is the class diagonal. Then there is another relative joining λ' on X^2 over ν such that*

$$h(\lambda') + \mu_1'(V) + \mu_2'(V) > h(\lambda) + \mu_1(V) + \mu_2(V)$$

where $\mu_1' = p_1(\lambda')$ and $\mu_2' = p_2(\lambda')$.

Proof. We may assume $0 < p := \lambda(D_1) < 1$. We can decompose λ into convex combination of two invariant measures:

$$\lambda = p\lambda_1 + (1-p)\lambda_2$$

where $\lambda_1(D_1) = 1$ and then both λ_i are relative joinings of μ_1, μ_2 over ν because μ_1, μ_2, ν are assumed ergodic. By the previous lemma, there is a relative joining λ_1' over ν such that

$$h(\lambda_1') + (p_1(\lambda_1'))(V) + (p_2(\lambda_1'))(V) > h(\lambda_1) + \mu_1(V) + \mu_2(V).$$

We write

$$\lambda' = p\lambda_1' + (1-p)\lambda_2$$

then λ' is a relative joining over ν.

It is an easy check that λ' satisfies the strict inequality in the conclusion. □

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

COROLLARY 8.3. *Let (X, Y, π) be a factor triple and ν an ergodic measure on Y. Let $V \in \mathrm{Bow}(X)$. Let λ be the relatively independent joining of distinct ergodic measures μ_1, μ_2 over ν where μ_1, μ_2 are both relative equilibrium states of V over ν. Then $\lambda(D_1) = 0$.*

Proof. Suppose $\lambda(D_1) > 0$ instead. The previous corollary then applies to produce another relative joining λ' on X^2 over ν such that

$$h(\lambda') + \mu_1'(V) + \mu_2'(V) > h(\lambda) + \mu_1(V) + \mu_2(V)$$

where $\mu_1' = p_1(\lambda')$ and $\mu_2' = p_2(\lambda')$. Note that $\mu_1, \mu_2, \mu_1', \mu_2'$ all project to ν.

Using the subadditivity of relative entropy and the fact that relative entropy is additive for relatively independent joining, we have

$$\begin{aligned} h(\lambda) + \mu_1(V) + \mu_2(V) &= h(\mu_1 \mid \nu) + h(\mu_2 \mid \nu) + h(\nu) + \mu_1(V) + \mu_2(V) \\ &= (h(\mu_1 \mid \nu) + \mu_1(V)) + (h(\mu_2 \mid \nu) + \mu_2(V)) + h(\nu) \\ &\geq (h(\mu_1' \mid \nu) + \mu_1'(V)) + (h(\mu_2' \mid \nu) + \mu_2'(V)) + h(\nu) \\ &= h(\mu_1' \mid \nu) + h(\mu_2' \mid \nu) + h(\nu) + \mu_1'(V) + \mu_2'(V) \\ &\geq h(\lambda') + \mu_1'(V) + \mu_2'(V) \end{aligned}$$

which contradicts our initial strict inequality. □

THEOREM 8.1. *Let (X, Y, π) be a factor triple and ν an ergodic measure on Y. Let $V \in \mathrm{Bow}(X)$. The number of ergodic relative equilibrium states of V over ν is at most the class degree of ν.*

Proof. Suppose d is the class degree of ν and that $\mu_1, \ldots, \mu_{d+1}$ are $d + 1$ distinct ergodic relative equilibrium states of V over ν. Form the $(d + 1)$-fold relatively independent joining of these $d + 1$ measures over ν. The fact that there are only d transition classes over ν-almost every y ensures the existence of distinct i, j such that the projection of the $(d + 1)$-fold joining to i, j violates the previous corollary. □

COROLLARY 8.4. *Let (X, Y, π) be a factor triple and ν an ergodic measure on Y with full support. Let $V \in \mathrm{Bow}(X)$. The number of ergodic relative equilibrium states of V over ν is at most the class degree of π.*

Proof. Since ν has full support, the class degree of ν is the class degree of π. □

We raise the following question.

Question 8.5. Let (X, Y, π) be a factor triple. Suppose X is irreducible and π is infinite-to-one. Is there always an ergodic measure ν on Y with full support such that the number of ergodic measures of relative maximal entropy over ν is the class degree of π? For each function $V \in \mathrm{Bow}(X)$, is there an ergodic measure ν on Y with full support such that the number of ergodic relative equilibrium states of V over ν is the class degree of π?

The infinite-to-one condition is included because if π is finite-to-one and X has infinitely many points, then there is an ergodic ν on Y with full support such that the number of ergodic measures over ν is the degree of π. A proof will be given in a forthcoming paper by the third author and Uijin Jung.

https://doi.org/10.1017/etds.2017.50 Published online by Cambridge University Press

Acknowledgements. The paper was written with support of BK21 PLUS SNU Mathematical Sciences Division and by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2012R1A6A3A01040839), Universidad de Santiago de Chile (USACH) Proyecto Basal USA 1555 and Fondecyt 11140369. The first author would like to express her sincere appreciation to Andrés Navas for useful discussions. The third author would like to thank Uijin Jung and Soonjo Hong for helping him understand transition classes more clearly and referees for their comments and patience in checking a lengthy proof.

REFERENCES

[**1**] M. Allahbakhshi, S. Hong and U. Jung. Structure of transition classes for factor codes on shifts of finite type. *Ergod. Th. & Dynam. Sys.* **35**(8) (2015), 2353–2370.

[**2**] M. Allahbakhshi and A. Quas. Class degree and relative maximal entropy. *Trans. Amer. Math. Soc.* **365**(3) (2013), 1347–1368.

[**3**] J. Antonioli. Compensation functions for factors of shifts of finite type. *Ergod. Th. & Dynam. Sys.* **36**(2) (2016), 375–389.

[**4**] R. Bowen. *Symbolic Dynamics for Hyperbolic Systems (Lecture Notes in Mathematics, 318)*. Springer, Berlin, 1975, pp. 51–58.

[**5**] F. Ledrappier and P. Walters. A relativised variational principle for continuous transformations. *J. Lond. Math. Soc. (2)* **16**(3) (1977), 568–576.

[**6**] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.

[**7**] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–66.

[**8**] K. Petersen, A. Quas and S. Shin. Measures of maximal relative entropy. *Ergod. Th. & Dynam. Sys.* **23**(1) (2003), 207–223.

[**9**] D. Ruelle. *Thermodynamic Formalism*. Addison-Wesley, Reading, MA, 1978.

[**10**] Y. Sinai. Gibbs measures in ergodic theory. *Uspekhi Mat. Nauk* **27** (1972), 21–64.

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

[**12**] J. Yoo. Measures of maximal relative entropy with full support. *Ergod. Th. & Dynam. Sys.* **31**(6) (2011), 1889–1899.

[**13**] J. Yoo. Multiplicity structure of preimages of invariant measures under finite-to-one factor maps. *Trans. Amer. Math. Soc.* to appear.