

# INTRINSIC ERGODICITY VIA OBSTRUCTION ENTROPIES

VAUGHN CLIMENHAGA AND DANIEL J. THOMPSON

ABSTRACT. Bowen showed that a continuous expansive map with specification has a unique measure of maximal entropy. We show that the conclusion remains true under weaker non-uniform versions of these hypotheses. To this end, we introduce the notions of obstructions to expansivity and specification, and show that if the entropy of such obstructions is smaller than the topological entropy of the map, then there is a unique measure of maximal entropy.

## 1. INTRODUCTION

Bowen showed that expansivity and specification imply intrinsic ergodicity [Bow75] – that is, existence of a unique measure of maximal entropy. In [CT], we introduced a non-uniform version of specification for a shift space which guarantees intrinsic ergodicity. In this paper, we establish our techniques in a non-symbolic setting. The key idea of [CT] is that the obstructions to specification should have less entropy than the whole space. Here we adapt this idea to a non-symbolic setting, and replace expansivity (which is automatically satisfied by a shift space) with a condition which says that obstructions to positive expansivity should have less entropy than the whole space (we work with positive expansivity rather than expansivity for convenience only, see remark after Definition 2.4).

After introducing precise definitions of  $h_{\text{spec}}^{\perp}(f)$ , which denotes the entropy of obstructions to specification, and  $h_{\text{exp}^+}^{\perp}(f)$ , which denotes the entropy of obstructions to expansivity, we establish the following result:

**Theorem A.** *Let  $X$  be a compact metric space and  $f: X \rightarrow X$  a continuous map. If  $h_{\text{spec}}^{\perp}(f) < h_{\text{top}}(f)$  and  $h_{\text{exp}^+}^{\perp}(f) < h_{\text{top}}(f)$ , then  $f$  is intrinsically ergodic.*

The obstruction entropies  $h_{\text{spec}}^{\perp}(f)$  and  $h_{\text{exp}^+}^{\perp}(f)$  are obtained as limits of quantities  $h_{\text{spec}}^{\perp}(f, \epsilon)$  and  $h_{\text{exp}^+}^{\perp}(f, \epsilon)$ , which are the entropies of obstructions to specification and expansivity at a fixed finite scale  $\epsilon$ . Theorem A is a consequence of the following result, in which we consider obstruction entropies only at suitable fixed scales.

---

*Date:* January 11, 2013.

V.C. was supported by an NSERC Postdoctoral Fellowship. D.T. is supported by NSF grant DMS-1101576.

**Theorem B.** *Let  $X$  be a compact metric space and  $f: X \rightarrow X$  a continuous map. Suppose  $\epsilon > 0$  is such that  $h_{\text{spec}}^\perp(f, \epsilon) < h_{\text{top}}(f)$  and  $h_{\text{exp}^+}^\perp(f, 28\epsilon) < h_{\text{top}}(f)$ . Then  $f$  is intrinsically ergodic.*

The hypotheses of Theorem B are a priori weaker and potentially easier to check than the scale-free hypotheses of Theorem A. It is unavoidable that we investigate expansivity at a larger scale than specification. This is due to changes of scale that occur in the course of the proof. In our setting, this cannot be eliminated by standard rescaling techniques of the type used by Bowen [Bow75]. (See remark at the end of Section 2.3.) This accounts for the (not necessarily optimal) factor of 28 in the statement of Theorem B.

Roughly speaking,  $h_{\text{exp}^+}^\perp(f, \epsilon)$  and  $h_{\text{spec}}^\perp(f, \epsilon)$  may be understood as follows. Positive expansivity is equivalent to having  $\bigcap_{n \geq 0} B_n(x, \epsilon) = \{x\}$  for all  $x \in X$  and all sufficiently small  $\epsilon > 0$ . We define  $h_{\text{exp}^+}^\perp(f, \epsilon)$  as the supremum of the entropies of ergodic measures giving positive weight to the set of points at which this condition fails.

Specification at scale  $\epsilon$  means that there exists a constant  $\tau$  such that every finite collection of finite orbit segments  $(x_i, \dots, f^{n_i}x_i)$  can be  $\epsilon$ -shadowed by a single orbit which takes  $\tau$  iterates to transition from one segment to the next. We look for a collection  $\mathcal{S}$  of orbit segments (obstructions to specification) such that any finite collection of finite orbit segments can be  $\epsilon$ -shadowed with gap size  $\tau$  *provided we are allowed to first remove elements of  $\mathcal{S}$  from the ends of each segment*. We define  $h_{\text{spec}}^\perp(f, \epsilon)$  to be the infimum of the entropy of such a collection  $\mathcal{S}$ .

Our goal for this note is to establish these techniques and concepts, particularly the use of  $h_{\text{exp}^+}^\perp$ . As a first novel application of Theorem B, we can establish intrinsic ergodicity for non-symbolic factors of  $\beta$ -shifts, subject to a weak expansivity condition, extending our results from [CT]. This includes the following result:

**Theorem C.** *Every positively expansive factor of a  $\beta$ -shift is intrinsically ergodic.*

This result suggests that arbitrary  $\beta$ -shifts may have value as coding spaces. The coding of (algebraic) dynamical systems by  $\beta$ -shifts has been well studied in the special case that  $\beta$  is a Pisot number, and hence the corresponding  $\beta$ -shift is sofic (see [Sch00, Sid03, LS04] for an extensive list of references).

## 2. DEFINITIONS

Let  $X$  be a compact metric space and  $f: X \rightarrow X$  a continuous map. We recall some standard definitions, and introduce the definitions of  $h_{\text{spec}}^\perp$  and  $h_{\text{exp}^+}^\perp$ . The notions that follow depend on the dynamics  $f$ , although our notation suppresses this dependence in order to simplify the presentation.

## 2.1. Entropy.

**Definition 2.1.** A set  $E \subset X$  is  $(n, \varepsilon)$ -separated for some  $n \in \mathbb{N}$  and  $\varepsilon > 0$  if for all  $x \neq y \in E$  we have  $y \notin B_n(x, \varepsilon)$ , where

$$B_n(x, \varepsilon) = \{y \in X \mid d_n(x, y) < \varepsilon\}$$

and  $d_n(x, y) = \max\{d(f^i x, f^i y) \mid 0 \leq i \leq n-1\}$ . Given  $Z \subset X$ , we write  $\Lambda(Z, n, \varepsilon)$  for the maximum cardinality of a  $(n, \varepsilon)$ -separated set  $E \subset Z$ . Let

$$(2.1) \quad \bar{h}(Z, \varepsilon) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \Lambda(Z, n, \varepsilon), \quad \bar{h}(Z) = \lim_{\varepsilon \rightarrow 0} \bar{h}(Z, \varepsilon),$$

and define  $\underline{h}$  analogously. If  $\underline{h}(Z, \varepsilon) = \bar{h}(Z, \varepsilon)$  then we write  $h(Z, \varepsilon)$  for the common value and call this the topological entropy of  $Z$  at scale  $\varepsilon$ .

*Remark.* In fact,  $\underline{h}$  and  $\bar{h}$  are the lower and upper capacity topological entropies [Pes97]. If  $Z$  is invariant then  $\underline{h}(Z, \varepsilon) = \bar{h}(Z, \varepsilon)$  for all  $\varepsilon > 0$  [Pes97, Theorem 11.5]. In particular,  $h(X, \varepsilon)$  exists for every  $\varepsilon$ , and  $h(X)$  is the standard definition of topological entropy for the system  $(X, f)$ . For unity of notation, from now on we will usually write  $h(X)$  in place of  $h_{\text{top}}(f)$ .

The variational principle states that  $h(X) = \sup_{\mu} h_{\mu}(f)$ , where  $h_{\mu}(f)$  is the measure-theoretic entropy of  $\mu$  and the supremum is taken over all Borel  $f$ -invariant probability measures on  $X$  [Wal82]. A measure achieving the supremum is a *measure of maximal entropy* (MME), and a system with a unique MME is called *intrinsically ergodic* [CT, Wei70, Hof79, Hof81, Buz97].

We will have occasion to consider the entropy of a *sequence* of sets  $Z_n$ . Such a sequence generates a subset  $\mathcal{D} \subset X \times \mathbb{N}$  in a natural way, as  $\mathcal{D} = \bigcup_n Z_n \times \{n\}$ . Conversely, given  $\mathcal{D} \subset X \times \mathbb{N}$  we obtain a sequence of sets  $\mathcal{D}_n = \{x \in X \mid (x, n) \in \mathcal{D}\}$ . We think of subsets  $\mathcal{D} \subset X \times \mathbb{N}$  as collections of points and times, so that given  $x \in X$ , the set  $\{n \in \mathbb{N} \mid (x, n) \in \mathcal{D}\}$  can be thought of as those times at which the orbit segment  $(x, f(x), \dots, f^n(x))$  satisfies a certain property.

**Definition 2.2.** Give  $\mathcal{D} \subset X \times \mathbb{N}$ , write  $\Lambda(\mathcal{D}, n, \varepsilon) = \Lambda(\mathcal{D}_n, n, \varepsilon)$ . The (upper) topological entropy of  $\mathcal{D}$  is

$$(2.2) \quad \bar{h}(\mathcal{D}, \varepsilon) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \Lambda(\mathcal{D}, n, \varepsilon), \quad \bar{h}(\mathcal{D}) = \lim_{\varepsilon \rightarrow 0} \bar{h}(\mathcal{D}, \varepsilon).$$

This procedure of taking the capacity entropy of a sequence of sets appears in the definition of coarse multifractal spectra, and a general discussion of the relationship between this quantity and other dimensional quantities is given in [Cli13].

**2.2. Expansivity.** Positive expansivity at scale  $\varepsilon$  is equivalent to the following property: for every  $x \in X$ , we have

$$(2.3) \quad \Phi_x^+(\varepsilon) := \bigcap_{n \geq 0} B_n(x, \varepsilon) = \{x\}.$$

**Definition 2.3.** Denote by  $\mathcal{N}^+(\varepsilon) := \{x \in X \mid \Phi_x^+(\varepsilon) \neq \{x\}\}$  the set of non-expansive points at scale  $\varepsilon$ . Following Buzzi and Fisher [BF11], we say that an  $f$ -invariant measure  $\mu$  is almost positively expansive at scale  $\varepsilon$  if  $\mu(\mathcal{N}^+(\varepsilon)) = 0$ ; in other words, if  $\Phi_x^+(\varepsilon) = \{x\}$  for  $\mu$ -a.e.  $x$ .

**Definition 2.4.** Writing  $\mathcal{M}_f^e$  for the set of ergodic  $f$ -invariant Borel probability measures on  $X$ , the entropy of obstructions to positive expansivity at scale  $\varepsilon$  is

$$\begin{aligned} h_{\text{exp}^+}^\perp(f, \varepsilon) &= \sup\{h_\mu(f) \mid \mu \in \mathcal{M}_f^e \text{ is not almost positively expansive}\} \\ &= \sup\{h_\mu(f) \mid \mu \in \mathcal{M}_f^e \text{ and } \mu(\mathcal{N}^+(\varepsilon)) > 0\}. \end{aligned}$$

We also define  $h_{\text{exp}^+}^\perp(f) = \lim_{\varepsilon \rightarrow 0} h_{\text{exp}^+}^\perp(f, \varepsilon)$ . The limit exists since  $h_{\text{exp}^+}^\perp(f, \varepsilon)$  is monotonic when considered as a function of  $\varepsilon$ .

*Remark.* Positive expansivity is the appropriate definition for non-invertible maps. For invertible maps, Buzzi and Fisher define almost expansivity using  $\Phi_x(\varepsilon) := \bigcap_{n \geq 0} f^n(B_{2n}(f^{-n}x, \varepsilon))$  and requiring that  $\mu$  give zero measure to the non-expansive set  $\mathcal{N}(\varepsilon) := \{x \in X \mid \Phi_x(\varepsilon) \neq \{x\}\}$ . Thus, when  $f$  is an expansive homeomorphism on  $X$ , every  $f$ -invariant measure is almost-expansive. We can define the *entropy of obstructions to expansivity at scale  $\varepsilon$*  for a homeomorphism  $f$  as

$$\begin{aligned} h_{\text{exp}}^\perp(\varepsilon) &= \sup\{h_\mu(f) \mid \mu \text{ is } f\text{-invariant and not almost expansive}\} \\ &= \sup\{h_\mu(f) \mid \mu \text{ is } f\text{-invariant and } \mu(\mathcal{N}(\varepsilon)) > 0\}. \end{aligned}$$

We do not consider the invertible case or  $h_{\text{exp}}^\perp$  in this note. This is purely for convenience. Our proofs go through with minimal modification in the situation when  $f$  is invertible and  $h_{\text{exp}}$  is less than the entropy of the space.

The following proposition gives a useful method for bounding  $h_{\text{exp}^+}^\perp$ .

**Proposition 2.1.**  $h_{\text{exp}^+}^\perp(f, \varepsilon) \leq \underline{h}(\mathcal{N}^+(\varepsilon))$ .

*Proof.* Given  $h < h_{\text{exp}^+}^\perp(f, \varepsilon)$ , let  $\mu \in \mathcal{M}_f^e$  be such that  $h_\mu(f) > h$  and  $\mu(\mathcal{N}^+(\varepsilon)) > 0$ . By the Brin–Katok local entropy formula [BK83], there exists  $N \in \mathbb{N}$  and  $Y \subset X$  such that  $\mu(Y) > 1 - \mu(\mathcal{N}^+(\varepsilon))$  and  $\mu(B_n(x, \varepsilon)) \leq e^{-nh}$  for all  $n \geq N$ . Taking  $Z = Y \cap \mathcal{N}^+(\varepsilon)$ , we get  $\mu(Z) > 0$ .

Let  $E_n \subset Z$  be any maximal  $(n, \varepsilon)$ -separated set. By maximality, we have  $Z \subset \bigcup_{x \in E_n} B_n(x, \varepsilon)$ , whence  $\mu(Z) \leq (\#E_n)e^{-nh}$ . Thus  $\#E_n \geq \mu(Z)e^{nh}$ , and it follows that  $\underline{h}(\mathcal{N}^+(\varepsilon)) \geq \underline{h}(Z, \varepsilon) \geq h$ . Since  $h < h_{\text{exp}^+}^\perp(f, \varepsilon)$  was arbitrary, we are done.  $\square$

The following two propositions are crucial to our approach.

**Proposition 2.2.** *If  $\mu$  is almost positively expansive at scale  $\varepsilon$  and  $\mathcal{A}$  is a measurable partition of  $X$  such that every element of  $\mathcal{A}$  is contained in  $B(x, \varepsilon)$  for some  $x \in X$ , then  $\mathcal{A}$  is (one-sided) generating for  $\mu$ .*

**Proposition 2.3.** *If  $h_{\text{exp}^+}^\perp(f, \varepsilon) < h(X)$ , then  $h(X, \varepsilon) = h(X)$ .*

Proposition 2.2 is immediate from the definitions. Before proving Proposition 2.3, we introduce a useful concept.

**Definition 2.5.** *Let  $E_n \subset X$  be an  $(n, \varepsilon)$ -separated set of maximum cardinality. We say that a partition  $\mathcal{A}^n$  is adapted to  $E_n$  if  $\mathcal{A}^n$  is a partition such that each atom  $w$  of the partition satisfies*

$$B_n\left(x, \frac{\varepsilon}{2}\right) \subset w \subset B_n(x, \varepsilon)$$

for some  $x \in E_n$ .

The existence of adapted partitions for any  $E_n$  follows from the fact that the sets  $B_n(x, \varepsilon/2)$  are disjoint and the sets  $B_n(x, \varepsilon)$  cover  $X$ . For every  $n \geq 1$ , the non-expansive set for the map  $f^n$  with respect to the Bowen metric  $d_n = \max_{0 \leq k < n} d \circ f^k$  coincides with the non-expansive set for  $f$  with respect to the original metric  $d$ . Thus, Proposition 2.2 shows that if  $\mu$  is almost positively expansive at scale  $\varepsilon$  and  $\mathcal{A}^n$  is an adapted partition for a maximal  $(n, \varepsilon)$ -separated set, then  $\mathcal{A}^n$  is generating for  $\mu$  under the map  $f^n$ .

*Proof of Proposition 2.3.* For each  $n \geq 1$ , let  $E_n$  be a maximal  $(n, \varepsilon)$ -separated set, and  $\mathcal{A}^n$  an adapted partition for  $E_n$ . Fix  $\mu \in \mathcal{M}_f^e$  with  $h_\mu(f) > h_{\text{exp}^+}^\perp(f, \varepsilon)$ , noting that such  $\mu$  exist by the variational principle. By the above remarks,  $\mathcal{A}^n$  is generating for  $\mu$  under the map  $f^n$ . In particular, we have

$$h_\mu(f) = \frac{1}{n} h_\mu(f^n) = \frac{1}{n} h_\mu(f^n, \mathcal{A}^n) \leq \frac{1}{n} \log \#\mathcal{A}^n = \frac{1}{n} \log \#E_n,$$

and sending  $n \rightarrow \infty$  gives  $h_\mu(f) \leq h(X, \varepsilon)$ . Taking a supremum over all such  $\mu$  and applying the variational principle gives the result.  $\square$

We will see in §3.5 that Proposition 2.3 implies the existence of an MME.

### 2.3. Specification.

**Definition 2.6.** *A collection of points and times  $\mathcal{G} \subset X \times \mathbb{N}$  has specification at scale  $\varepsilon$  if there exists  $\tau \in \mathbb{N}$  such that for every  $\{(x_j, n_j) \mid 0 \leq j \leq k\} \subset \mathcal{G}$ ,*

$$(2.4) \quad \bigcap_{j=0}^k f^{-\sum_{i=0}^{j-1} (n_i + \tau)} B_{n_j}(x_j, \varepsilon) \neq \emptyset.$$

*If the intersection always contains a periodic point with period  $\sum_{i=0}^k (n_i + \tau)$ , we say that  $\mathcal{G}$  has (Per)-specification at scale  $\varepsilon$ .*

*Remark.* There are two standard definitions of specification for the dynamical system  $(X, f)$ , which differ from each other only in whether or not one requires that the shadowing orbit be periodic. Using our terminology, these properties are equivalent to the requirement that the entire set  $X \times \mathbb{N}$  has specification, or (Per)-specification, for every  $\varepsilon > 0$ .

**Definition 2.7.** A triple  $(\mathcal{P}, \mathcal{G}, \mathcal{S}) \subset (X \times \mathbb{N})^3$  is a decomposition for  $(X, f)$  if for every  $x \in X$  and  $n \in \mathbb{N}$  there exist  $p, g, s \in \mathbb{N}$  with  $p + g + s = n$  so that

$$(x, p) \in \mathcal{P}, \quad (f^p x, g) \in \mathcal{G}, \quad (f^{p+g} x, s) \in \mathcal{S}.$$

That is, every orbit segment can be decomposed into a ‘prefix’ from  $\mathcal{P}$ , a ‘good’ core from  $\mathcal{G}$ , and a ‘suffix’ from  $\mathcal{S}$ . Given a decomposition  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$ , we will also need to consider the following collections of points and times:

$$\mathcal{G}^M := \{(x, n) \in X \times \mathbb{N} \mid p(x, n) \leq M, s(x, n) \leq M\}.$$

*Remark.* Note that  $X \times \mathbb{N} = \bigcup_M \mathcal{G}^M$ , so that a decomposition  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$  defines a filtration of the collection of all points and times. We are interested in the situation when each level  $\mathcal{G}^M$  of this filtration has specification. In the course of the proof we will see that every MME  $\mu$  has the property that  $\inf_n \mu(\mathcal{G}_n^M) \rightarrow 1$  as  $M \rightarrow \infty$ . This suggests that the collections  $\mathcal{G}^M$  can be thought of as regular sets analogous to the Pesin sets of non-uniform hyperbolicity theory.

**Definition 2.8.** The entropy of obstructions to specification at scale  $\varepsilon$  is

$$h_{\text{spec}}^\perp(f, \varepsilon) := \inf \{ \bar{h}(\mathcal{P} \cup \mathcal{S}, 3\varepsilon) \mid \mathcal{P}, \mathcal{S} \subset X \times \mathbb{N} \text{ are such that} \\ \text{there exists a decomposition } (\mathcal{P}, \mathcal{G}, \mathcal{S}) \text{ for } (X, f) \\ \text{for which every } \mathcal{G}^M \text{ has specification at scale } \varepsilon \}.$$

We define  $h_{\text{spec}}^\perp(f, \varepsilon) = \lim_{\varepsilon \rightarrow 0} h_{\text{spec}}^\perp(f, \varepsilon)$ . The limit exists since  $h_{\text{spec}}^\perp(f, \varepsilon)$  is monotonic when considered as a function of  $\varepsilon$ .

*Remark.* A priori,  $\mathcal{G}$  could have specification at scale  $\varepsilon$ , but not at a smaller scale  $\delta$ . Thus,  $h_{\text{spec}}^\perp(f, \varepsilon)$  may increase as  $\varepsilon$  decreases, which is why the hypotheses in Theorem B are a priori weaker than those in Theorem A. This contrasts with the classical setting of Bowen, where specification at a fixed scale and expansivity at that same scale together imply the specification property at all smaller scales. This can be shown by using expansivity and compactness to find  $N = N(\delta, \varepsilon)$  such that  $B_N(x, \varepsilon) \subset B(x, \delta)$  for every  $x$ , so that specification at scale  $\varepsilon$  with gap size  $\tau$  implies specification at scale  $\delta$  with gap size  $\tau + N(\delta, \varepsilon)$ . This argument does not work in our setting, because although such an  $N = N(x, \delta, \varepsilon)$  exists for each  $x \notin \mathcal{N}^+(\varepsilon)$ , it cannot be chosen uniformly.

### 3. PROOFS

Throughout this section,  $(X, d)$  is a compact metric space and  $f: X \rightarrow X$  is continuous. Large parts of the proof work without the full strength of the hypotheses of Theorem B, so for each partial result we state precisely which conditions are needed. Our method is inspired by Bowen’s original proof from [Bow75], incorporating the modifications from [CT] that allow us to use the weakened version of specification.

We outline the strategy of the proof. First, we obtain a series of combinatorial estimates which yield bounds on the maximum cardinalities of various  $(n, \delta)$ -separated sets (Lemmas 3.1-3.8). These estimates allow us to construct a measure of maximal entropy with a certain weak Gibbs property (Lemma 3.10), which we show is ergodic (Proposition 3.13). We use the weak Gibbs property, together with a crucial combinatorial estimate on positive measure sets (Lemma 3.12), to rule out the existence of any other measure of maximal entropy.

**3.1. Lower bounds on  $X$ .** We obtain a lower bound on the maximum cardinality of  $(n, \delta)$ -separated sets for  $X$  using a completely general, and rather standard, argument which is essentially contained in Bowen [Bow75].

**Lemma 3.1.**  $\Lambda(X, \sum_{j=1}^k n_j, 2\delta) \leq \prod_{j=1}^k \Lambda(X, n_j, \delta)$  for every  $\delta > 0$  and  $n_j$ .

*Proof.* Let  $E$  be maximal  $(\sum n_j, 2\delta)$ -separated and let  $E_j$  be maximal  $(n_j, \delta)$ -separated. Then  $E_j$  is  $(n_j, \delta)$ -spanning, and we define a map  $\pi: E \rightarrow E_1 \times \cdots \times E_k$  by the condition that  $f^{n_1 + \cdots + n_{j-1}}(x) \in B_{n_j}(\pi_j(x), \delta)$  for each  $j$ . This map is injective.  $\square$

**Lemma 3.2.** For every  $n \in \mathbb{N}$  and  $\delta > 0$ , we have  $\Lambda(X, n, \delta) \geq e^{nh(X, 2\delta)}$ .

*Proof.* By Lemma 3.1 we have  $\Lambda(X, kn, 2\delta) \leq (\Lambda(X, n, \delta))^k$ . It follows that  $\log \Lambda(X, n, \delta) \geq n(\frac{1}{kn} \log \Lambda(X, kn, 2\delta))$ , and we send  $k \rightarrow \infty$ .  $\square$

**3.2. Upper bounds on  $\mathcal{G}$ .** Using a standard argument based on the specification property of  $\mathcal{G}$ , we obtain upper bounds on the cardinality of an  $(n, \delta)$ -separated set in  $\mathcal{G}_n$ .

**Lemma 3.3.** If  $\mathcal{G} \subset X \times \mathbb{N}$  has specification at scale  $\delta$  with gap size  $\tau$ , then

$$\Lambda(X, n_1 + \cdots + n_k + (k-1)\tau, \delta) \geq \prod_{j=1}^k \Lambda(\mathcal{G}, n_j, 3\delta)$$

for every  $n_1, \dots, n_k$ .

*Proof.* Let  $E_j \subset \mathcal{G}_{n_j}$  be a  $(n_j, 3\delta)$ -separated set of maximum cardinality, and use the specification property to define a map  $\pi: E_1 \times \cdots \times E_k \rightarrow X$  such that  $\pi(x_1, \dots, x_k)$  is contained in the intersection from (2.4). A computation similar to the proof of Lemma 3.1 shows that the image of  $\pi$  is  $(n_1 + \cdots + n_k + (k-1)\tau, \delta)$ -separated.  $\square$

**Lemma 3.4.** If  $\mathcal{G} \subset X \times \mathbb{N}$  has specification at scale  $\delta$  with gap size  $\tau$ , then  $\Lambda(\mathcal{G}, n, 3\delta) \leq e^{(n+\tau)h(X, \delta)}$  for every  $n \in \mathbb{N}$ .

*Proof.* Every  $(m, \delta)$ -separated set is  $(n, \delta)$ -separated for  $n > m$ . Thus Lemma 3.3 implies  $\Lambda(X, k(n+\tau), \delta) \geq (\Lambda(\mathcal{G}, n, 3\delta))^k$ , and the result follows by sending  $k \rightarrow \infty$  in the resulting inequality  $\log \Lambda(\mathcal{G}, n, 3\delta) \leq (n+\tau)(\frac{1}{k(n+\tau)} \log \Lambda(X, k(n+\tau), \delta))$ .  $\square$

**3.3. Lower bounds on  $\mathcal{G}$  and related sets.** First, we prove the following useful summability result.

**Lemma 3.5.** *If  $\mathcal{P}, \mathcal{S} \subset X \times \mathbb{N}$  are such that  $\bar{h}(\mathcal{P} \cup \mathcal{S}, \delta) < h(X, \delta)$ , then  $\sum_{i \geq M} \Lambda(\mathcal{P} \cup \mathcal{S}, i, \delta') e^{-ih(X, \delta)} \rightarrow 0$  as  $M \rightarrow \infty$  for every  $\delta' \geq \delta$ .*

*Proof.* Write  $h := h(X, \delta)$  and let  $\gamma > 0$  be such that  $\bar{h}(\mathcal{P} \cup \mathcal{S}, \delta) < h - 2\gamma$ . Then there exists a constant  $C > 0$  such that  $\Lambda(\mathcal{P} \cup \mathcal{S}, n, \delta) < Ce^{n(h-\gamma)}$  for all  $n$ , and so

$$e^{-nh} \Lambda(\mathcal{P} \cup \mathcal{S}, n, \delta') \leq e^{-nh} \Lambda(\mathcal{P} \cup \mathcal{S}, n, \delta) < Ce^{-nh} e^{n(h-\gamma)} = Ce^{-n\gamma}.$$

Thus  $\sum_{n=0}^{\infty} e^{-nh} \Lambda(\mathcal{P} \cup \mathcal{S}, n, \delta') < C \sum_{n=0}^{\infty} (e^{-\gamma})^n < \infty$ , and the tail of the series converges to 0.  $\square$

The following result holds for decompositions where  $\mathcal{G}$  has specification, without any conditions on  $\mathcal{G}^M$  or on expansivity. The result should be interpreted as a lower bound on the growth rate of a ‘fattened up’ version of  $\mathcal{G}$ .

**Lemma 3.6.** *Let  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$  be a decomposition for  $(X, f)$  such that*

- (1)  $\mathcal{G}$  has specification at scale  $\delta$  with gluing time  $\tau$ , and
- (2)  $\bar{h}(\mathcal{P} \cup \mathcal{S}, \delta) < h(X, \delta)$ .

*Then for every  $\gamma_1, \gamma_2 > 0$  there exists  $M \in \mathbb{N}$  such that the following is true: For any  $\mathcal{D} \subset X \times \mathbb{N}$  such that  $\Lambda(\mathcal{D}, n, 6\delta) \geq \gamma_1 e^{nh(X, \delta)}$  for all  $n$ , we have  $\Lambda(\mathcal{D} \cap \mathcal{G}^M, n, 6\delta) \geq (1 - \gamma_2) \Lambda(\mathcal{D}, n, 6\delta)$  for all  $n$ .*

*Proof.* Let  $D_n \subset \mathcal{D}$  be a  $(n, 6\delta)$ -separated set of maximum cardinality, and given  $i, j, k \in \mathbb{N}$ , let

$$D(i, j, k) = \{x \in D_n \mid n = i + j + k, p(x, n) = i, g(x, n) = j, s(x, n) = k\},$$

where  $p, g, s$  are as in Definition 2.7. Observe that  $\bigcup \{D(i, j, k) \mid i, k \leq M\} \subset \mathcal{D} \cap \mathcal{G}^M$  for every  $M$ , and in particular

$$D_n^M := \bigcup \{D(i, j, k) \mid i + j + k = n, i, k \leq M\}$$

is an  $(n, 6\delta)$ -separated subset of  $\mathcal{D} \cap \mathcal{G}^M$ , so  $\Lambda(\mathcal{D} \cap \mathcal{G}^M, n, 6\delta) \geq \#D_n^M$ .

Let  $E_i^P \subset \mathcal{P}_i$  be maximal  $(i, 3\delta)$ -separated, and similarly for  $E_j^G \subset \mathcal{G}_j$  and  $E_k^S \subset \mathcal{S}_k$ . As in Lemma 3.1, there is an injection  $\pi: D(i, j, k) \rightarrow E_i^P \times E_j^G \times E_k^S$ , and we obtain

$$\Lambda(\mathcal{D}, n, 6\delta) \leq \sum_{\substack{i+j+k=n \\ i, k \leq M}} \#D(i, j, k) + \sum_{\substack{i+j+k=n \\ i \vee k \geq M}} \Lambda(\mathcal{P}, i, 3\delta) \Lambda(\mathcal{G}, j, 3\delta) \Lambda(\mathcal{S}, k, 3\delta).$$

The first sum is equal to  $\#D_n^M$ . By Condition (1), Lemma 3.4 gives

$$\Lambda(\mathcal{D}, n, 6\delta) \leq \Lambda(\mathcal{D} \cap \mathcal{G}^M, n, 6\delta) + \sum_{\substack{i+j+k=n \\ i \vee k \geq M}} \Lambda(\mathcal{P}, i, 3\delta) \Lambda(\mathcal{S}, k, 3\delta) e^{(j+\tau)h(X, \delta)}.$$

Write  $a_i = \Lambda(\mathcal{P} \cup \mathcal{S}, i, 3\delta)e^{-ih(X,\delta)}$  and observe that by Condition (2) and Lemma 3.5, we have  $b_M := \sum_{i \geq M} a_i \rightarrow 0$  as  $M \rightarrow \infty$ . Together with the condition on  $\mathcal{D}$ , this yields

$$\begin{aligned} \Lambda(\mathcal{D}, n, 6\delta) &\leq \Lambda(\mathcal{D} \cap \mathcal{G}^M, n, 6\delta) + e^{(n+\tau)h(X,\delta)} \sum_{i \vee k \geq M} a_i a_k \\ &\leq \Lambda(\mathcal{D} \cap \mathcal{G}^M, n, 6\delta) + \gamma_1^{-1} \Lambda(\mathcal{D}, n, 6\delta) e^{\tau h(X,\delta)} b_M^2. \end{aligned}$$

Choosing  $M$  such that  $\gamma_1^{-1} e^{\tau h(X,\delta)} b_M^2 < \gamma_2$ , the proof is complete.  $\square$

*Remark.* If we assume the hypotheses of Theorem B, then we have  $h_{\text{spec}}^\perp(f, \delta) < h_{\text{top}}(f) = h_{\text{top}}(f, \delta)$  for every  $\delta \in [\epsilon, 28\epsilon]$ , where the equality uses Proposition 2.3. This gives the existence of a decomposition  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$  for  $(X, f)$  so that  $\mathcal{G}^M$  has specification for each  $M$ , and  $\bar{h}(\mathcal{P} \cup \mathcal{S}, \delta) < h(X, \delta)$ . Thus, the results of Lemmas 3.5 and 3.6 apply to this decomposition. This is the only place in the proof of Theorem B where we use the assumption  $h_{\text{spec}}^\perp(f, \epsilon) < h_{\text{top}}(f)$ .

**3.4. Upper bounds on  $\Lambda(X, n, 6\delta)$ .** From now on, we assume that  $h_{\text{spec}}^\perp(f, \delta) < h(X, \delta)$ . Thus, there is a decomposition  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$  for  $(X, f)$  satisfying

- (1) each  $\mathcal{G}^M$  has specification at scale  $\delta$  with gluing time  $\tau_M$ , and
- (2)  $\bar{h}(\mathcal{P} \cup \mathcal{S}, \delta) < h(X, \delta)$ .

In this section, we also assume that

$$(3.1) \quad h(X, 12\delta) = h(X, \delta),$$

which, by Proposition 2.3, is a weaker assumption than  $h_{\text{exp}^+}^\perp(f, 12\delta) < h(X)$ .

**Lemma 3.7.** *Under the above conditions, for every  $\gamma > 0$  there exists  $M$  such that  $\Lambda(\mathcal{G}^M, n, 6\delta) \geq (1 - \gamma)\Lambda(X, n, 6\delta)$  for all  $n$ .*

*Proof.* Lemma 3.2 together with (3.1) gives  $\Lambda(X, n, 6\delta) \geq e^{nh(X,\delta)}$ , so we can apply Lemma 3.6.  $\square$

**Lemma 3.8.** *There exists  $C_1 \in \mathbb{R}$  such that  $e^{nh(\delta)} \leq \Lambda(X, n, 6\delta) \leq C_1 e^{nh(\delta)}$  for all  $n$ .*

*Proof.* As above, the first inequality follows from Lemma 3.2 and (3.1). The second inequality follows by applying Lemma 3.7 with  $\gamma = \frac{1}{2}$  and then applying Lemma 3.4 to  $\mathcal{G}^M$ , so that

$$\Lambda(X, n, 6\delta) \leq 2\Lambda(\mathcal{G}^M, n, 6\delta) \leq (2e^{\tau_M h(X,\delta)}) e^{nh(X,\delta)}. \quad \square$$

**3.5. An MME with a Gibbs property.** We replace (3.1) with the stronger assumption that for sufficiently small  $\delta > 0$ ,

$$(3.2) \quad h(X, \delta) = h(X).$$

This ensures the existence of an MME, constructed using the following standard argument. Let  $E_n \subset X$  be a maximal  $(n, \delta)$ -separated set, and take

$$\nu_n := \frac{1}{\#E_n} \sum_{x \in E_n} \delta_x.$$

In order to obtain invariant measures, we consider the measures

$$(3.3) \quad \mu_n := \frac{1}{n} \sum_{k=0}^{n-1} f_*^k \nu_n$$

and let  $\mu$  be a weak\* limit of the sequence  $\{\mu_n\}$ .

**Lemma 3.9.**  *$\mu$  is a measure of maximal entropy.*

*Proof.* The second part of [Wal82, Theorem 8.6] shows that  $h_\mu(f) \geq h(X, \delta)$ . Thus, by (3.2), if  $\delta$  is sufficiently small,  $\mu$  is an MME.  $\square$

So far we have used  $\delta$  to denote the scale at which we work, since the results will be used for various values of  $\delta$ . From now on, we fix the scale  $\epsilon$  as in Theorem B. We construct  $\mu$  as above using  $\delta = 6\epsilon$  in the selection of the sets  $E_n$ . By Proposition 2.3 and the assumption  $h_{\text{exp}^+}^\perp(f, 28\epsilon) < h_{\text{top}}(f)$ , (3.2) is satisfied for this value of  $\delta$ , so  $\mu$  is indeed an MME.

We now also assume that  $h_{\text{spec}}^\perp(f, \epsilon) < h_{\text{top}}(f)$ , and show that  $\mu$  satisfies a certain weak Gibbs property.

**Lemma 3.10.** *For sufficiently large  $M$ , there exists  $K_M > 0$  such that for every  $(x, n) \in \mathcal{G}^M$ , we have*

$$(3.4) \quad \mu(B_n(x, 7\epsilon)) \geq K_M e^{-nh(X)}.$$

*Proof.* Let  $M$  be large enough to satisfy the conclusion of Lemma 3.7 with  $\gamma = \frac{1}{2}$ . Then applying Lemma 3.2 with  $\delta = \frac{14}{6}\epsilon$ , we have  $\Lambda(\mathcal{G}^M, n, 14\epsilon) \geq \frac{1}{2}e^{nh(X)}$  for all  $n$ , and so there exists  $E_n^G \subset \mathcal{G}_n^M$  such that  $\#E_n^G \geq \frac{1}{2}e^{nh(X)}$  and  $E_n^G$  is  $(n, 14\epsilon)$ -separated.

To estimate  $\mu(B_n(x, 7\epsilon))$ , we first estimate  $\nu_m(f^{-k}B_n(x, 7\epsilon))$  for  $m \gg n$  and (most values of)  $0 \leq k \leq m$ . Let  $\tau = \tau_M \in \mathbb{N}$  be provided by the specification property. Fix  $\tau \leq k \leq m - n - \tau$ ; let  $\ell_1 = k - \tau$  and  $\ell_2 = m - k - \tau - n$ , so that  $\ell_1 + \tau + n + \tau + \ell_2 = m$ .

Using specification, for each  $y \in E_{\ell_1}^G$  and  $z \in E_{\ell_2}^G$  there exists

$$\varphi(y, z) \in B_{\ell_1}(y, \epsilon) \cap f^{-\ell_1 - \tau} B_n(x, \epsilon) \cap f^{-n - \ell_1 - 2\tau} B_{\ell_2}(z, \epsilon);$$

furthermore, because  $E_n^G$  is  $(n, 14\epsilon)$ -separated,  $\varphi$  is injective and  $\varphi(E_{\ell_1}^G \times E_{\ell_2}^G)$  is  $(m, 12\epsilon)$ -separated.

Because the set  $E_m$  used to construct  $\mu$  is  $(m, 6\epsilon)$ -spanning, we can define  $\psi: \varphi(E_{\ell_1}^G \times E_{\ell_2}^G) \rightarrow E_m$  such that  $d_m(w, \psi(w)) < 6\epsilon$ , and hence  $\psi \circ \varphi: E_{\ell_1}^G \times E_{\ell_2}^G \rightarrow E_m$  is injective. Furthermore,  $\varphi(y, z) \in f^{-k}B_n(x, \epsilon)$  for all  $y, z$ , and

hence  $\psi \circ \varphi(y, z) \in f^{-k}B_n(x, 7\epsilon)$ . Therefore

$$\nu_m(f^{-k}B_n(x, 7\epsilon)) \geq \frac{(\#E_{\ell_1}^G)(\#E_{\ell_2}^G)}{\#E_m} \geq \frac{e^{(\ell_1+\ell_2)h(X)}}{4C_1e^{mh(X)}},$$

where the bound on the denominator comes from Lemma 3.8. This holds for all  $\tau \leq k \leq m - n - \tau$ , and averaging over  $0 \leq k < m$  gives

$$\mu_m(B_n(x, 7\epsilon)) \geq \left( \frac{m - n - 2\tau}{4C_1m} \right) e^{-(2\tau+n)h(X)}.$$

Sending  $m \rightarrow \infty$  gives the result with  $K_M = (4C_1)^{-1}e^{-2\tau_M h(X)}$ .  $\square$

Later on, in the proof of ergodicity, we will need a very similar lemma that estimates the measure of a set of points whose trajectories are prescribed not on a single interval, but on two disjoint intervals.

**Lemma 3.11.** *For sufficiently large  $M$ , there exists  $K'_M > 0$  such that for every  $(x_1, n_1), (x_2, n_2) \in \mathcal{G}^M$  and  $q \geq 2\tau_M$ , we have*

$$\mu(B_{n_1}(x_1, 7\epsilon) \cap f^{-n_1-q}B_{n_2}(x_2, 7\epsilon)) \geq K'_M e^{-(n_1+n_2)h(X)}.$$

*Proof.* The proof closely resembles that of Lemma 3.10. Let  $M$  be large enough to satisfy the conclusion of Lemma 3.7 with  $\gamma = \frac{1}{2}$  and let  $\tau = \tau_M$ . As before, let  $E_n^G \subset \mathcal{G}_n^M$  be  $(n, 14\epsilon)$ -separated with  $\#E_n^G \geq \frac{1}{2}e^{nh(X)}$ .

Given  $m \gg n_1 + n_2 + q$  and  $\tau \leq k \leq m - n_2 - q - n_1 - \tau$ , let  $\ell_1 = k - \tau$  and  $\ell_2 = m - n_2 - q - n_1 - \tau - k$ , so that

$$m = \ell_1 + \tau + n_1 + \tau + \bar{q} + \tau + n_2 + \tau + \ell_2,$$

where  $\bar{q} = q - 2\tau$ . A similar argument to the one in Lemma 3.10 shows that

$$\begin{aligned} \nu_m(f^{-k}B_{n_1}(x_1, 7\epsilon) \cap f^{-k-n_1-q}B_{n_2}(x_2, 7\epsilon)) &\geq \frac{(\#E_{\ell_1}^G)(\#E_{\bar{q}}^G)(\#E_{\ell_2}^G)}{\Lambda(X, m, 6\epsilon)} \\ &\geq \frac{e^{\ell_1+\bar{q}+\ell_2}}{8C_1e^{mh(X)}}. \end{aligned}$$

Averaging over  $0 \leq k < m$  gives

$$\begin{aligned} \mu_m(f^{-k}B_{n_1}(x_1, 7\epsilon) \cap f^{-k-n_1-q}B_{n_2}(x_2, 7\epsilon)) \\ \geq \left( \frac{m - n_1 - q - n_1 - 2\tau}{8C_1m} \right) e^{-(4\tau+n_1+n_2)h(X)}, \end{aligned}$$

and sending  $m \rightarrow \infty$  gives the result with  $K'_M = (8C_1)^{-1}e^{-4\tau_M h(X)}$ .  $\square$

**3.6. Adapted partitions and positive measure sets for MMEs.** From now on we need the full strength of the hypotheses in Theorem B. In particular, we assume that every MME  $\nu$  is almost positively expansive at scale  $28\epsilon$ . We recall that the definition of an adapted partition is given in §2.2.

**Lemma 3.12.** *For every  $\gamma \in (0, 1)$  there exists  $C_\gamma > 0$  such that if  $\nu$  is an MME and  $\mathcal{D} \subset X \times \mathbb{N}$  satisfies  $\nu(\mathcal{D}_n) \geq \gamma$  for every  $n$ , then  $\mathcal{D}_n$  has non-empty intersection with at least  $C_\gamma e^{nh(X)}$  atoms of  $\mathcal{A}^n$  for any partition  $\mathcal{A}^n$  adapted to a maximal  $(n, 14\epsilon)$ -separated set.*

*Proof.* Because  $\nu$  is almost positively expansive, we see that for  $\nu$ -a.e.  $x \in X$  the following holds: for all  $y \neq x$  there exists  $m$  such that  $d(f^m y, f^m x) \geq 28\epsilon$ . Writing  $m = qn + r$  for  $q, r \in \mathbb{N}$ ,  $r < n$ , we see that  $d_n((f^n)^q y, (f^n)^q x) \geq 28\epsilon$ , and it follows that  $\mathcal{A}^n$  is generating for  $f^n$ . In particular, this implies that

$$\begin{aligned} h_\nu(f^n) &= h_\nu(f^n, \mathcal{A}^n) = \inf_{q \geq 1} \frac{1}{q} H_\nu(\mathcal{A}^n \vee (f^n)^{-1} \mathcal{A}^n \vee \dots \vee (f^n)^{-q+1} \mathcal{A}^n) \\ &\leq H_\nu(\mathcal{A}^n) = \sum_{w \in \mathcal{A}^n} -\nu(w) \log \nu(w). \end{aligned}$$

Without loss of generality, we assume that  $\mathcal{D}_n$  is a union of atoms of  $\mathcal{A}^n$ , and write this collection of atoms as  $\mathcal{W}_n$ . Writing  $\mathcal{W}_n^c = \mathcal{A}^n \setminus \mathcal{W}_n$  and  $\mathcal{D}_n^c = X \setminus \mathcal{D}_n$ , we recall that  $h_\nu(f^n) = nh_\nu(f) = nh(X)$  and obtain

$$nh(X) \leq \sum_{w \in \mathcal{W}_n} -\nu(w) \log \nu(w) + \sum_{w \in \mathcal{W}_n^c} -\nu(w) \log \nu(w).$$

Normalising each sum yields

$$\begin{aligned} (3.5) \quad nh(X) &\leq \nu(\mathcal{D}_n) \left( \sum_{w \in \mathcal{W}_n} -\frac{\nu(w)}{\nu(\mathcal{D}_n)} \log \left( \frac{\nu(w)}{\nu(\mathcal{D}_n)} \right) \right) \\ &\quad + \nu(\mathcal{D}_n^c) \left( \sum_{w \in \mathcal{W}_n^c} -\frac{\nu(w)}{\nu(\mathcal{D}_n^c)} \log \left( \frac{\nu(w)}{\nu(\mathcal{D}_n^c)} \right) \right) \\ &\quad + (-\nu(\mathcal{D}_n) \log \nu(\mathcal{D}_n) - \nu(\mathcal{D}_n^c) \log \nu(\mathcal{D}_n^c)). \end{aligned}$$

Recall that for any non-negative numbers  $a_1, \dots, a_k$  summing to 1, we have

$$\sum_{i=1}^k -a_i \log a_i \leq \log k.$$

Applying this to the first sum in (3.5) with the quantities  $a_i$  replaced by  $\frac{\nu(w_x)}{\nu(\mathcal{D}_n)}$ , and to the second sum with  $a_i$  replaced by  $\frac{\nu(w_x)}{\nu(\mathcal{D}_n^c)}$ , we see that

$$nh(X) \leq \nu(\mathcal{D}_n) \log \#\mathcal{W}_n + \nu(\mathcal{D}_n^c) \log \#\mathcal{W}_n^c + H(\nu(\mathcal{D}_n)),$$

where we write  $H(t) = -t \log t - (1-t) \log(1-t)$ . Lemma 3.8 implies that  $\#\mathcal{W}_n^c \leq \#S_n \leq C_1 e^{nh(X)}$ , and so writing  $C_2 = \log C_1 + \max_{t \in [0,1]} H(t)$ , we have

$$\begin{aligned} nh(X) &\leq \nu(\mathcal{D}_n) \log \#\mathcal{W}_n + (1 - \nu(\mathcal{D}_n))(\log C_1 + nh(X)) + H(\nu(\mathcal{D}_n)) \\ &= \nu(\mathcal{D}_n) \log \#\mathcal{W}_n + nh(X) - \nu(\mathcal{D}_n)(\log C_1 + nh(X)) + C_2, \end{aligned}$$

which gives

$$\begin{aligned}\nu(\mathcal{D}_n) \log \#\mathcal{W}_n &\geq \nu(\mathcal{D}_n)(\log C_1 + nh(X)) - C_2, \\ \log \#\mathcal{W}_n &\geq \log C_1 + nh(X) - \frac{C_2}{\gamma}.\end{aligned}$$

Thus it suffices to take  $C_\gamma = C_1 e^{-C_2/\gamma}$ .  $\square$

**3.7. Ergodicity.** Now we consider the MME constructed in §3.5, and show that it is ergodic. This is a consequence of the following proposition.

**Proposition 3.13.** *If two measurable sets  $P, Q \subset X$  both have positive  $\mu$ -measure, then  $\underline{\lim}_{n \rightarrow \infty} \mu(P \cap \sigma^{-n}(Q)) > 0$ .*

*Proof.* In fact, we show that for every  $\gamma > 0$  there exists  $\alpha > 0$  such that if  $\min\{\mu(P), \mu(Q)\} \geq 2\gamma$ , then

$$\underline{\lim}_{n \rightarrow \infty} \mu(P \cap \sigma^{-n}(Q)) \geq \alpha.$$

As in §3.6, we take a sequence of partitions  $\mathcal{A}^n$ , each of which is adapted to a maximal  $(n, 14\epsilon)$ -separated set  $S_n$ . Recall that each  $x \in S_n$  is identified with a partition element  $w_x \in \mathcal{A}^n$  such that

$$B_n(x, 7\epsilon) \subset w_x.$$

Fix  $\delta \in (0, \gamma)$ . Because  $\Phi_x(28\epsilon) = \{x\}$  for  $\mu$ -a.e.  $x \in X$ , we can assume WLOG that this holds for every  $x \in P, Q$ . Thus  $\bigcap_n B_n(x, 14\epsilon) = \{x\}$  for all  $x \in P, Q$ .

In what follows, to simplify notation, we sometimes use the same symbol (such as  $U$ ) to denote both a set of partition elements ( $U \subset \mathcal{A}^n$ ) and the union of those partition elements ( $U \subset X$ ).

**Lemma 3.14.** *Let  $\mu$  be a finite Borel measure on a compact metric space  $X$  and let  $P \subset X$  be measurable. Suppose  $\mathcal{A}^n$  is a sequence of partitions such that  $\bigcap_n \mathcal{A}^n(x) = \{x\}$  for every  $x \in P$ , and let  $\delta > 0$ . Then for all sufficiently large  $n$  there exists a collection  $U \subset \mathcal{A}^n$  such that  $\mu(U \triangle P) < \delta$ .*

*Proof.* This is a standard argument. Let  $R \subset P$  and  $S \subset X \setminus P$  be compact such that  $\mu(X \setminus (R \cup S)) < \delta/2$ . Now let  $\eta > 0$  be such that  $d(x, y) > \eta$  for all  $x \in R$  and  $y \in S$ . For all sufficiently large  $n$  there exists  $R' \subset R$  such that  $\mu(R \setminus R') < \delta/2$  and  $\text{diam } \mathcal{A}^n(x) < \eta$  for all  $x \in R'$ .

Let  $U = \{w \in \mathcal{A}^n \mid w \cap R' \neq \emptyset\}$ . Then  $R' \subset U \subset X \setminus S$ , and the result follows since  $U \triangle P \subset (X \setminus (R \cup S)) \cup (R \setminus R')$ .  $\square$

Returning to the proof of Proposition 3.13, we see that Lemma 3.14 allows us to choose, for all sufficiently large  $n$ , collections  $U, V \subset \mathcal{A}^n$  such that

$$\mu(U \triangle P) < \delta \text{ and } \mu(V \triangle Q) < \delta.$$

In particular  $\mu(U) > \mu(P) - \delta > \gamma$ , and  $\mu(V) > \mu(Q) - \delta > \gamma$ . Thus, by Lemma 3.12, we have

$$\#U \geq C_\gamma e^{nh(X)} \text{ and } \#V \geq C_\gamma e^{nh(X)}.$$

Let  $M$  be such that Lemma 3.6 holds with  $\gamma_1 = C_\gamma$  and  $\gamma_2 = \frac{1}{2}$ ; write  $U' = \{w_x \in U \mid x \in \mathcal{G}^M\}$  and similarly for  $V'$ . Then

$$(3.6) \quad \#U' \geq \frac{1}{2}C_\gamma e^{nh(X)} \text{ and } \#V' \geq \frac{1}{2}C_\gamma e^{nh(X)}.$$

For every  $w_{x_1} \in U'$  and  $w_{x_2} \in V'$  and all  $m \geq 2\tau + n$ , Lemma 3.11 implies that

$$\mu(w_{x_1} \cap f^{-m}w_{x_2}) \geq \mu(B_n(x_1, 7\epsilon) \cap f^{-m}B_n(x_2, 7\epsilon)) \geq K'_M e^{-2nh(X)}.$$

It follows that

$$\mu(U' \cap f^{-m}V') \geq \left(\frac{1}{4}C_\gamma^2 e^{2nh(X)}\right) (K'_M e^{-2nh(X)}) = \frac{C_\gamma^2 K'_M}{4} =: \alpha.$$

Now since  $U \supset U'$  and  $V \supset V'$ , and since for any  $m$ ,

$$|\mu(U \cap \sigma^{-m}V) - \mu(P \cap \sigma^{-m}Q)| < \delta,$$

we have

$$\varliminf_{m \rightarrow \infty} \mu(P \cap \sigma^{-m}Q) \geq \varliminf_{m \rightarrow \infty} \mu(U \cap \sigma^{-m}V) - \delta \geq \alpha - \delta.$$

Furthermore, since  $\delta > 0$  was arbitrary, we conclude that

$$\varliminf_{m \rightarrow \infty} \mu(P \cap \sigma^{-m}Q) \geq \alpha. \quad \square$$

**3.8. Contradiction if there is another mme.** Let  $\mu$  be the ergodic MME constructed in the previous sections, and suppose that there exists another ergodic MME  $\nu \perp \mu$ . Let  $P \subset X$  be such that  $\mu(P) = 0$  and  $\nu(P) = 1$ , and let  $\mathcal{A}^n$  be a sequence of partitions adapted to maximal  $(n, 14\epsilon)$ -separated sets  $S_n$ . By Lemma 3.14, there exists  $U_n \subset \mathcal{A}^n$  such that  $(\mu + \nu)(U_n \triangle P) \rightarrow 0$ . In particular, we have  $\nu(U_n) \rightarrow 1$  and  $\mu(U_n) \rightarrow 0$ .

Now  $\gamma := \inf_n \nu(U_n) > 0$ , and so by Lemma 3.12, we have  $\#U_n \geq C_\gamma e^{nh(X)}$  for all  $n$ . As in the proof of Proposition 3.13, fix  $M$  and let  $U'_n = \{w_x \in U_n \mid x \in \mathcal{G}^M\}$ ; by Lemma 3.6, there exists  $M$  such that  $\#U'_n \geq \frac{1}{2}C_\gamma e^{nh(X)}$  for all  $n$ .

Finally, we use the Gibbs property in Lemma 3.10 to observe that

$$\mu(U_n) \geq \mu(U'_n) \geq (\#U'_n)K_M e^{-nh(X)} \geq \frac{1}{2}C_\gamma K_M > 0,$$

which contradicts the fact that  $\mu(U_n) \rightarrow 0$ . This contradiction implies that any MME  $\nu$  is absolutely continuous with respect to  $\mu$ , and since  $\mu$  is ergodic, this in turn implies that  $\nu = \mu$ , which completes the proof of the theorem.

#### 4. APPLICATION TO FACTORS

We prove that every positively expansive topological factor of any  $\beta$ -shift has a unique measure of maximal entropy. The key point is that a decomposition  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$  for a system  $(X, f)$  induces a decomposition for a factor  $(Y, g)$  in a natural way. We demonstrate this phenomenon, which was

worked out in the symbolic case in [CT], in the special case of factors of the  $\beta$ -shift.

For a more detailed description of the structure of the  $\beta$ -shift  $(X, \sigma)$ , we refer to [CT] and the references therein; here we only state the properties we need for the proof. The key is that there exists a distinguished sequence  $w \in \Sigma^+ := \{0, 1, \dots, b\}^{\mathbb{N}}$  (the  $\beta$ -expansion of 1) such that  $z \in \Sigma^+$  is in  $X$  if and only if  $\sigma^n z \preceq w$  for all  $n \geq 0$ , where  $\preceq$  denotes the lexicographic order.

The language  $\mathcal{L}$  of the  $\beta$ -shift is the collection of all finite words that appear in sequences  $z \in X$ . We recall the decomposition of  $\mathcal{L}$  given in [CT]:  $\tilde{\mathcal{C}}^s = \{w_1 \cdots w_n \mid n \geq 1\}$  is the collection of prefixes of  $w$ , and  $\tilde{\mathcal{G}}$  is the collection of words in  $\mathcal{L}$  that do not end with any prefix of  $w$ . That is,

$$\tilde{\mathcal{G}} = \{v_1 \cdots v_m \in \mathcal{L} \mid v_{m-n+1} \cdots v_{m-1} v_m \neq w_1 \cdots w_n \text{ for all } 1 \leq n \leq m\}.$$

Every word  $v \in \mathcal{L}$  can be decomposed as  $v = v_1 \cdots v_m w_1 \cdots w_n$ , where  $v_1 \cdots v_m \in \tilde{\mathcal{G}}$ . We will use the following properties of words in  $\tilde{\mathcal{G}}$ , which are easily seen from the order-theoretic characterisation above or from the graph presentation of  $X$  described in [CT].

- (1) Given any  $v \in \mathcal{L}$ , the word  $v0^k$  is in  $\tilde{\mathcal{G}}$  for all sufficiently large  $k$ .
- (2) For any  $u \in \tilde{\mathcal{G}}$  and any  $v \in \mathcal{L}$ , the word  $uv \in \mathcal{L}$ .

Let  $(Y, g)$  be a topological factor of  $(X, \sigma)$ , and let  $\pi: X \mapsto Y$  be the factor map.

**Lemma 4.1.**  *$(Y, g)$  has positive topological entropy or  $Y$  is a single point.*

*Proof.* Assume that  $Y$  is not a single point. Then there exists  $x, y, \delta_0$  so that  $d(x, y) = 3\delta_0$  (where  $d$  is the metric on  $Y$ ). By uniform continuity of  $\pi$ , there exists  $\delta$  so that  $\pi(B(z, \delta)) \subset B(\pi z, \delta_0)$  for all  $z \in X$ . Fix  $\tilde{x} \in \pi^{-1}(x)$  and  $\tilde{y} \in \pi^{-1}(y)$ .

For sufficiently large  $n$ , there are words  $v_1, v_2 \in \tilde{\mathcal{G}}$  of length  $n$  such that  $[v_1] \subset B(\tilde{x}, \delta)$  and  $[v_2] \subset B(\tilde{y}, \delta)$ . To see this, take prefixes of  $\tilde{x}$  and  $\tilde{y}$  with sufficient length that the first condition holds, then append the symbol 0 to these prefixes enough times that the resulting words have the same length and are both in  $\tilde{\mathcal{G}}$ , using property (1).

Now, for any  $\underline{i} = i_1 \dots i_m \in \{1, 2\}^m$ , we can choose  $y_i \in [v_{i_1} v_{i_2} \dots v_{i_m}]$  using property (2), and define  $x_i = \pi y_i$ . The points  $\{x_i\}$  form a  $(nm, \delta_0)$ -separated set in  $Y$ . Thus  $\Lambda(Y, nm, \delta_0) \geq 2^m$  for every  $m$ . This shows that the topological entropy of  $(Y, g)$  is positive.  $\square$

**Proposition 4.2.**  $h_{\text{spec}}^\perp(f, \epsilon) = 0$  for every  $\epsilon > 0$ .

*Proof.* We obtain a decomposition  $(\mathcal{P}, \mathcal{G}, \mathcal{S})$  for  $(Y, g)$  from the decomposition  $\mathcal{L} = \tilde{\mathcal{G}}\tilde{\mathcal{C}}^s$ , then show that  $\mathcal{G}^M$  has specification for every  $\epsilon > 0$ . (Of course the transition time  $\tau$  depends on  $\epsilon$  and  $M$ .) let

$$\begin{aligned} \mathcal{G} &= \{(x, n) \mid x = \pi \tilde{x} \text{ for some } \tilde{x} \in X \text{ such that } \tilde{x}_1 \dots \tilde{x}_n \in \tilde{\mathcal{G}}\}, \\ \mathcal{S} &= \{(x, n) \mid x = \pi \tilde{x} \text{ for some } \tilde{x} \in X \text{ such that } \tilde{x}_1 \dots \tilde{x}_n \in \tilde{\mathcal{C}}^s\}. \end{aligned}$$

Setting  $\mathcal{P} = \emptyset$ , this defines a decomposition for  $Y$ .

**Lemma 4.3.**  $\mathcal{G}^M$  has the specification property for all  $\epsilon > 0$  and  $M \in \mathbb{N}$ .

*Proof.* Fix  $\epsilon$  and take  $\delta$  so that  $\pi(B(x, \delta)) \subset B(\pi x, \epsilon)$  for all  $x \in X$ . There exists  $m = m(\delta)$  so that for all  $\tilde{y} \in X$ , we have  $[\tilde{y}_1 \cdots \tilde{y}_m] \in B(y, \delta)$ . Then for every  $\tilde{z} \in [\tilde{y}_1 \cdots \tilde{y}_{m+n}]$ , we have  $d_n(\tilde{y}, \tilde{z}) < \delta$ , and so for  $y = \pi\tilde{y}$  and  $z = \pi\tilde{z}$  we have  $d_n(y, z) < \epsilon$ , where  $d_n(y, z) := \max_{0 \leq k \leq n} d(g^k(y), g^k(z))$  is the Bowen metric.

Given any collection  $\{(x^1, n_1), \dots, (x^k, n_k)\} \subset \mathcal{G}^M$ , let  $\tilde{x}^i \in \pi^{-1}x_i$  be such that  $\tilde{x}_1^i \cdots \tilde{x}_{n_i+m}^i \in \tilde{\mathcal{G}}^M$ . Since  $\tilde{\mathcal{G}}^M$  has specification with gap length  $\tau_M$ , we can find connecting words  $v^i$  of length  $\tau_M$  such that the cylinder

$$[\tilde{x}_1^1 \cdots \tilde{x}_{n_1+m}^1 v^1 \tilde{x}_1^2 \cdots \tilde{x}_{n_2+m}^2 v^2 \cdots \tilde{x}_1^k \cdots \tilde{x}_{n_k+m}^k]$$

is non-empty. In particular, any point  $z$  in the image of this cylinder under  $\pi$  verifies the specification property for the collection  $\{(x^1, n_1), \dots, (x^k, n_k)\}$ , where the gap length is  $m + \tau_M$ .  $\square$

**Lemma 4.4.**  $h(\mathcal{S}) = 0$ .

*Proof.* Let  $w = w_1 w_2 \dots$  be the  $\beta$ -expansion of 1, and let  $z = \pi(w)$ . Then, by definition,

$$\mathcal{S} \subset \{(x, n) \mid x = \pi\tilde{x}, \tilde{x}_1 \dots \tilde{x}_n = w_1 \dots w_n\}.$$

Take  $\delta$  so that  $\pi(B(x, \delta)) \subset B(\pi x, \epsilon)$  for all  $x \in X$ . There exists  $k$  so that if  $y \in \mathcal{S}_{n+k}$ , then there exists  $\tilde{y} \in \pi^{-1}(y) \cap B_n(w, \delta)$ , and thus  $y \in B_n(z, \epsilon)$ . Thus, for all  $n$ ,  $\mathcal{S}_{n+k} \subset B_n(z, \epsilon)$ . In particular,  $\Lambda(\mathcal{S}_{n+k}, n+k, \epsilon) \leq \Lambda(X, k, \epsilon)$ , and  $h(\mathcal{S}, \epsilon) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \Lambda(X, k, \epsilon) = 0$ . Since  $\epsilon$  was arbitrary, we are done.  $\square$

Proposition 4.2 follows from Lemmas 4.3 and 4.4.  $\square$

Using Proposition 4.2, we can apply Theorem B to  $(Y, g)$  as long as  $h_{\text{exp}^+}^\perp(g, \epsilon) < h(Y)$  for some  $\epsilon$ . In particular, we can apply Theorem B when  $(Y, g)$  is positively expansive.

## REFERENCES

- [BF11] Jerome Buzzi and Todd Fisher, *Entropic stability beyond partial hyperbolicity*, Preprint, 27 pages, [arXiv:1103.2707](https://arxiv.org/abs/1103.2707), 2011.
- [BK83] M. Brin and A. Katok, *On local entropy*, Geometric dynamics (Rio de Janeiro, 1981), Lecture Notes in Math., vol. 1007, Springer, Berlin, 1983, pp. 30–38. MR 730261 (85c:58063)
- [Bow75] Rufus Bowen, *Some systems with unique equilibrium states*, Math. Systems Theory **8** (1974/75), no. 3, 193–202. MR 0399413 (53 #3257)
- [Buz97] J. Buzzi, *Intrinsic ergodicity of smooth interval maps*, Israel J. Math. **100** (1997), no. 1, 125–161.
- [Cli13] Vaughn Climenhaga, *Topological pressure of simultaneous level sets*, Nonlinearity **26** (2013), no. 1, 241–268.
- [CT] Vaughn Climenhaga and Daniel J. Thompson, *Intrinsic ergodicity beyond specification:  $\beta$ -shifts,  $S$ -gap shifts, and their factors*, Israel J. Math., 26 pages, to appear, [arXiv:1011.2780](https://arxiv.org/abs/1011.2780).

- [Hof79] F. Hofbauer, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy*, Israel J. Math. **34** (1979), no. 3, 213–237.
- [Hof81] ———, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy II*, Israel J. Math. **38** (1981), no. 1-2, 107–115.
- [LS04] E. Lindenstrauss and K. Schmidt, *Invariant sets and measures of nonexpansive group automorphisms*, Israel J. Math. **144** (2004), 29–60.
- [Pes97] Yakov B. Pesin, *Dimension theory in dynamical systems*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1997, Contemporary views and applications. MR 1489237 (99b:58003)
- [Sch00] Klaus Schmidt, *Algebraic coding of expansive group automorphisms and two-sided beta-shifts*, Monatsh. Math. **129** (2000), no. 1, 37–61. MR 1741033 (2001f:54043)
- [Sid03] N. Sidorov, *Arithmetic dynamics*, Topics in Dynamics and Ergodic Theory, LMS Lecture Notes Ser., vol. 310, 2003, pp. 145–189.
- [Wal82] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982. MR 648108 (84e:28017)
- [Wei70] B. Weiss, *Intrinsically ergodic systems*, Bull. Amer. Math. Soc. **76** (1970), no. 6, 1266–1269.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON, HOUSTON, TEXAS 77204

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, 100 MATH TOWER,  
231 WEST 18TH AVENUE, COLUMBUS, OHIO 43210

*E-mail address:* `climinha@math.uh.edu`

*E-mail address:* `thompson@math.osu.edu`