Ergod. Th. & Dynam. Sys. (2019), **39**, 2456–2480 © *Cambridge University Press, 2018*
doi:10.1017/etds.2017.135

One-sided almost specification and intrinsic ergodicity

VAUGHN CLIMENHAGA† and RONNIE PAVLOV‡

† *Department of Mathematics, University of Houston, 4800 Calhoun St., Houston, TX 77204, USA*
(e-mail: climenha@math.uh.edu)

‡ *Department of Mathematics, University of Denver, 2290 S. York St., Denver, CO 80208, USA*
(e-mail: rpavlov@du.edu)

(*Received* 18 *July* 2016 *and accepted in revised form* 3 *November* 2017)

Abstract. We define a new property called *one-sided almost specification*, which lies between the properties of specification and almost specification, and prove that it guarantees intrinsic ergodicity (i.e. uniqueness of the measure of maximal entropy) if the corresponding mistake function g is bounded. We also show that uniqueness may fail for unbounded g such as $\log\log n$. Our results have consequences for almost specification: we prove that almost specification with $g \equiv 1$ implies one-sided almost specification (with $g \equiv 1$) and hence uniqueness. On the other hand, the second author showed recently that almost specification with $g \equiv 4$ does not imply uniqueness.

1. *Introduction*

The notion of 'entropy' in dynamics can be defined as an invariant of a map preserving a probability measure (measure-theoretic entropy) and of a continuous map on a compact metric space (topological entropy). These are related by the variational principle, which states that the topological entropy of a topological dynamical system (X, T) is the supremum of the measure-theoretic entropies taken over all probability measures preserved by T. A measure on X that achieves this supremum is called a *measure of maximal entropy (MME)*. If there is a unique MME, the system (X, T) is said to be *intrinsically ergodic* [**20**].

Existence and uniqueness of MMEs is a central question in thermodynamic formalism, which relates ergodic theory and topological dynamics. It is often the case that the unique MME for an intrinsically ergodic system has strong statistical properties, and related thermodynamic considerations (equilibrium states for non-zero potentials) are connected to properties of 'physical measures' for smooth systems.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

We study intrinsic ergodicity in the context of symbolic dynamics over a finite alphabet, where existence of an MME is automatic, and so the real question is uniqueness. Intrinsic ergodicity for mixing subshifts of finite type (SFTs) was proved by Parry [**10**] using Perron–Frobenius theory; a different proof for subshifts with the *specification property* (which includes the class of mixing SFTs) was given by Bowen [**3**]. This property requires that any sequence of orbit segments can be shadowed by a single orbit that transitions from each segment to the next in uniformly bounded time; in symbolic dynamics, orbit segments correspond to *words* in the language of the shift.

Recently, a number of weakened versions of the specification property have been used to study various questions in ergodic theory. This includes *almost specification* [**8**, **11**, **13**, **14**, **18**], which allows one to concatenate arbitrary words in the language into a new word in the language if a 'small' number of letters are allowed to change in each word. The number of changes is controlled by a sublinear 'mistake function' $g(n)$; see §2 for a formal definition. Almost specification is enough to establish various results in large-deviation theory [**14**] and multifractal analysis [**13**, **18**], but it was unknown for some time whether almost specification also implies intrinsic ergodicity. This question was answered in the negative by (independently) [**9**] and [**11**]; in fact, [**11**, Theorem 1.2] shows that intrinsic ergodicity may fail even with the constant mistake function $g \equiv 4$. (Here and elsewhere, $g \equiv C$ means that g is the constant function C.)

One motivation for almost specification is the fact that many natural examples satisfy it for small g, such as the classical β-shifts, which have almost specification with $g \equiv 1$ and do satisfy intrinsic ergodicity. In fact, β-shifts satisfy a slightly stronger property; when one wishes to concatenate some words $w^{(1)}, \ldots, w^{(n)}$ together, it suffices to make the permitted number of changes to the words $w^{(1)}, \ldots, w^{(n-1)}$, and leave the final word $w^{(n)}$ untouched. Though this might seem like an incremental strengthening, it proves quite important. We call this stronger notion *left almost specification (LAS)*; our main result is that this property actually does imply intrinsic ergodicity if the mistake function g is bounded.

THEOREM 1.1. *If X is a subshift with left almost specification for a bounded mistake function, then it has a unique measure of maximal entropy μ. Moreover, μ is the limiting distribution of periodic orbits; finally, the system (X, σ, μ) is Bernoulli and has exponential decay of correlations and the central limit theorem for Hölder observables.*

We prove Theorem 1.1 in §3. This theorem covers the case when X has the usual specification property (see Lemma 2.15). We show in §4.1 that it also covers the case when X has the usual almost specification property with $g \equiv 1$, and deduce the following.

COROLLARY 1.2. *If X is a subshift with almost specification for the mistake function $g \equiv 1$, then it has a unique measure of maximal entropy μ, and this measure satisfies the conclusions of Theorem 1.1.*

Together with the result from [**11**] that $g \equiv 4$ is not enough to guarantee uniqueness, we are led to the following open problem.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Question 1.3. Does almost specification with mistake function $g \equiv 2$ or $g \equiv 3$ guarantee intrinsic ergodicity?

We observe that the approach of Corollary 1.2 does not immediately address Question 1.3, since once we allow $g > 1$ we can no longer deduce one-sided almost specification.

Returning to shifts with LAS, we observe that the requirement of bounded mistake function cannot be relaxed too far; we prove that even doubly logarithmic growth of g does not imply intrinsic ergodicity.

THEOREM 1.4. *There is a subshift X with LAS with $g(n) = 1 + 2\lfloor \log_2 \log_2 n \rfloor$ and multiple measures of maximal entropy (with disjoint supports).*

This immediately leads to the following open problem.

Question 1.5. Is there any unbounded function $g(n)$ such that LAS with mistake function g implies intrinsic ergodicity?

We omit details here, but, using the example of Theorem 1.4 as a starting point, one can use higher-power presentations just as in [**11**] to create examples satisfying LAS with $g(n) < C \log \log n$ and multiple MMEs for arbitrarily small C. So, the mistake functions g for which we do not know whether LAS implies uniqueness of the MME are those which are unbounded, but for which $g(n)/(\log \log n) \to 0$ as $n \to \infty$. See Remark 3.9 and §4.5 for the point at which our proof breaks down when g is unbounded.

The example from Theorem 1.4 has exactly two ergodic MMEs, whose supports are disjoint. It is actually possible for a subshift to have two ergodic MMEs that are *both* fully supported—this is the case, for example, with the Dyck shift. This leads to the following natural question.

Question 1.6. Is there a subshift X with LAS for some sublinear mistake function for which there are multiple fully supported ergodic MMEs?

The mechanism for uniqueness in Theorem 1.1 is a result from [**5**] that uses a certain non-uniform specification condition (superficially quite different from almost specification) to model X with a countable-state Markov shift. The conditions in [**5**] are mild variants of conditions introduced in [**6**], where they were used to prove that not only are β-shifts intrinsically ergodic, but so are their subshift factors. Although the conditions from [**6**] are well behaved under passing to factors, their behaviour upon passing to products is not immediately clear. On the other hand, as we will show, the hypotheses of Theorem 1.1 are stable under both products and factors.

PROPOSITION 1.7. *If X, Y are subshifts satisfying LAS with bounded g, then so is their product $X \times Y$. Similarly, if X is a subshift satisfying LAS with bounded g, and Z is a symbolic factor of X, then Z satisfies LAS with bounded g.*

Remark 1.8. Since β-shifts satisfy LAS with $g \equiv 1$, Proposition 1.7 implies that any shift created by taking iterated products/factors of β-shifts also has a unique MME. For example, given any β, β', we can define $X_{\beta,\beta'}$ to be the subshift consisting of all sequences created by coordinate-wise sums of points from the β-shifts X_β and $X_{\beta'}$. It is clear

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

that $X_{\beta,\beta'}$ is a factor of the direct product $X_\beta \times X_{\beta'}$ (induced by the letter-to-letter map $(i, j) \mapsto i + j$) and so every $X_{\beta,\beta'}$ also has a unique MME. It is quite possible that the methods of [**6**] or [**5**] could also be used to prove uniqueness of MME for these shifts, but it would be more difficult.

In §2 we give all the relevant definitions. Theorem 1.1 is proved in §3 and the remainder of the proofs are given in §4.

2. *Definitions and preliminaries*

2.1. *Symbolic dynamics.*

Definition 2.1. Let A be a finite set, which we call an *alphabet*. The *full shift* over A is the set $A^{\mathbb{Z}} = \{\cdots x(-1)\, x(0)\, x(1) \cdots : x(i) \in A\}$, which is a compact metric space with the metric ρ defined by $\rho(x, y) = 2^{-\min\{|k|:x(k)\neq y(k)\}}$ for $x \neq y$.

A *word* over A is an element of $A^* := \bigcup_{n=0}^{\infty} A^n$; here $A^0 = \{\varepsilon\}$, where ε is the so-called empty word. Given a word $w \in A^n$, we write $|w| = n$ for the *length* of the word. (We will also use $|S|$ to refer to cardinalities of finite sets; however, the usage should always be clear from context.) Given $x \in A^{\mathbb{Z}}$ and $i < j \in \mathbb{Z}$, we write $x([i, j]) = x(i)x(i + 1) \cdots x(j)$ for the word of length $j - i + 1$ that begins at position i and ends at position j.

Definition 2.2. The *shift action*, denoted by $\{\sigma^n\}_{n\in\mathbb{Z}}$, is the $\mathbb{Z}$-action on a full shift $A^{\mathbb{Z}}$ defined by $(\sigma^n x)(m) = x(m + n)$ for $m, n \in \mathbb{Z}$. A *subshift over* A is a closed subset of the full shift $A^{\mathbb{Z}}$ that is invariant under the shift action. Every subshift is a compact metric space with respect to the induced metric from $A^{\mathbb{Z}}$.

The shift $\sigma := \sigma^1$ is a homeomorphism on any subshift X, making (X, σ) a topological dynamical system. A subshift is characterized by its list of forbidden words $\mathcal{F} \subset A^*$: given any such $\mathcal{F}$, the set $X_{\mathcal{F}} := \{x \in A^{\mathbb{Z}} : x([i, j]) \notin \mathcal{F}\ \forall i, j \in \mathbb{Z}, i < j\}$ is closed and shift-invariant, and any subshift can be represented in this way.

Definition 2.3. The *language* of a subshift X, denoted by $\mathcal{L}(X)$, is the set of all words which appear in points of X. For any $n \in \mathbb{N}$, we write $\mathcal{L}_n(X) := \mathcal{L}(X) \cap A^n$ for the set of words in the language of X with length n. We will also need to deal with collections of words $\mathcal{D} \subset \mathcal{L}(X)$, and given such a collection we write $\mathcal{D}_n = \mathcal{D} \cap \mathcal{L}_n(X)$ for the set of words in $\mathcal{D}$ with length n.

A collection of words $\mathcal{D} \subset A^*$ is said to be *factorial* if it is closed under passing to subwords. The language $\mathcal{L}(X)$ of a subshift is factorial, but we will also have to deal with subsets $\mathcal{D} \subset \mathcal{L}(X)$ that are not factorial. Observe that if $\mathcal{D}$ is factorial, then

$$|\mathcal{D}_{m+n}| \leq |\mathcal{D}_m| \cdot |\mathcal{D}_n| \quad \text{for all } m, n \in \mathbb{N}. \tag{2.1}$$

Definition 2.4. For any subshift X and word $w \in \mathcal{L}_n(X)$, the *cylinder set* $[w]$ is the set of all $x \in X$ with $x([1, n]) = w$. The shift X is *irreducible* (or *topologically transitive*) if for any $u, v \in \mathcal{L}(X)$, there exists $n \in \mathbb{N}$ so that $[u] \cap \sigma^{-n}[v] \neq \varnothing$, or equivalently if there exists a word w so that $uwv \in \mathcal{L}(X)$.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

2.2. *Thermodynamic formalism for subshifts.* We recall some of the main elements of thermodynamic formalism as they appear in the context of subshifts; further details and results can be found in [**19**].

Definition 2.5. The *topological entropy* of a subshift X is

$$h(X) := \lim_{n\to\infty} \frac{1}{n} \ln |\mathcal{L}_n(X)| = \inf_{n\in\mathbb{N}} \frac{1}{n} \ln |\mathcal{L}_n(X)|, \tag{2.2}$$

where existence of the limit and the second equality follow from (2.1) and a standard lemma regarding subadditive sequences. We will also need to consider the entropy of a subset of the language: given $\mathcal{D} \subset \mathcal{L}(X)$, we write

$$h(\mathcal{D}) := \limsup_{n\to\infty} \frac{1}{n} \ln |\mathcal{D}_n|, \tag{2.3}$$

where in general the limit may not exist.

We will make occasional use of the following elementary fact:

$$h(\mathcal{C} \cup \mathcal{D}) = \max(h(\mathcal{C}), h(\mathcal{D})). \tag{2.4}$$

If $\mathcal{D} \subset A^*$ is factorial, then just as with $\mathcal{L}(X)$, (2.1) shows that the limit in (2.3) exists and is equal to $\inf_n (1/n) \ln |\mathcal{D}_n|$. The following consequence of this is useful enough to be worth stating formally.

LEMMA 2.6. *For any factorial $\mathcal{D} \subset A^*$, we have $|\mathcal{D}_n| \geq e^{nh(\mathcal{D})}$ for every $n \in \mathbb{N}$. In particular, for any subshift X, we have $|\mathcal{L}_n(X)| \geq e^{nh(X)}$.*

Given a factorial set of words $\mathcal{D} \subset A^*$, one can define a subshift $X(\mathcal{D})$ by the condition that $x \in X(\mathcal{D})$ if and only if every subword of x is in $\mathcal{D}$. The language of $X(\mathcal{D})$ is contained in $\mathcal{D}$, so clearly $h(X(\mathcal{D})) \leq h(\mathcal{D})$. However, we may have $\mathcal{L}(X(\mathcal{D})) \subsetneq \mathcal{D}$, since there could be words in $\mathcal{D}$ that do not appear as subwords of arbitrarily long words of $\mathcal{D}$. Accordingly, we will need the following result, which is proved in §4.4 and applies even when $\mathcal{L}(X(\mathcal{D}))$ is smaller than $\mathcal{D}$.

LEMMA 2.7. *Let $\mathcal{D}$ be a factorial set of words. Then $h(X(\mathcal{D})) = h(\mathcal{D})$.*

Now we recall some definitions from measure-theoretic dynamics. We write $\mathcal{M}(A^{\mathbb{Z}})$ for the space of σ-invariant Borel probability measures on the full shift, and similarly $\mathcal{M}(X)$ will denote the elements of $\mathcal{M}(A^{\mathbb{Z}})$ that give full weight to X.

Definition 2.8. The *measure-theoretic entropy* of $\mu \in \mathcal{M}(A^{\mathbb{Z}})$ is

$$h(\mu) := \lim_{n\to\infty} \frac{-1}{n} \sum_{w\in A^n} \mu([w]) \ln \mu([w]),$$

where terms with $\mu([w]) = 0$ are omitted from the sum. (Existence of the limit is a standard result and again uses subadditivity.)

The *variational principle* [**19**, Theorem 8.6] relates the two kinds of entropy: for every subshift X, we have $h(X) = \sup\{h(\mu) : \mu \in \mathcal{M}(X)\}$.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Definition 2.9. A *measure of maximal entropy (MME)* for a subshift X is an invariant measure $\mu \in \mathcal{M}(X)$ for which $h(\mu) = h(X)$. The shift X is said to be *intrinsically ergodic* if it has a unique MME.

We will need the following standard result on construction of measures with large (or full) entropy, which follows from the second half of the proof of [**19**, Theorem 8.6].

LEMMA 2.10. *Let X be a subshift and $n_k \to \infty$. Given $\mathcal{D}_{n_k} \subset \mathcal{L}_{n_k}(X)$ and any choice of $x_v \in [v]$ for each $v \in \mathcal{D}_{n_k}$, consider the measures $\nu_k = (1/|\mathcal{D}_{n_k}|) \sum_{v \in \mathcal{D}_{n_k}} \delta_{x_v}$ and $\mu_k = (1/n_k) \sum_{i=0}^{n_k-1} \sigma_*^i \nu_k$. Then any weak* limit point μ of the sequence μ_k is a σ-invariant measure with $h(\mu) \geq \liminf (1/n_k) \ln |\mathcal{D}_{n_k}|$. In particular, if $\liminf (1/n_k) \ln |D_{n_k}| = h(X)$, then any weak* limit point of the sequence μ_k is an MME for X.*

Our main goal is to prove that certain subshifts are intrinsically ergodic. A key intermediate step in our approach will be to establish the property of *entropy minimality*, first defined in [**7**].

Definition 2.11. A subshift X is *entropy minimal* if every non-empty proper subshift $X' \subsetneq X$ has topological entropy less than $h(X)$.

Two equivalent formulations are easily checked and will be useful later. Firstly, X is clearly entropy minimal if and only if for all $w \in \mathcal{L}(X)$, the subshift

$$X_w := \{x \in X : x \text{ does not contain } w \text{ as a subword}\}$$

has topological entropy less than $h(X)$. Also, X is entropy minimal if and only if all measures of maximal entropy of X are *fully supported* (meaning that every open set in X has positive measure). In particular, entropy minimality holds whenever X is intrinsically ergodic and the unique MME is fully supported. Such shifts include all mixing SFTs and, more generally, the class of shifts that we will study.

2.3. *Specification properties.* By upper semi-continuity of the entropy function $h\colon \mathcal{M}(X) \to [0, \ln |A|]$ (or by applying Lemma 2.10 with $\mathcal{D} = \mathcal{L}(X)$), every subshift X has at least one MME. We will be interested in the question of uniqueness, taking the following definition as our starting point.

Definition 2.12. A subshift X has the *specification* property if there is $\tau \in \mathbb{N}$ such that for every $v, w \in \mathcal{L}(X)$, there is $u \in \mathcal{L}(X)$ such that $|u| = \tau$ and $vuw \in \mathcal{L}(X)$.

It was shown by Bowen [**2**] that specification implies uniqueness of the MME†. More recently, various weakenings of the specification property have been introduced. The version we will focus on was introduced by Pfister and Sullivan [**14**] as the 'g-almost product property'; we will follow the convention of [**18**] and call this property *almost specification*.

In this definition and the remainder of the paper, the distance between two words v, w (which is only defined when $|v| = |w|$) is always assumed to be the *Hamming distance*

† Bowen's original definition is made in a more general setting and has a few more hypotheses, including periodicity of the 'glued' word; in the setting of subshifts it turns out that our definition is equivalent to his [**1**].

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

$d_H(v, w) := |\{i : v(i) \neq w(i)\}|$. We will frequently use the associated *Hamming balls* (of radius m)

$$B_m(w) := \{v \in \mathcal{L}_{|w|}(X) : d_H(v, w) \leq m\}. \tag{2.5}$$

Definition 2.13. A subshift X has *almost specification (or AS) with mistake function* $g(n)$ if:

- $g(n)/n \to 0$;
- for any words $w^{(1)}, w^{(2)}, \ldots, w^{(k)} \in \mathcal{L}(X)$, there exist words $\bar{w}^{(1)}, \bar{w}^{(2)}, \ldots, \bar{w}^{(k)} \in \mathcal{L}(X)$ so that $d_H(\bar{w}^{(i)}, w^{(i)}) \leq g(|w^{(i)}|)$ for $1 \leq i \leq k$, and the concatenation $\bar{w}^{(1)}\bar{w}^{(2)} \ldots \bar{w}^{(k)}$ is in $\mathcal{L}(X)$.

We will also consider the following slightly stronger property.

Definition 2.14. A subshift X has *left almost specification (or LAS) with mistake function* $g(n)$ if:

- $g(n)/n \to 0$;
- for any words $w^{(1)}, w^{(2)} \in \mathcal{L}(X)$, there exists a word $\bar{w}^{(1)}$ in $\mathcal{L}(X)$, where $d_H(\bar{w}^{(1)}, w^{(1)}) \leq g(|w^{(1)}|)$ and $\bar{w}^{(1)}w^{(2)} \in \mathcal{L}(X)$.

We first quickly demonstrate that specification does in fact imply LAS with bounded g.

LEMMA 2.15. *If X has specification, then it has LAS with bounded g.*

Proof. Suppose that X has specification; then there exists τ so that for all $v, w \in \mathcal{L}(X)$, there exists $u \in \mathcal{L}(X)$ with $|u| = \tau$ such that $vuw \in \mathcal{L}(X)$. We claim that X has LAS with $g \equiv \tau$. To see this, choose any $w^{(1)}, w^{(2)} \in \mathcal{L}(X)$. If $|w^{(1)}| \leq \tau$, then choose any word $\bar{w}^{(1)}$ with length $|w^{(1)}|$ which can precede $w^{(2)}$ in a point of X; then trivially $d_H(w^{(1)}, \bar{w}^{(1)}) \leq \tau$ and $\bar{w}^{(1)}w^{(2)} \in \mathcal{L}(X)$. If $|w^{(1)}| > \tau$, then define v to be the word obtained by removing the final τ letters of $w^{(1)}$. Then, by specification, there exists u with $|u| = \tau$ so that $vuw^{(2)} \in \mathcal{L}(X)$. Again $d_H(vu, w^{(1)}) \leq \tau$. In both cases, we found $\bar{w}^{(1)}$ with $d_H(w^{(1)}, \bar{w}^{(1)}) \leq \tau$ and $\bar{w}^{(1)}w^{(2)} \in \mathcal{L}(X)$, so X has LAS with $g \equiv \tau$. □

The reader may check that if X has LAS with $g(n)$, it also has LAS with $g'(n)$ defined by $g'(n) = \min\{g(k) : k \geq n\}$, and so we always assume without loss of generality that g is non-decreasing. There are many subshifts known to satisfy AS and/or LAS; for instance, any β-shift has LAS with gap function $g \equiv 1$ (see [**13**]); since β-shifts only have specification for a Lebesgue-null set of values of β (see [**4**, **16**]), this also demonstrates that the converse of Lemma 2.15 fails, and so LAS with bounded g is a more general property than specification. Many of the so-called S-gap shifts satisfy AS (with gap function dependent on S). Obviously there is a corresponding notion of *right almost specification (RAS)*, and all of our proofs carry over to that case in a standard way. (See the proof of Corollary 1.2 for more details.) So, though our results are stated for subshifts with LAS, they really apply to subshifts satisfying either type of 'one-sided' almost specification.

We observe that LAS implies AS.

LEMMA 2.16. *If a subshift has LAS with mistake function $g(n)$, then it has AS with mistake function $g(n)$.*

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Proof. We claim that LAS with mistake function $g(n)$ implies the following statement, which implies AS with mistake function $g(n)$: for any words $w^{(1)}, w^{(2)}, \ldots, w^{(k)} \in \mathcal{L}(X)$, there exist words $\bar{w}^{(1)}, \bar{w}^{(2)}, \ldots, \bar{w}^{(k-1)} \in \mathcal{L}(X)$ so that $d_H(\bar{w}^{(i)}, w^{(i)}) \leq g(|w^{(i)}|)$ for $1 \leq i \leq k-1$, and $\bar{w}^{(1)}\bar{w}^{(2)} \ldots \bar{w}^{(k-1)}w^{(k)}$ is in $\mathcal{L}(X)$. (Here, note that $\bar{w}^{(k)}$ was not defined; $w^{(k)}$ does not need to be changed in the concatenation.)

This is proved via induction on k. For $k = 2$, this is exactly the definition of LAS, so the base case is proved. If the statement above is true for k, then given any words $w^{(1)}, w^{(2)}, \ldots, w^{(k+1)} \in \mathcal{L}(X)$, first use LAS to define $\bar{w}^{(k)}$ with $d_H(\bar{w}^{(k)}, w^{(k)}) \leq g(|w^{(k)}|)$ so that $\bar{w}^{(k)}w^{(k+1)} \in \mathcal{L}(X)$, and then apply the inductive hypothesis to the k words $w^{(1)}, w^{(2)}, \ldots, w^{(k-1)}, \bar{w}^{(k)}w^{(k+1)}$ to yield the desired $\bar{w}^{(i)}$ for $1 \leq i \leq k-1$. □

Almost specification has been used in the literature to study statistical behaviour such as large deviations [**13**] and multifractal properties [**14**, **18**]. In [**6**], the first author and Thompson introduced a new non-uniform version of the specification property that guarantees intrinsic ergodicity, and asked whether or not almost specification could be used to prove intrinsic ergodicity. It was shown by the second author in [**11**] that almost specification does not imply intrinsic ergodicity, even for $g \equiv 4$. To prove Theorem 1.1, we will show that *left* almost specification with any *constant* mistake function implies a version of the non-uniform specification property from [**5**, **6**]. This property requires the existence of $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)$ such that:

[I] there exists $\tau \in \mathbb{N}$ such that for every $v, w \in \mathcal{G}$, there exists $u \in \mathcal{L}(X)$ such that $|u| = \tau$ and $vuw \in \mathcal{G}$ (we say that '$\mathcal{G}$ has specification') and, moreover, this u has the property that $v'uw' \in \mathcal{G}$ whenever $v' \in \mathcal{G}$ is a final segment of v and $w' \in \mathcal{G}$ is an initial segment of w;

[II] $h(\mathcal{C}^p \cup \mathcal{C}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s)) < h(X)$.

In light of [I], the collections $\mathcal{C}^p$, $\mathcal{C}^s$, and $\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s$ are thought of as *obstructions to specification*, since words in $\mathcal{C}^p\mathcal{G}\mathcal{C}^s$ can be glued together (as in the specification property) provided we are first allowed to remove an element of $\mathcal{C}^p$ from the front of the word, and an element of $\mathcal{C}^s$ from the end. Thus, [I]–[II] can be informally stated as the requirement that 'obstructions to specification have small entropy'. These conditions appeared in [**6**] (in a mildly different form), where a third condition was also required that controls how quickly words of the form uvw with $u \in \mathcal{C}^p$, $v \in \mathcal{G}$, $w \in \mathcal{C}^s$ with $|u|, |w| \leq M$ can be extended to words in $\mathcal{G}$. In our setting of LAS with bounded g, we produce $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s$ satisfying [I]–[II], but it is not clear whether the collections we produce satisfy this third condition, and so we cannot apply the results from [**6**]. Rather, we use the following conditions on $\mathcal{G}$ that were introduced in [**5**], which we are able to verify in our setting; roughly speaking, these ask that intersections and unions of words in $\mathcal{G}$ are again in $\mathcal{G}$ (under some mild conditions).

[III_a] There is $L \in \mathbb{N}$ such that if $x \in X$ and $i \leq j \leq k \leq \ell$ are such that $k - j \geq L$ and $x([i, k]), x([j, \ell]) \in \mathcal{G}$, then $x([j, k]) \in \mathcal{G}$.

[III_b] There is $L \in \mathbb{N}$ such that if $x \in X$ and $a \leq i \leq j \leq k \leq \ell$ are such that $k - j \geq L$ and $x([i, k]), x([j, \ell]), x([a, \ell]) \in \mathcal{G}$, then $x([i, \ell]) \in \mathcal{G}$.

It was shown in [**5**] that [I]–[III_b] guarantee intrinsic ergodicity, as well as strong statistical properties for the unique MME, which we describe next. We note that although

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

[**5**, Theorem 1.1] is stated using a stronger version of [III_a]–[III_b] (in which no assumption on $x([a, \ell])$ is given in [III_b]), the condition we use here is sufficient; the result we need is stated as Theorem 2.22 below, and proved in §3.4 using [**5**, Theorems 3.1(B) and 3.2].

2.4. *Statistical properties.* The unique MMEs that [**5**, Theorem 1.1] produces will have the following strong statistical properties.

Definition 2.17. A measure $\mu \in \mathcal{M}(X)$ satisfies an *upper Gibbs bound* if there is $Q_1 > 0$ such that $\mu([w]) \le Q_1 e^{-|w|h(X)}$ for all $w \in \mathcal{L}(X)$.

Definition 2.18. We say that $\mu \in \mathcal{M}(X)$ is the *limiting distribution of periodic orbits* if it is the weak* limit of the *periodic orbit measures*

$$\mu_n = \frac{1}{|\operatorname{Per}_n(X)|} \sum_{x \in \operatorname{Per}_n(X)} \delta_x, \tag{2.6}$$

where $\operatorname{Per}_n(X) = \{x \in X \mid \sigma^n x = x\}$ is the set of n-periodic points.

Definition 2.19. The *Bernoulli scheme* over a state space S with probability vector $p = (p_a)_{a \in S}$ is the measure-preserving system $(S^{\mathbb{Z}}, \sigma, \mu_p)$, where σ is the left shift map and $\mu_p([w]) = \prod_{i=1}^{|w|} p_{a_i}$ for every $w \in S^*$. We say that $\mu \in \mathcal{M}(X)$ has the *Bernoulli property* if (X, σ, μ) is measure-theoretically isomorphic to a Bernoulli scheme.

Definition 2.20. A measure $\mu \in \mathcal{M}(X)$ has *exponential decay of correlations* for Hölder observables if there is $\theta \in (0, 1)$ such that for every pair of Hölder continuous functions $\psi_1, \psi_2 \colon X \to \mathbb{R}$, there exists a constant $K(\psi_1, \psi_2) > 0$ such that

$$\left| \int (\psi_1 \circ \sigma^n) \psi_2 \, d\mu - \int \psi_1 \, d\mu \int \psi_2 \, d\mu \right| \le K(\psi_1, \psi_2) \theta^{|n|} \quad \text{for every } n \in \mathbb{Z}.$$

Definition 2.21. We say that $\mu \in \mathcal{M}(X)$ satisfies the *central limit theorem* for Hölder observables if for every Hölder continuous $\psi \colon X \to \mathbb{R}$ with $\int \psi \, d\mu = 0$, the quantity $(1/\sqrt{n}) S_n \psi = (1/\sqrt{n}) \sum_{k=0}^{n-1} \psi \circ \sigma^k$ converges in distribution to a normal distribution $\mathcal{N}(0, \sigma_\psi)$ for some $\sigma_\psi \ge 0$; that is, if

$$\lim_{n \to \infty} \mu \left\{ x \in X \;\middle|\; \frac{1}{\sqrt{n}} S_n \psi(x) \le \tau \right\} = \frac{1}{\sigma_\psi \sqrt{2\pi}} \int_{-\infty}^{\tau} e^{-t^2/(2\sigma_\psi^2)} \, dt$$

for every $\tau \in \mathbb{R}$. (When $\sigma_\psi = 0$, the convergence is to the Heaviside function.)

Recall that a function ψ is *cohomologous to a constant* if there are a measurable function $u \colon X \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that $\psi(x) = u(x) - u(\sigma x) + c$ for μ-almost every $x \in X$. It is typically true that in the central limit theorem the variance σ_ψ^2 is 0 if and only if ψ is cohomologous to a constant. This will hold for us as well.

2.5. *A uniqueness result with non-uniform specification.* To prove Theorem 1.1, we will need the following uniqueness result, which combines several theorems and remarks from [**5**]; see §3.4 for details.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

THEOREM 2.22. *Let X be a subshift on a finite alphabet and suppose that there are $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)$ satisfying [I], [II], [III_a], and [III_b]. Then X has a unique MME μ and μ satisfies the central limit theorem for Hölder observables.*

If in addition τ from [I] is such that $\gcd\{|w| + \tau : w \in \mathcal{G}\} = 1$, then μ is the limiting distribution of periodic orbits, the system (X, σ, μ) is Bernoulli, and it has exponential decay of correlations for Hölder observables.

3. *Proof of Theorem 1.1*

To prove our main result on LAS with bounded g, we need to begin by showing that even if X is not irreducible, we can pass to an irreducible subshift that has the same periodic points and invariant measures as X, and which retains the same LAS property as X. We do this in §3.1; then in §3.2 we show that irreducibility and LAS with bounded g are enough to imply entropy minimality (Theorem 3.8). Finally, in §3.3 we complete the proof of Theorem 1.1 by producing $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s$ that satisfy conditions [I]–[III_b] and hence allow us to apply the results on [**5**]. Entropy minimality is the crucial step in verifying condition [II].

3.1. *Irreducible subshifts.*

Intuitively, irreducibility might 'feel' like a weaker property than LAS with bounded g, since it places no restrictions on the distance required to concatenate two words. However, LAS does not quite imply irreducibility due to a simple degenerate case; there could be words in $\mathcal{L}(X)$ which can only appear finitely many times in points of X. For instance, take $X = \{x \in \{0, 1\}^{\mathbb{Z}} : |\{n : x(n) = 1\}| \le 1\}$, the set of 0–1 sequences with at most one 1. This is clearly not irreducible, but satisfies LAS with $g \equiv 1$; any word can be changed to a string of 0s with at most one change, and a string of 0s can legally precede any word in $\mathcal{L}(X)$.

To avoid such situations, we pass to a natural subshift: recall that the *measure center* $\hat{X} \subset X$ is the union of the supports of all invariant measures on X. This a closed shift-invariant subset of X, which has the same simplex of invariant measures as X by definition, and hence also has the same set of periodic points (since each periodic orbit supports a unique invariant measure).

THEOREM 3.1. *If X has LAS with mistake function $g(n)$, then its measure center $\hat{X}$ is irreducible and has LAS with mistake function $g(n)$.*

Proof. We start by using LAS to get a preliminary version of irreducibility.

LEMMA 3.2. *For every $u \in \mathcal{L}(\hat{X})$, there is $N = N_u \in \mathbb{N}$ such that for every $v \in \mathcal{L}(X)$, there is a word $z \in A^*$ such that $|z| \le N$ and $uzv \in \mathcal{L}(X)$.*

Proof. Since $u \in \mathcal{L}(\hat{X})$, there is an ergodic measure μ such that $\mu([u]) > 0$. Let $\epsilon = \frac{1}{2}\mu([u])$. By the ergodic theorem, there is $N_0 \in \mathbb{N}$ such that for all $N \ge N_0$, there is $w \in \mathcal{L}_N(X)$ with

$$\#\{1 \le k \le |w| - |u| : w([k, k + |u|)) = u\} \ge \epsilon N. \tag{3.1}$$

Since $g(n)/n \to 0$, we may choose $N \ge N_0$ such that $g(N) < \epsilon N/|u|$. Fix $w \in \mathcal{L}_N(X)$ satisfying (3.1). Then every $\bar{w} \in B_{g(|w|)}(w)$ has u as a subword, since changing $g(|w|)$ symbols destroys at most $|u|g(|w|) = |u|g(N) < \epsilon N$ occurrences of u; thus, every such

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

$\bar{w}$ has the form yuz for some $y, z \in A^*$. By LAS for X there is some such $\bar{w}$ with $\bar{w}v = yuzv \in \mathcal{L}(X)$ and thus $uzv \in \mathcal{L}(X)$, which proves the lemma. □

Now we use Lemma 3.2 iteratively to prove Theorem 3.1. Given $u, v \in \mathcal{L}(\hat{X})$, let $N = \max\{N_u, N_v\}$. Use LAS to get $\bar{u}v \in \mathcal{L}(X)$ for some $\bar{u} \in B_{g(|u|)}(u)$, and apply Lemma 3.2 to the words u and $\bar{u}v$ to get $z^{(1)} \in A^*$ such that $|z^{(1)}| \le N$ and $uz^{(1)}\bar{u}v \in \mathcal{L}(X)$. Then apply Lemma 3.2 to the words v and $uz^{(1)}\bar{u}v$ to obtain $y^{(1)} \in A^*$ such that $|y^{(1)}| \le N$ and $vy^{(1)}uz^{(1)}\bar{u}v$. Iterating, we obtain $y^{(i)}, z^{(i)} \in A^*$ with lengths $\le N$ such that

$$w^{(j)} := vy^{(j)}uz^{(j)}\bar{u}vy^{(j-1)}uz^{(j-1)}\bar{u}v \cdots vy^{(1)}uz^{(1)}\bar{u}v \in \mathcal{L}(X) \quad \text{for all } j \in \mathbb{N}.$$

By compactness, there is $x \in X$ such that

$$x_{(-\infty,0]} = \cdots vy^{(j)}uz^{(j)}\bar{u}v \cdots vy^{(1)}uz^{(1)}\bar{u}v.$$

Let $\mu_n = (1/n)\sum_{k=0}^{n-1} \delta_x \circ \sigma^k$. Each $z^{(j)}$ is chosen from the finite set $\bigcup_{n=0}^{N} A^n$, so there is $z \in \bigcup_{n=0}^{N} A^n$ for which $\{j \in \mathbb{N} : z^{(j)} = z\}$ has positive upper density, and thus there is some subsequence $n_i \to \infty$ such that $\liminf_i \mu_{n_i}([uz\bar{u}v]) > 0$. By passing to a further subsequence if necessary, we can consider the weak* limit μ of the sequence μ_{n_i}, which is σ-invariant by standard arguments and has $\mu([uz\bar{u}v]) > 0$ by our choice of n_i. We conclude that $uz\bar{u}v \in \mathcal{L}(\hat{X})$, implying irreducibility. Clearly, the subword $\bar{u}v \in \mathcal{L}(\hat{X})$ as well, which implies LAS and completes the proof of Theorem 3.1. □

Remark 3.3. The proof of Lemma 3.2 shows that for every $u \in \mathcal{L}(\hat{X})$, there is a word $w \in \mathcal{L}(X)$ that contains at least $g(|w|) + 1$ disjoint instances of u as subwords, and the paragraphs below that lemma show that this property in fact characterizes words in $\mathcal{L}(\hat{X})$. In particular, when $g \equiv m$, we can write $\mathcal{O}$ for the set of words $u \in \mathcal{L}(X)$ such that no point of X contains $m + 1$ disjoint occurrences of u, and conclude that $\hat{X}$ is the subshift of points of X containing no word from $\mathcal{O}$.

3.2. *Entropy minimality.* Before proving that irreducibility and LAS with bounded g are enough to obtain entropy minimality, we need to establish some counting estimates. Recall from Lemma 2.6 that $|\mathcal{L}_n(X)| \ge e^{nh(X)}$ for every n. In general the definition of $h(X)$ gives the upper bound $|\mathcal{L}_n(X)| \le C_n e^{nh(X)}$ for some subexponential C_n. An important part of the uniqueness proof in [**3**, **6**] is to prove that C_n can be taken to be bounded. At this stage of our proof we do not yet get quite this bound, but we can prove that C_n grows at most polynomially for subshifts with almost specification; note that by Lemma 2.16, the same is true for subshifts with LAS.

LEMMA 3.4. *If X is a subshift with almost specification with constant $g \equiv m$, then, for every n,*

$$|\mathcal{L}_n(X)| \le |A|^{2m} n^{2m} e^{nh(X)}. \tag{3.2}$$

Proof. Suppose that X is a subshift satisfying AS with $g \equiv m$. Choose a maximal $(2m + 1)$-separated subset $S_n \subset \mathcal{L}_n(X)$ with respect to Hamming distance; since the cardinality of a Hamming ball of radius $2m$ is clearly bounded from above by $|A|^{2m} n^{2m}$, we have

$$|S_n| \ge |\mathcal{L}_n(X)| \cdot |A|^{-2m} n^{-2m}. \tag{3.3}$$

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

For any $k \in \mathbb{N}$, almost specification lets us define $f\colon (S_n)^k \to \mathcal{L}_{kn}(X)$ by $f(w^{(1)}, \ldots, w^{(k)}) = \bar{w}^{(1)} \ldots \bar{w}^{(k-1)}\bar{w}^{(k)} \in \mathcal{L}_{kn}(X)$, where $\bar{w}^{(i)} \in B_m(w^{(i)})$ for $1 \le i \le k$. Then the map f is clearly injective, since for any $s \ne s'$ in S, the Hamming balls $B_m(s)$ and $B_m(s')$ are disjoint. Using (3.3), this gives

$$|\mathcal{L}_{kn}(X)| \ge |S_n|^k \ge |\mathcal{L}_n(X)|^k |A|^{-2mk} n^{-2mk} \quad \text{and so}$$
$$\frac{1}{kn} \ln |\mathcal{L}_{kn}(X)| \ge \frac{1}{n} \ln |\mathcal{L}_n(X)| - \frac{2m}{n} \ln(|A|n).$$

Sending $k \to \infty$ gives $(1/n) \ln |\mathcal{L}_n(X)| \le h(X) + (2m/n) \ln(|A|n)$; then multiplying by n and taking exponentials gives (3.2). □

The following technical theorem will be a main tool in our proof that LAS and irreducibility imply entropy minimality, and may be of some independent interest. It shows that shifts with entropy minimality satisfy a sort of weakened Gibbs counting bound on the number of words that end with a given word w, as long as the overall word complexity function satisfies a polynomial upper bound of the sort just proved.

THEOREM 3.5. *Suppose that X is a subshift for which there exist $C, r > 0$ so that $|\mathcal{L}_n(X)| \le Cn^r e^{nh(X)}$ for every n, and that μ is a measure of maximal entropy on X. Then, for every $w \in \mathcal{L}(X)$ with $\mu([w]) > 0$, there exists $\epsilon > 0$ so that $|\mathcal{L}_n(X) \cap \mathcal{L}(X)w| > \epsilon n^{-1/2} e^{nh(X)}$ for all $n \ge |w|$.*

Proof. Suppose that X, C, r, μ, and w are as in the theorem, and denote by A the alphabet of X. By ergodic decomposition, we may assume without loss of generality that μ is ergodic. Use the ergodic theorem to choose N so that $\mu\big(\bigcup_{i=0}^{N-1} \sigma^i[w]\big) > 2r/(2r+1)$.

Then, for each $n \ge N + |w|$, define $S_n = \mathcal{L}_n(X) \cap \bigcup_{i=n-|w|-N+1}^{n-|w|} \sigma^{-i}[w]$, the set of n-letter words in $\mathcal{L}(X)$ which contain a w somewhere in the last $|w| + N - 1$ letters. We use the notation $\mu(S_n)$ to refer to $\mu(\bigcup_{v \in S_n}[v])$; then $\mu(S_n) > 2r/(2r+1)$ for all $n > N + |w|$. Note that this implies that $(1 - \mu(S_n))/\mu(S_n) < 1/2r$.

Now we will apply Theorem 4.7 from [**12**]. That result is phrased in terms of general expansive systems (not just subshifts) and for topological pressure for a general potential ϕ (not just topological entropy). However, once restricting to our setting (by setting $\phi = 0$), it says that there exists a constant M so that

$$|S_n| \ge M(e^{nh(X)})^{1/\mu(S_n)} |\mathcal{L}_n(X)|^{1-(1/\mu(S_n))} 2^{-1/\mu(S_n)}.$$

Since $|\mathcal{L}_n(X)| \le Cn^r e^{nh(X)}$ and μ is an MME on X, this implies that

$$\begin{aligned} |S_n| &\ge M e^{nh(X)} (Cn^r)^{-(1-\mu(S_n))/\mu(S_n)} 2^{-1/\mu(S_n)} \\ &> M e^{nh(X)} (Cn^r)^{-1/(2r)} 2^{-(2r+1)/2r} = \epsilon' n^{-1/2} e^{nh(X)} \end{aligned} \tag{3.4}$$

for $\epsilon' = MC^{-1/(2r)} 2^{-(2r+1)/2r}$.

We now note that since each word in S_n ends with w, followed by a word of length less than N, we have

$$|S_n| \le \sum_{i=0}^{N-1} |\mathcal{L}_{n-i}(X) \cap \mathcal{L}(X)w| |A|^i \le N|A|^{N-1} |\mathcal{L}_n(X) \cap \mathcal{L}(X)w|. \tag{3.5}$$

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

(For the inequality, note that every word in $\mathcal{L}_{n-i}(X) \cap \mathcal{L}(X)w$ can be extended on the left to a word in $\mathcal{L}_n(X) \cap \mathcal{L}(X)w$.) Now combining (3.4) and (3.5) yields

$$|\mathcal{L}_n(X) \cap \mathcal{L}(X)w| > \frac{\epsilon'}{N|A|^{N-1}} n^{-1/2} e^{nh(X)}. \tag{3.6}$$

The desired bound has now been shown for $n \geq N + |w|$ (and $\epsilon = \epsilon'/(N|A|^{N-1})$) and, since $|\mathcal{L}_n(X) \cap \mathcal{L}(X)w| > 0$ for the finite set of n in $[|w|, N + |w|)$, such a bound trivially still holds for $n \geq |w|$. □

Remark 3.6. In fact Theorem 3.5 could be generalized by using any upper bound of the form $|\mathcal{L}_n(X)| \leq f(n)e^{nh(X)}$ as a hypothesis and deriving, for any k, a lower bound of the form $|\mathcal{L}_s(X) \cap \mathcal{L}(X)w| > \epsilon f(n)^{-1/k} e^{nh(X)}$ as a conclusion (in the choice of N at the start of the proof, replace $2r$ with k). In particular, we note that we could bound $|\mathcal{L}_n(X) \cap \mathcal{L}(X)w|$ from below by some constant multiple of $n^{-\beta} e^{nh(X)}$ for any desired $\beta > 0$; for our purposes, $\beta = 1/2$ will suffice.

The following corollary is immediate, since every MME for an entropy minimal subshift is fully supported.

COROLLARY 3.7. *Suppose that X is an entropy minimal subshift and that there exist $C, r > 0$ so that $|\mathcal{L}_n(X)| \leq Cn^r e^{nh(X)}$ for every n. Then, for every $w \in \mathcal{L}(X)$, there exists $\epsilon > 0$ so that $|\mathcal{L}_n(X) \cap \mathcal{L}(X)w| > \epsilon n^{-1/2} e^{nh(X)}$ for all $n \geq |w|$.*

Now we can prove that LAS (with bounded g) and irreducibility together imply entropy minimality.

THEOREM 3.8. *If X is an irreducible subshift with left almost specification with constant $g \equiv m$, then X is entropy minimal.*

Remark 3.9. This result fails if g is unbounded; see §4.5. This is the point at which our proof breaks down for unbounded mistake functions.

Proof of Theorem 3.8. Take X a subshift as in the theorem. By [**15**, Proposition 2.8], X contains an entropy minimal subshift Y with $h(Y) = h(X)$. We suppose for a contradiction that Y is a proper subset of X and so there exists $u \in \mathcal{L}(X) \setminus \mathcal{L}(Y)$. Define i to be the maximum number of changes necessary to a word in $\mathcal{L}(Y)$ when using left almost specification to append a word of $\mathcal{L}(X)$ to its right and yield a word in $\mathcal{L}(X)$, i.e.

$$i = \max_{v \in \mathcal{L}(Y), w \in \mathcal{L}(X)} \min\{j : \exists v' \in B_j(v) \text{ for which } v'w \in \mathcal{L}(X)\}.$$

Clearly, $i \leq m$ since X has LAS with $g \equiv m$. Choose words $v \in \mathcal{L}(Y)$ and $w \in \mathcal{L}(X)$ achieving the maximum i, i.e. for $j < i$, there exists no $y \in B_j(v)$ for which $yw \in \mathcal{L}(X)$.

By Lemma 3.4, there exist $C, r > 0$ such that for all n, we have $|\mathcal{L}_n(X)| \leq Cn^r e^{nh(X)}$. Since $Y \subseteq X$, the same bound holds for $|\mathcal{L}_n(Y)|$ and, since Y is entropy minimal, we can apply Corollary 3.7 to Y. Since $h(X) = h(Y)$, this yields $\epsilon > 0$ so that for all $n \geq |v|$,

$$M_n := |\mathcal{L}_n(Y) \cap \mathcal{L}(Y)v| > \epsilon n^{-1/2} e^{nh(Y)} = \epsilon n^{-1/2} e^{nh(X)}, \tag{3.7}$$

where v is the 'maximally changing word' from above.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

By irreducibility of X, define $z \in \mathcal{L}(X)$ containing $m+1$ disjoint occurrences of w, followed by $m+1$ disjoint occurrences of u. Denote by N the length of z and assume without loss of generality that $N \geq |v|$. Then, by (3.7), for all $i \in \mathbb{N}$, $|M_{iN}| > \epsilon(iN)^{-1/2}e^{iNh(X)}$. Since $t!$ grows superexponentially, we may fix t such that $\sqrt{t!} > 2(\sqrt{N}e^{Nh(X)}/\epsilon)^t$. Write $M = \sum_{i=1}^{t} iN$.

We will make many words in $\mathcal{L}(X)$ by using left almost specification to almost concatenate words in various $\mathcal{L}_{iN}(Y)$ with copies of z in an alternating fashion. Specifically, for any k, we create words in $\mathcal{L}_{k(M+tN)}(X)$ in the following way:

(a) choose any permutations $\pi_1, \pi_2, \ldots, \pi_k$ of $\{1, \ldots, t\}$ and words $w^{(i,j)} \in \mathcal{L}_{iN}(Y) \cap \mathcal{L}(Y)v$ for $1 \leq i \leq t$ and $1 \leq j \leq k$;

(b) define words $u^{(\ell)}$ for $1 \leq \ell \leq kt$ by $u^{(i+t(j-1))} = w^{(\pi_j(i),j)}$ for $1 \leq i \leq t$ and $1 \leq j \leq k$, so the list $u^{(1)}, u^{(2)}, \ldots, u^{(kt)}$ consists of the words $w^{(i,1)}$ permuted according to π_1, followed by the words $w^{(i,2)}$ permuted according to π_2, and so on;

(c) given any $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ and $\mathbf{w} = (w^{(i,j)})$, define a word $f(\boldsymbol{\pi}, \mathbf{w}) \in \mathcal{L}_{k(M+tN)}(X)$ using LAS iteratively on $u^{(1)}, \ldots, u^{(\ell)}$ from right to left, as described in detail below.

After defining $f(\boldsymbol{\pi}, \mathbf{w})$, we will show that the map f is injective, which leads to a lower bound on $\mathcal{L}_{k(M+tN)}(X)$ and hence a lower bound on $h(X)$. By an appropriate choice of the parameter t, we will be able to show that $h(X) > h(Y)$ and hence obtain a contradiction.

Now we define $f(\boldsymbol{\pi}, \mathbf{w}) \in \mathcal{L}_{k(M+tN)}(X)$. Begin with $u^{(tk)}$ and then use left almost specification to find $z^{(tk-1)} \in B_m(z)$ for which $z^{(tk-1)}u^{(tk)} \in \mathcal{L}(X)$. Since z contained $m+1$ occurrences of w, at least one of them remains; delete the portion to the left of the leftmost remaining occurrence out of those to create a word $\bar{z}^{(tk-1)}$ with length less than or equal to N. Then:

- $\bar{z}^{(tk-1)}u^{(tk)} \in \mathcal{L}(X)$;
- $\bar{z}^{(tk-1)}$ begins with w; and
- $\bar{z}^{(tk-1)}$ contains u (since z had $m+1$ disjoint instances of u, and the part deleted to get $\bar{z}^{(tk-1)}$ from $z^{(tk-1)}$ was to the left of all of them).

Extend $u^{(tk-1)}$ on the left by $N - |\bar{z}^{(tk-1)}|$ letters to make a new word $\bar{u}^{(tk-1)} \in \mathcal{L}(Y)$, which still ends with v. Use left almost specification to find $U^{(tk-1)} \in B_i(\bar{u}^{(tk-1)})$ for which $U^{(tk-1)}\bar{z}^{(tk-1)}u^{(tk)} \in \mathcal{L}(X)$; we needed less than or equal to i changes to $\bar{u}^{(tk-1)}$ since $\bar{u}^{(tk-1)} \in \mathcal{L}(Y)$. However, recall that $\bar{u}^{(tk-1)}$ ends with v and $\bar{z}^{(tk-1)}$ begins with w, and that at least i changes are required to v to concatenate w on its right. Therefore, all i of the changes made in changing $\bar{u}^{(tk-1)}$ to $U^{(tk-1)}$ occurred within the terminal copy of v.

Observe that the length of $U^{(tk-1)}\bar{z}^{(tk-1)}u^{(tk)}$ is equal to $|u^{(tk-1)}| + N + |u^{(tk)}|$; in other words, even though $\bar{z}^{(tk-1)}$ may have length smaller than N, $U^{(tk-1)}$ 'corrects' the loss so that the length of $U^{(tk-1)}\bar{z}^{(tk-1)}u^{(tk)}$ is forced. Continuing in this fashion, we eventually arrive at a word

$$f(\boldsymbol{\pi}, \mathbf{w}) = U^{(1)}\bar{z}^{(1)}U^{(2)}\bar{z}^{(2)} \ldots U^{(tk-1)}\bar{z}^{(tk-1)}u^{(tk)} \in \mathcal{L}_{k(M+tN)}(X). \tag{3.8}$$

At each step, let us say that we choose the lexicographically minimal possible word to append, so that the procedure described is deterministic.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

We claim that f is injective. Suppose that $(\boldsymbol{\pi}_1, \mathbf{w}_1) \neq (\boldsymbol{\pi}_2, \mathbf{w}_2)$, that $(\boldsymbol{\pi}_1, \mathbf{w}_1)$ induces $u_1^{(1)}, \dots, u_1^{(tk)}$ as in 1 above, and that $u_2^{(1)}, \dots, u_2^{(tk)}$ is similarly induced by $(\boldsymbol{\pi}_2, \mathbf{w}_2)$. We break into two cases.

Case 1: $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_2$. In this case, $|u_1^{(\ell)}| = |u_2^{(\ell)}|$ for $1 \le \ell \le tk$. We may then choose maximal ℓ so that $u_1^{(\ell)} \neq u_2^{(\ell)}$ (such an ℓ must exist since $(\boldsymbol{\pi}_1, \mathbf{w}_1) \neq (\boldsymbol{\pi}_2, \mathbf{w}_2)$).

During the creation of $f(\boldsymbol{\pi}_1, \mathbf{w}_1)$ and $f(\boldsymbol{\pi}_2, \mathbf{w}_2)$, the first $tk - \ell - 1$ steps will be identical and will yield the same word $\bar{z}^{(\ell+1)} U^{(\ell+1)} \dots \bar{z}^{(tk-1)} u^{(tk)}$. Then, at the next step, the unequal words $u_1^{(\ell)}$ and $u_2^{(\ell)}$ are extended to the left by $N - |\bar{z}^{(\ell+1)}|$ units, which yields unequal words $\bar{u}_1^{(\ell)}$ and $\bar{u}_2^{(\ell)}$, and then left almost specification is used to find $U_1^{(\ell)}$ and $U_2^{(\ell)}$, which can be appended to the left of $\bar{z}^{(\ell+1)} U^{(\ell+1)} \dots \bar{z}^{(tk-1)} u^{(tk)}$. However, as noted above, the only changes made are inside the terminal copies of v within $\bar{u}_1^{(\ell)}$ and $\bar{u}_2^{(\ell)}$ and so the appended words $U_1^{(\ell)}$ and $U_2^{(\ell)}$ are still unequal, meaning that the final words $f(\boldsymbol{\pi}_1, \mathbf{w}_1)$ and $f(\boldsymbol{\pi}_2, \mathbf{w}_2)$ are also unequal.

Case 2: $\boldsymbol{\pi}^{(1)} \neq \boldsymbol{\pi}^{(2)}$. In this case, there must exist ℓ so that $|u_1^{(\ell)}| \neq |u_2^{(\ell)}|$; take ℓ to be maximal with this property. Without loss of generality, assume that $|u_1^{(\ell)}| > |u_2^{(\ell)}|$. By assumption on S, $|u_2^{(\ell)} - u_1^{(\ell)}| \ge N$. Again, in the process of creating $f(\boldsymbol{\pi}_1, \mathbf{w}_1)$ and $f(\boldsymbol{\pi}_2, \mathbf{w}_2)$, the first $tk - \ell - 1$ steps will be identical, yielding the same word $\bar{z}^{(\ell+1)} U^{(\ell+1)} \dots \bar{z}^{(tk-1)} u^{(tk)}$.

Then, in the construction of $f(\boldsymbol{\pi}_2, \mathbf{w}_2)$, once $U_2^{(\ell)}$ is appended to the left of $\bar{z}^{(\ell+1)} U^{(\ell+1)} \dots \bar{z}^{(tk-1)} u^{(tk)}$, somewhere in the N letters immediately to the left an occurrence of u will be added at the next step (inside some $\bar{z}_2^{(\ell)} \in B_m(z)$). However, the corresponding locations in $U_1^{(\ell)} \bar{z}^{(\ell+1)} U^{(\ell+1)} \dots \bar{z}^{(tk-1)} u^{(tk)}$ are part of $U_1^{(\ell)} \in \mathcal{L}(Y)$ since $|u_1^{(\ell)}| \ge |u_2^{(\ell)}| + N$ and so contain no occurrences of u since $u \notin \mathcal{L}(Y)$. Thus, there is a location at which $f(\boldsymbol{\pi}_2, \mathbf{w}_2)$ contains a u and $f(\boldsymbol{\pi}_1, \mathbf{w}_1)$ does not, meaning that they are unequal.

We have shown that f is one-to-one and so it generates a set within $\mathcal{L}_{k(M+tN)}(X)$ with cardinality at least $(t! \prod_{i=1}^{t} M_{iN})^k$, where we recall that M_{iN} was defined in (3.7) and satisfies $M_{iN} \ge \epsilon (iN)^{-1/2} e^{iNh(X)}$. Thus,

$$\begin{aligned} |\mathcal{L}_{k(M+tN)}(X)| &\ge \left(t! \prod_{i=1}^{t} M_{iN} \right)^k \ge \left(t! \prod_{i=1}^{t} \epsilon (iN)^{-1/2} e^{iNh(X)} \right)^k \\ &= (e^{Mh(X)} \sqrt{t!} (\epsilon/\sqrt{N})^t)^k > 2^k e^{k(M+tN)h(X)}. \end{aligned} \tag{3.9}$$

(Recall that t was chosen so that $\sqrt{t!} > 2(\sqrt{N} e^{Nh(X)}/\epsilon)^t$.)

Now we take logarithms in (3.9), divide by $k(M + tN)$, and let $k \to \infty$, which implies that $h(X) \ge h(Y) + \log 2/(M + tN)$, contradicting the earlier statement that $h(Y) = h(X)$. Therefore, our original assumption that $Y \neq X$ was false and we conclude that $Y = X$; since Y was entropy minimal, we have shown that X is entropy minimal. □

3.3. *Completion of the proof.* Now we use the results from §§3.1 and 3.2 to complete the proof of Theorem 1.1. By Theorem 3.1, we may assume without loss of generality that

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

X is irreducible, since all desired properties are dependent only on periodic points of X and measures on X. Define any m so that $g(n) \leq m$ for all n.

Define i to be the maximum number of changes necessary to a word in $\mathcal{L}(X)$ when using left almost specification to append a word of $\mathcal{L}(X)$ to its right and yield a word in $\mathcal{L}(X)$, i.e.

$$i = \max_{y \in \mathcal{L}(X), v \in \mathcal{L}(X)} \min\{j : \exists y' \in B_j(y) \text{ for which } y'v \in \mathcal{L}(X)\}. \tag{3.10}$$

We have $i \leq m$ from the LAS property. Choose words $y \in \mathcal{L}(X)$ and $v^{(0)} \in \mathcal{L}(X)$ achieving this maximum, i.e. for $j < i$, there exists no $y' \in B_j(y)$ for which $y'v^{(0)} \in \mathcal{L}(X)$. Let $\mathcal{L}^y = \{w \in \mathcal{L} : wy \in \mathcal{L}\}$. Given $w \in \mathcal{L}^y$ and $v \in \mathcal{L}$, let $D(w, v) = \{y' \in B_i(y) : wy'v \in \mathcal{L}\}$; this is non-empty because LAS gives $x \in B_i(wy)$ such that $xv \in \mathcal{L}$ and, by our choice of y and v, all the changes between wy and x must happen in the last $|y|$ symbols, so $x = wy'$ for some $y' \in B_i(y)$.

Observe that if $v' \in v\mathcal{L} \cap \mathcal{L}$ and $w' \in \mathcal{L}w \cap \mathcal{L}^y$, then $D(w', v') \subset D(w, v)$. Now define a sequence of words $w^{(k)}$, $v^{(k)}$ for $k \geq 1$ iteratively as follows: if there are $w' \in \mathcal{L}w^{(k-1)} \cap \mathcal{L}^y$ and $v' \in v^{(k-1)}\mathcal{L} \cap \mathcal{L}$ such that $D(w', v') \neq D(w^{(k-1)}, v^{(k-1)})$, then let $w^{(k)} = w'$ and $v^{(k)} = v'$; if there are no such v', w', terminate the construction. Because the sets $D(w^{(k)}, v^{(k)})$ are finite, nested, and non-empty, the construction terminates at some k; putting $v = v^{(k)}$ and choosing $y' \in D(w^{(k)}, v)$, we see that for every $x \in \mathcal{L}^y \cap \mathcal{L}w^{(k)}$ and $z \in \mathcal{L} \cap v\mathcal{L}$, we have $xy'z \in \mathcal{L}$. We will write $u = w^{(k)}y$, $u' = w^{(k)}y'$, and use this in the following form:

$$\text{if } xu \in \mathcal{L} \quad \text{and} \quad z \in v\mathcal{L} \cap \mathcal{L} \quad \text{then } xu'z \in \mathcal{L}. \tag{3.11}$$

Consider the collection

$$\mathcal{G} = \{w \in \mathcal{L}(X) \cap v\mathcal{L}(X) : wu \in \mathcal{L}(X)\}$$

of words in $\mathcal{L}(X)$ starting with v which can be legally followed by u.

To apply the results from [**5**], we define the 'prefix' and 'suffix' collections

$$\mathcal{C}^p = \{w \in \mathcal{L}(X) : w \text{ does not contain } v \text{ as a subword}\},$$
$$\mathcal{C}^s = \{w \in \mathcal{L}(X) \cap u\mathcal{L}(X) : w \text{ contains } u \text{ only once as a subword}\}.$$

Let $\mathcal{C}^p\mathcal{G}\mathcal{C}^s$ denote the set of words that can be decomposed as a concatenation of a word in $\mathcal{C}^p$, followed by a word in $\mathcal{G}$, followed by a word in $\mathcal{C}^s$, and let $\mathcal{B} := \mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s$ be the set of words that do not admit such a decomposition. Finally, define $\mathcal{C} = \mathcal{C}^p \cup \mathcal{C}^s \cup \mathcal{B}$.

Condition [I] says that $\mathcal{G}$ has specification and follows from our choice of $\mathcal{G}$ and (3.11). Given any w, $w' \in \mathcal{G}$, we have wu, $w'u \in \mathcal{L} \cap v\mathcal{L}$ by the definition of $\mathcal{G}$ and so (3.11) gives $wu'w'u \in \mathcal{L}$. Moreover, $wu'w'$ begins with v (since $w \in \mathcal{G}$) and so $wu'w' \in \mathcal{G}$, which suffices to prove [I] since the 'gluing word' u' does not depend on the choice of w, $w' \in \mathcal{G}$.

Now we prove [II] by showing that $h(\mathcal{C}) < h(X)$. First, we note that by Theorem 3.8, X is entropy minimal. If we define X_v to be the subshift of points of X containing no v as a subword, then, by entropy minimality, $h(X_v) < h(X)$. Since $\mathcal{C}^p$ is clearly factorial, $h(\mathcal{C}^p) = h(X_v)$ by Lemma 2.7. Therefore, $h(\mathcal{C}^p) < h(X)$. To treat $\mathcal{C}^s$, note that if we define

$$\mathcal{C}' = \{w \in \mathcal{L}(X) : w \text{ does not contain } u \text{ as a subword}\},$$

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

then removing u from the beginning of words in $\mathcal{C}^s$ places $(\mathcal{C}^s)_n$ in bijective correspondence with $(\mathcal{C}')_{n-|u|}$ for $n > |u|$. Then clearly $h(\mathcal{C}^s) = h(\mathcal{C}')$. If we define X_u to be the subshift of points of X containing no u as a subword, then, as above, by entropy minimality, $h(X_u) < h(X)$. Since $\mathcal{C}'$ is clearly factorial, Lemma 2.7 implies that $h(\mathcal{C}') \le h(X_u)$ and so $h(\mathcal{C}^s) < h(X)$. It remains only to consider $\mathcal{B}$. If w contains disjoint occurrences of v and u where the v occurs to the left of the u, then we could write $w = xvyuz$, where x contains no v and z contains no u. Then $vy \in \mathcal{G}$, $x \in \mathcal{C}^p$, and $uz \in \mathcal{C}^s$, meaning that $w \notin \mathcal{B}$. We have then shown that words in $\mathcal{B}$ either contain no u at all or can be written as xuy, where x contains no v and y contains no u. This clearly implies that

$$|\mathcal{B}_n| \le |(\mathcal{C}')_n| + \sum_{i=0}^{n-|u|} |(\mathcal{C}^p)_i| \cdot |(\mathcal{C}')_{n-i}|. \tag{3.12}$$

Choose $\epsilon > 0$ so that $h(\mathcal{C}^p), h(\mathcal{C}') < h(X) - 2\epsilon$. Then there is K such that $|\mathcal{C}^p_n|, |\mathcal{C}'_n| \le K e^{n(h(X)-\epsilon)}$ for every $n \in \mathbb{N}$. Combining this with (3.12) yields

$$\begin{aligned} |\mathcal{B}_n| &\le K e^{n(h(X)-\epsilon)} + \sum_{i=0}^{n-|u|} K e^{i(h(X)-\epsilon)} K e^{(n-i)(h(X)-\epsilon)} \\ &= K e^{n(h(X)-\epsilon)} + \sum_{i=0}^{n-|u|} K^2 e^{n(h(X)-\epsilon)} \le (K + nK^2) e^{n(h(X)-\epsilon)}. \end{aligned}$$

Taking logarithms, dividing by n, and letting $n \to \infty$ yields $h(\mathcal{B}) \le h(X) - \epsilon < h(X)$. Since $\mathcal{C} = \mathcal{C}^p \cup \mathcal{C}^s \cup \mathcal{B}$, we have shown [II].

Finally, we prove [III$_\mathrm{a}$] and [III$_\mathrm{b}$] with $L = |v|$. For [III$_\mathrm{a}$], suppose that $x \in X$ and $i \le j \le k \le \ell$ are such that $x([i,k]), x([j,\ell]) \in \mathcal{G}$ and $k - j \ge L$. Then v is a prefix of $x([j,\ell])$; it is also a prefix of $x([j,k])$ since $k - j \ge L = |v|$. Moreover, $x([i,k])u \in \mathcal{L}$ and hence $x([j,k])u \in \mathcal{L}$ as well, which shows that $x([j,k]) \in \mathcal{G}$. For [III$_\mathrm{b}$], we see that v is a prefix of $x([i,\ell])$ since it is a prefix of $x([i,k])$. To show that $x([i,\ell])u \in \mathcal{L}$, we need to use the assumption that there is $a \le i$ with $x([a,\ell]) \in \mathcal{G}$, which implies that $x([a,\ell])u \in \mathcal{L}$ and hence $x([i,\ell])u \in \mathcal{L}$. Thus, $x([i,\ell]) \in \mathcal{G}$, proving [III$_\mathrm{b}$].

We have verified the hypotheses of the first part of Theorem 2.22, which is enough to establish uniqueness and the central limit theorem. The remaining conclusions of Theorem 1.1 will follow from the second half of Theorem 2.22 once we prove that

$$\gcd\{|w| + |u| : w \in \mathcal{G}\} = 1. \tag{3.13}$$

To see this, we produce words w, $w' \in \mathcal{G}$ whose lengths differ by exactly 1. First let $x \in \mathcal{L}$ be a word that starts and ends with v, and contains at least $2m + 1$ disjoint occurrences of the word v; that is, $x = vz^{(1)}vz^{(2)}v \cdots vz^{(2m)}v$. Note that such an x exists by irreducibility. Now let a be a letter such that $y = xa \in \mathcal{L}$. By LAS, there are $x' \in B_m(x)$ and $y' \in B_m(y)$ such that $x'u$, $y'u \in \mathcal{L}$. Thus, at least $m + 1$ of the instances of v in x survive in x', and similarly in y'. In particular, there is some instance of v that survives in both x' and y'; say this occurs as $x_{[i,i+|v|)}$. Then we can take $w = vx'_{[i+|v|,|x|]}u \in \mathcal{L}$ and $w' = vy'_{[i+|v|,|x|+1]}u \in \mathcal{L}$; by definition, both of these words are in $\mathcal{G}$, and their lengths clearly differ by 1. This establishes (3.13) and completes the proof of Theorem 1.1, modulo the derivation in the next section of Theorem 2.22 from the results in [**5**].

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

3.4. *Proof of Theorem 2.22.* Theorem 2.22 is a consequence of results in [**5**]. Since it is not stated in exactly this form there, we explain the necessary steps. First note that [**5**] is given in terms of a potential function φ, while we consider only MMEs, which correspond to the case $\varphi = 0$. Also note that our condition [I] is denoted by [I′] in [**5**, Theorem 3.2].

The main result in that paper is [**5**, Theorem 1.1], which is given in terms of [I], [II], and a condition [III] that we do not state here, but which implies both [III_a] and [III_b]. It is shown there that if [I], [II], and [III] hold, then X has a unique MME μ, and that μ satisfies the stronger statistical properties listed in Theorem 1.1, with two caveats:

- the result in [**5**] allows for a period d in the Bernoulli property and exponential decay of correlations;
- the definition of periodic orbit measures in [**5**] involves a sum over all orbits of length $\leq n$, instead of length exactly n.

Thus, to prove Theorem 2.22, we must explain three things:

(1) why [III_a] and [III_b] are sufficient to replace [III];
(2) why $d = 1$ in our setting;
(3) why our definition of periodic orbit measures still gives the equidistribution result.

This requires a more careful examination of the results in [**5**, §3], which establish [**5**, Theorem 1.1]. In [**5**, Theorem 3.2], it is shown that [I], [II], [III_a], and [III_b] guarantee existence of $\mathcal{F} \subset \mathcal{L}(X)$ such that the following are true.

- $\mathcal{F}$ is closed under concatenation: if $v, w \in \mathcal{F}$, then $vw \in \mathcal{F}$. In particular, $\mathcal{F}$ satisfies [I] with $\tau = 0$.
- $\mathcal{F}$ satisfies [II]; there are $\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}(X)$ such that $h(\mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L}(X) \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s)) < h(X)$.
- $\gcd\{|v| : v \in \mathcal{F}\} = \gcd\{|w| + \tau : w \in \mathcal{G}\}$.

In fact, [**5**, Theorem 3.2] guarantees two further conditions [II′] and [III*], which we do not state here but which are used in [**5**, Theorem 3.1] to give the conclusions of [**5**, Theorem 1.1] with $d = \gcd\{|v| : v \in \mathcal{F}\} = \gcd\{|w| + \tau : w \in \mathcal{G}\}$. This demonstrates that under the hypotheses of Theorem 2.22, we have $d = 1$ and moreover that [III_a] and [III_b] can replace [III] in [**5**, Theorem 1.1], which deals with the first two issues raised above.

It only remains to address the issue concerning periodic orbit measures. By [**5**, Lemma 4.5] applied to $\mathcal{F}$, there are $Q > 0$ and $K, N \in \mathbb{N}$ such that for $n \geq K$ there is $j \in [n - N, n]$ with $|\mathcal{F}_j| \geq Qe^{jh(X)}$. Since $\gcd\{|v| : v \in \mathcal{F}\} = 1$ and since $\mathcal{F}$ is closed under concatenation, there is $M \in \mathbb{N}$ such that $\mathcal{F}_m \neq \emptyset$ for all $m \geq M$. Given $n \geq K + M + N$, there is $j \in [n - M - N, n - M]$ such that

$$|\mathcal{F}_j| \geq Qe^{jh(X)} \geq Qe^{(n-M-N)h(X)}.$$

Now $\mathcal{F}_j \mathcal{F}_{n-j} \subset \mathcal{F}_n$ (by the fact that $\mathcal{F}$ is closed under concatenation) and $\mathcal{F}_{n-j} \neq \emptyset$ (since $n - j \geq M$), so

$$|\text{Per}_n(X)| \geq |\mathcal{F}_n| \geq |\mathcal{F}_j| \geq Qe^{-(M+N)h(X)} e^{nh(X)}, \tag{3.14}$$

where the first inequality uses the fact that every word $w \in \mathcal{F}_n$ generates a periodic point of period n by the free concatenation property. Now let $\mu_n = 1/(|\text{Per}_n(X)|) \sum_{x \in \text{Per}_n(X)} \delta_x$ be the periodic orbit measures defined in (2.6), and let $\tilde{\mu} = \lim_k \mu_{n_k}$ be any weak* limit point of the sequence μ_n. Note that $\sigma_* \mu_n = \mu_n$, so this is a case of the construction from

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Lemma 2.10. By (3.14), clearly $\liminf (1/n_k) \ln |\operatorname{Per}_{n_k}(X)| = h(X)$ and we conclude that $h(\tilde{\mu}) = h(X)$. Since μ is the unique MME, this implies that $\tilde{\mu} = \mu$. This holds for every limit point $\tilde{\mu}$ of μ and therefore $\mu_n \to \mu$.

4. *Other proofs*

4.1. *Proof of Corollary 1.2.* The key observation for the proof of Corollary 1.2 is that usual almost specification with $g \equiv 1$ implies either left or right almost specification.

LEMMA 4.1. *If X is irreducible and has almost specification with $g \equiv 1$, then either X has left almost specification with $g \equiv 1$ or X has right almost specification with $g \equiv 1$.*

Proof. Suppose that X is an irreducible subshift which has neither LAS with $g \equiv 1$ nor RAS with $g \equiv 1$. Then there exist $u, v \in \mathcal{L}(X)$ for which $u'v \notin \mathcal{L}(X)$ for all $u' \in B_1(u)$, and there exist $U, V \in \mathcal{L}(X)$ for which $UV' \notin \mathcal{L}(X)$ for all $V' \in B_1(V)$. Use irreducibility to construct a word of the form $vwU \in \mathcal{L}(X)$.

We now claim that X cannot have almost specification with $g \equiv 1$. Choose any $u' \in B_1(u)$ and $V' \in B_1(V)$. Then $u'v, UV' \notin \mathcal{L}(X)$. Then, if y has the same length as vwU and $u'yV' \in \mathcal{L}(X)$, both the initial v and terminal U from vwU must have been changed in y. This means that $y \notin B_1(vwU)$, and so X does not have almost specification with $g \equiv 1$.

By contrapositive, if X is irreducible and has almost specification with $g \equiv 1$, it must have either LAS or RAS with $g \equiv 1$, completing the proof. □

We also need the following analogue of Theorem 3.1.

THEOREM 4.2. *If X has almost specification with mistake function $g(n)$, then its measure center $\hat{X}$ is irreducible and has almost specification with mistake function $g(n)$.*

Proof. Since X has almost specification, it follows from [**21**, Theorem 6.7] that $\hat{X}$ does as well. Then irreducibility follows from [**8**, Theorems 3.5, 3.7, and 4.3]. Note that the last of these requires an invariant measure whose support is $\hat{X}$; to build such a measure, enumerate the elements of $\mathcal{L}(\hat{X})$ as $v^{(1)}, v^{(2)}, \ldots$, choose invariant measures μ_n with $\mu_n([v^{(n)}]) > 0$, and let $\mu = \sum_{n=1}^{\infty} 2^{-n}\mu_n$.

Because [**21**, Theorem 6.7] does not explicitly state the mistake function for $\hat{X}$, we give a direct proof that $\hat{X}$ has almost specification with mistake function $g(n)$. Given $w^{(1)}, \ldots, w^{(k)} \in \mathcal{L}(\hat{X})$, let $w^{(mk+r)} = w^{(r)}$ for all $m \in \mathbb{N}$ and $r \in \{1, \ldots, k\}$; for each n, we apply the almost specification property of X to $w^{(1)}, \ldots, w^{(nk)}$ to obtain $x^{(n)} \in X$ such that $x^{(n)}_{[1,\infty)} = \bar{w}^{(n,1)} \ldots \bar{w}^{(n,nk)}$, where $\bar{w}^{(n,i)} \in B_{g(|w^{(i)}|)}(w^{(i)})$ for each $1 \le i \le nk$. By compactness, there is $n_j \to \infty$ such that $x = \lim_{j\to\infty} x^{(n_j)} \in X$ exists. Moreover, we have $x_{[1,\infty)} = \mathbf{v}^{(1)}\mathbf{v}^{(2)} \cdots$ for some $\mathbf{v}^{(m)} = v^{(m,1)}v^{(m,2)} \cdots v^{(m,k)}$ with $v^{(m,r)} \in B_{g(|w^{(r)}|)}(w^{(r)})$. Because there are only finitely many possibilities for each $\mathbf{v}^{(m)}$, there is $\bar{\mathbf{v}} = \bar{v}^{(1)} \cdots \bar{v}^{(k)}$ such that $M = \{m \in \mathbb{N} : \mathbf{v}^{(m)} = \bar{\mathbf{v}}\}$ has positive upper density. By the same argument as in the proof of Theorem 3.1, there is an ergodic measure μ with $\mu([\bar{\mathbf{v}}]) > 0$, and we conclude that $\bar{\mathbf{v}} \in \mathcal{L}(\hat{X})$, which completes the proof of Theorem 4.2. □

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Now we complete the proof of Corollary 1.2. By Theorem 4.2, we may assume without loss of generality that X is irreducible. Then, by Theorem 4.1, X either has LAS with $g \equiv 1$ or RAS with $g \equiv 1$. In the former case, the proof is complete by applying Theorem 1.1.

If X has RAS with $g \equiv 1$, then define a map $\rho\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ by

$$\rho(\ldots x(-1)x(0)x(1)\ldots) = \ldots x(1)x(0)x(-1)\ldots.$$

Then $\rho(X)$ is a subshift and, since X was irreducible and had RAS with $g \equiv 1$, it is clear that $\rho(X)$ is irreducible and has LAS with $g \equiv 1$. Therefore, $\rho(X)$ has a unique measure of maximal entropy satisfying the conclusions of Theorem 1.1. Since ρ is bijective on periodic points of X and induces an entropy-preserving bijection between $\mathcal{M}(X)$ and $\mathcal{M}(\rho(X))$, there is a corresponding unique measure of maximal entropy on X with the same properties.

4.2. *Proof of Theorem 1.4.* Having shown that LAS with bounded g implies intrinsic ergodicity, it is natural to wonder whether this result can be strengthened, i.e. whether other upper bounds on g may also imply uniqueness. We can now prove Theorem 1.4, which shows that multiple MMEs can occur within subshifts with LAS for g on the order of $\log\log n$. We will use the following technical lemma from [**11**].

LEMMA 4.3. [**11**, Lemma 4.4] *For every alphabet A and positive integer n, there exists a set $U_n \subseteq A^n$ such that $|U_n| \le (16/n^2)|A|^n$ and U_n is 2-spanning with respect to the Hamming metric, i.e. for all $w \in A^n$, there exists $u \in U_n$ with $d_H(w, u) \le 2$.*

We will actually exhibit a subshift with RAS rather than LAS; as argued in the proof of Corollary 1.2, this obviously can be turned into an LAS example by simply reflecting sequences about the origin. Our subshift will have alphabet

$$A = \{-N, \ldots, -1, 1, \ldots, N\},$$

where $N > 2^{17} + 4$. We begin by using Lemma 4.3 to fix, for every n, a set $U_n \subseteq \{1, \ldots, N\}^n$ which is 2-spanning in $\{1, \ldots, N\}^n$ and with $|U_n| \le (16/n^2)N^n$. We will use U_n to define certain collections $T^+ \subset \{1, \ldots, N\}^*$ and $T^- \subset \{-1, \ldots, -N\}^*$, and then take X to be the coded shift generated by $T := T^+ \cup T^-$, i.e. the closure of the set of all bi-infinite concatenations of words from T.

For every $n \ge 4$, define $k = k(n) = \lfloor \log_2 \log_2 n \rfloor$ (note that $2^{2^k} \le n < 2^{2^{k+1}}$) and define T_n^+ to be the set

$$\{w \in \{1, \ldots, N\}^n : w(1) = 1, \forall i \in [0, k-1],\ w([2^{2^i}+1, 2^{2^{i+1}}]) \in U_{2^{2^{i+1}}-2^{2^i}}\}.$$

Define $T^+ = \bigcup_n T_n^+$. We first note that the collection T^+ is prefix-closed, meaning that $w \in T^+,\ w = uv \Longrightarrow u \in T^+$; this follows immediately from the definition of the sets T_n^+. Also, for every n, T_n^+ is $(1 + 2\lfloor \log_2 \log_2 n \rfloor)$-spanning in $\{1, \ldots, N\}^n$; given a word $w \in \{1, \ldots, N\}^n$, one only needs to change the first letter to 1 and make at most two changes to each $w([2^{2^i}+1, 2^{2^{i+1}}])$, $i \in [0, k-1]$, to place it in $U_{2^{2^{i+1}}-2^{2^i}}$.

Finally, for every n, $|T_n^+|$ is bounded above by

$$N^{n-1} \cdot \prod_{i=0}^{k-1} \frac{16}{(2^{2^{i+1}} - 2^{2^i})^2} \le N^{n-1} \cdot \prod_{i=0}^{k-1} \frac{16}{(2^{2^{i+1}-1})^2} = N^n \left(\frac{1}{N \cdot 2^{2^{k+2}-6k-4}} \right).$$

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Then, for all n,

$$|T_n^+| \le \frac{N^n}{N}. \tag{4.1}$$

When $n \ge 2^{16}$, then $k \ge 4$, so $2^{k+2} - 6k - 4 > 1.125 \cdot 2^{k+1}$. This means that for $n \ge 2^{16}$, we get the stronger bound

$$|T_n^+| \le \frac{N^n}{N \cdot n^{1.125}}. \tag{4.2}$$

Now, for every n, define $T_n^- \subseteq \{-N, \dots, -1\}^n$ by $T_n^- = -T_n^+ = \{-w : w \in T_n^+\}$, and define $T^- = \bigcup_n T_n^-$. Clearly, T^- is prefix-closed since T^+ is. Similarly, T_n^- is $(1 + 2\lfloor \log_2 \log_2 n \rfloor)$-spanning in $\{-N, \dots, -1\}^n$ and has the same cardinality bound as that of (4.2). Define, for every n, $T_n = T_n^- \cup T_n^+$; clearly, T_n is $(1 + 2\lfloor \log_2 \log_2 n \rfloor)$-spanning within the set of n-letter words with constant sign. Then define $T = T^+ \cup T^- = \bigcup_n T_n$. We are finally prepared to define our subshift X; it is simply the coded system defined by T.

We first note that $\{1, \dots, N\}^{\mathbb{Z}} \subset X$. To see this, note that for every k, the final $2^{2^{k+1}} - 2^{2^k} - 1 > k$ letters of words in $T^+_{2^{2^{k+1}}-1}$ are completely unconstrained. In other words, every word in $\{1, \dots, N\}^*$ is the suffix of some word in T^+ and so, by taking limits as $k \to \infty$, we see that $\{1, \dots, N\}^{\mathbb{Z}} \subset X$. Similarly, every word in $\{-N, \dots, -1\}^*$ is the suffix of some word in T^- and so $\{-N, \dots, -1\}^{\mathbb{Z}} \subset X$. Trivially, this implies that $h(X) \ge \log N$. We will now bound $|\mathcal{L}_n(X)|$ from above similarly to [**11**] to show that in fact $h(X) = \log N$.

For every n and every $w \in \mathcal{L}_n(X)$, we can decompose w as $w = w^{(1)} \dots w^{(k)}$, where $w^{(1)}$ is the suffix of a word in T, $w^{(2)}, \dots, w^{(k-1)} \in T$, and $w^{(k)}$ is the prefix of a word in T. Recall that T is prefix-closed and that every word of constant sign is the suffix of some word in T. Therefore, we can rephrase by saying that $w^{(1)}$ has constant sign and that $w^{(2)}, \dots, w^{(k)} \in T$. Defining $n_i = |w^{(i)}|$ for $1 \le i \le k$, we see that

$$\begin{aligned} |\mathcal{L}_n(X)| &\le \sum_{k=1}^{n} \sum_{\sum_{i=1}^k n_i = n} 2N^{n_1} \prod_{i=2}^{k} |T_{n_i}| \\ &\le 2N^n \sum_{k=1}^{n} \sum_{\sum_{i=1}^k n_i = n} \prod_{i=2}^{k} \frac{|T_{n_i}|}{N^{n_i}} \le 2N^n \sum_{k=1}^{n} \left(\sum_{j=1}^{\infty} \frac{|T_j|}{N^j} \right)^{k-1}. \end{aligned} \tag{4.3}$$

By (4.1) and (4.2), we have

$$\sum_{j=1}^{\infty} \frac{|T_j|}{N^j} \le \sum_{j=1}^{2^{16}} \frac{2}{N} + \sum_{2^{16}+1}^{\infty} \frac{2}{N \cdot j^{1.125}} < \frac{2^{17}}{N} + \int_{2^{16}}^{\infty} \frac{2}{N \cdot x^{1.125}}\, dx = \frac{2^{17}+4}{N}.$$

Since $N > 2^{17} + 4$, we can define $\alpha := ((2^{17} + 4)/N) < 1$, and then (4.3) shows that $|\mathcal{L}_n(X)| < 2N^n(1/(1-\alpha))$. Taking logarithms, dividing by n, and letting $n \to \infty$ shows that $h(X) \le \log N$ and so $h(X) = \log N$. This immediately implies that X has at least two measures of maximal entropy (with disjoint supports), namely the uniform Bernoulli measures on the full shifts $\{1, \dots, N\}^{\mathbb{Z}}$ and $\{-N, \dots, -1\}^{\mathbb{Z}}$, both of which are contained in X.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

It only remains to show that X has RAS with $g(n) = 1 + \lfloor 2 \log_2 \log_2 n \rfloor$. To see this, consider any $v, w \in \mathcal{L}(X)$. Decompose each into maximal words of constant sign as $v = v^{(1)}v^{(2)} \dots v^{(m)}$ and $w = w^{(1)}w^{(2)} \dots w^{(n)}$, i.e. every $v^{(j)}$ and $w^{(j)}$ has constant sign, which alternates as j varies. Then $v^{(1)}$ is a suffix of a word t in T, and all $v^{(j)}$ for $j > 1$ are in T (recall that T is prefix-closed). Similarly, $w^{(j)}$ is in T for $j > 1$. Let $\ell = |w^{(1)}|$. Since T_ℓ is $(1 + 2\lfloor \log_2 \log_2 \ell \rfloor)$-spanning in $\{-N, \dots, -1\}^\ell \cup \{1, \dots, N\}^\ell$ and $w^{(1)}$ has constant sign, there exists $u^{(1)} \in B_{1+2\lfloor \log_2 \log_2 \ell \rfloor}(w^{(1)})$ which is also in T. Since it is a concatenation of words in T, $tv^{(2)} \dots v^{(m)}u^{(1)}w^{(2)} \dots w^{(n)} \in \mathcal{L}(X)$. Then its subword $v^{(1)} \dots v^{(m)}u^{(1)}w^{(2)} \dots w^{(n)}$ must also be in $\mathcal{L}(X)$. Finally, since $d_H(u^{(1)}w^{(2)} \dots w^{(m)}, w) = d_H(u^{(1)}, w^{(1)}) \le 1 + 2\lfloor \log_2 \log_2 \ell \rfloor \le 1 + 2 \log_2 \log_2 |w|$, we have shown that X has RAS with $g(n) = 1 + 2\lfloor \log_2 \log_2 n \rfloor$.

4.3. *Proof of Proposition 1.7.* Now we prove that the property of having LAS with bounded g is preserved under products and factors.

Suppose that X, Y, with alphabets A, B, respectively, have LAS with $g \equiv m$ and $g \equiv n$, respectively. Then $X \times Y$ is a subshift with alphabet $A \times B$. Consider any words $w^{(1)}, w^{(2)} \in \mathcal{L}(X \times Y)$. Let $u^{(1)} \in A^*$ and $v^{(1)} \in B^*$ denote the first and second coordinates of $w^{(1)}$, respectively; by definition, $u^{(1)} \in \mathcal{L}(X)$ and $v^{(1)} \in \mathcal{L}(Y)$. Similarly, define $u^{(2)} \in \mathcal{L}(X)$ and $v^{(2)} \in \mathcal{L}(Y)$ via $w^{(2)}$. Then, by LAS of X, there exist $\bar{u}^{(1)} \in B_m(u^{(1)})$, $\bar{v}^{(1)} \in B_n(v^{(1)})$ so that $\bar{u}^{(1)}u^{(2)} \in \mathcal{L}(X)$ and $\bar{v}^{(1)}v^{(2)} \in \mathcal{L}(Y)$. Then, if we define $\bar{w}^{(1)}$ to have first and second coordinates $\bar{u}^{(1)}$ and $\bar{v}^{(1)}$, we get $\bar{w}^{(1)}w^{(2)} \in \mathcal{L}(X \times Y)$ and $\bar{w}^{(1)} \in B_{m+n}(w^{(1)})$. Therefore, $X \times Y$ has LAS with $g \equiv m + n$.

Now consider a factor map ϕ on X. Then ϕ is a sliding block code with some radius r, i.e. $x([i - r, i + r])$ uniquely determines $(\phi(x))(i)$ for every $x \in X$, $i \in \mathbb{Z}$. Take any words $w^{(1)} \in \mathcal{L}_{n_1}(\phi(X))$ and $w^{(2)} \in \mathcal{L}_{n_2}(\phi(X))$. There then exist words $v^{(1)} \in \mathcal{L}_{n_1+2r}(X)$, $v^{(2)} \in \mathcal{L}_{n_2+2r}(X)$ so that $\phi(v^{(1)}) = w^{(1)}$ and $\phi(v^{(2)}) = w^{(2)}$. Create $u^{(1)} \in \mathcal{L}_{n_1}(X)$ by removing the final $2r$ letters of $v^{(1)}$. Then, by definition of sliding block code, $p^{(1)} := \phi(u^{(1)})$ is a prefix of $w^{(1)}$ of length $n_1 - 2r$.

Then, by LAS of X, there exists $\bar{u}^{(1)} \in B_m(u^{(1)})$ for which $\bar{u}^{(1)}v^{(2)} \in \mathcal{L}_{n_1+n_2+2r}(X)$. Define $y = \phi(\bar{u}^{(1)}v^{(2)})$; then $y \in \mathcal{L}_{n_1+n_2}(\phi(X))$. Since $v^{(2)}$ is a suffix of $\bar{u}^{(1)}v^{(2)}$, and $\phi(v^{(2)}) = w^{(2)}$, $y = \phi(\bar{u}^{(1)}v^{(2)})$ has $w^{(2)}$ as a suffix. Also, since $\bar{u}^{(1)} \in B_m(u^{(1)})$, and ϕ has radius r, $\bar{p}^{(1)} := \phi(\bar{u}^{(1)}) \in B_{m(2r+1)}(\phi(u^{(1)})) = B_{m(2r+1)}(p^{(1)})$. (The only differences in $\phi(\bar{u}^{(1)})$ and $\phi(u^{(1)})$ must be at locations within distance r from a difference within $\bar{u}^{(1)}$ and $u^{(1)}$.) Since $\bar{u}^{(1)}$ is a prefix of $\bar{u}^{(1)}v^{(2)}$, $y = \phi(\bar{u}^{(1)}v^{(2)})$ has $\bar{p}^{(1)}$ as a prefix. But then we can write $y = \bar{p}^{(1)}xw^{(2)}$ for some $x \in \mathcal{L}_{2r}(\phi(X))$. Since $\bar{p}^{(1)} \in B_{m(2r+1)}(p^{(1)})$, $\bar{p}^{(1)}x \in B_{2r+m(2r+1)}(w^{(1)})$. Since $y \in \mathcal{L}(\phi(X))$, we have proved that $\phi(X)$ has LAS with $g \equiv 2r + m(2r + 1)$.

4.4. *Proof of Lemma 2.7.* We will show that for a factorial set $\mathcal{D}$ of words, the entropy of $\mathcal{D}$ is the same as that of the subshift 'generated by' $\mathcal{D}$.

Given a word $w \in A^*$, a sufficient condition to have $w \in \mathcal{L}(X(\mathcal{D}))$ is that there are infinitely many $k \in \mathbb{N}$ and $u, v \in A^k$ such that $uwv \in \mathcal{D}$. Given $i, j \in \mathbb{N}$, consider the set

$$\mathcal{D}_i^{(j)} = \{w \in \mathcal{D}_i : \text{there are } u, v \in \mathcal{D}_j \text{ such that } uwv \in \mathcal{D}\}.$$

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

Note that by definition, for every $n \in \mathbb{N}$,

$$\mathcal{D}_n = \mathcal{D}_n^{(0)} \supset \mathcal{D}_n^{(n)} \supset \mathcal{D}_n^{(2n)} \supset \cdots .$$

Since these sets are finite, this sequence stabilizes, i.e. there exists $K \in \mathbb{N}$ so that for all $k \geq K$,

$$\mathcal{D}_n^{(kn)} = \bigcap_{j \in \mathbb{N}} \mathcal{D}_n^{(jn)} \subseteq \mathcal{L}_n(X(\mathcal{D})). \tag{4.4}$$

Our goal is to prove that for any $n \in \mathbb{N}$ and $\epsilon > 0$, we have

$$|\mathcal{D}_n^{(kn)}| \geq e^{n(h(\mathcal{D})-2\epsilon)} \tag{4.5}$$

for all sufficiently large k; using this, we will produce 'enough' words in $\mathcal{L}_n(X(\mathcal{D}))$.

Since $\mathcal{D}$ is factorial, Lemma 2.6 gives $|\mathcal{D}_n| \geq e^{nh(\mathcal{D})}$ for all n. Fix $\epsilon > 0$ and let $N = N(\epsilon)$ be such that

$$e^{nh(\mathcal{D})} \leq |\mathcal{D}_n| \leq e^{n(h(\mathcal{D})+\epsilon)} \tag{4.6}$$

for all $n \geq N$. For any $k, n \in \mathbb{N}$, we observe that if $w \in \mathcal{D}_{kn}^{(kn)}$, then

$$w([1, n]), w([n + 1, 2n]), \ldots, w([(k - 1)n + 1, kn]) \in \mathcal{D}_n^{(kn)}$$

by factoriality of $\mathcal{D}$. In particular, this gives

$$|\mathcal{D}_{kn}^{(kn)}| \leq |\mathcal{D}_n^{(kn)}|^k. \tag{4.7}$$

On the other hand, every $w \in \mathcal{D}_{3kn}$ has $w([kn + 1, 2kn]) \in \mathcal{D}_{kn}^{(kn)}$ and so

$$e^{3knh(\mathcal{D})} \leq |\mathcal{D}_{3kn}| \leq |\mathcal{D}_{kn}|^2 |\mathcal{D}_{kn}^{(kn)}| \leq e^{2kn(h(\mathcal{D})+\epsilon)} |\mathcal{D}_n^{(kn)}|^k$$

for every k, n with $kn > N$, where the last inequality uses (4.6) and (4.7). Dividing by $e^{2kn(h(\mathcal{D})+\epsilon)}$ and taking a kth root gives (4.5) for all $k > N/n$.

Now, by (4.4) and (4.5),

$$|\mathcal{L}_n(X(\mathcal{D}))| \geq \left| \bigcap_{j \in \mathbb{N}} \mathcal{D}_n^{(jn)} \right| \geq e^{n(h(\mathcal{D})-2\epsilon)}.$$

Dividing by n, taking logarithms, and letting $n \to \infty$ yields $h(X(\mathcal{D})) \geq h(\mathcal{D}) - 2\epsilon$. Since $\epsilon > 0$ was arbitrary, we have $h(X(\mathcal{D})) \geq h(\mathcal{D})$, which completes the proof of Lemma 2.7.

4.5. *Unbounded mistake functions.* As mentioned in Remark 3.9, Theorem 3.8 fails as soon as g is unbounded. To demonstrate this, we give a family of irreducible examples with LAS for arbitrary unbounded g which are not entropy minimal, and in fact even have zero entropy.

Consider any unbounded $g\colon \mathbb{N} \to \mathbb{N}$, where $g(n)/n \to 0$ as $n \to \infty$. We may assume without loss of generality that g is non-decreasing since any subshift with LAS with mistake function $g(n)$ also has LAS for any larger mistake function. Then define X to have alphabet $\{0, 1\}$ and consist of all $x \in \{0, 1\}^{\mathbb{Z}}$ such that for every n and every n-letter subword w of x, the number of 1 symbols in w is less than or equal to $g(n)$. (These shifts were defined in [**17**] under the name *bounded density shifts.*) We note that $\mathcal{L}(X)$ is exactly

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

the set of words v where for every n and every subword w of length n, the number of 1s in w is less than or equal to $g(n)$; any such word v is in $\mathcal{L}(X)$ since $0^\infty v 0^\infty \in X$.

We first show that X has LAS with mistake function $g(n)$. For any $v, w \in \mathcal{L}(X)$, the number of 1s in v is at most $g(|v|)$ and so $v' = 0^{|v|} \in B_{g(|v|)}(v)$. Then we claim that $v'w \in \mathcal{L}(X)$. To see this, note that for every subword u of $v'w$, the number of 1s in u is equal to the number of 1s in u', the maximal subword of u contained in w. Then, since $w \in \mathcal{L}(X)$, this is less than or equal to $g(|u'|) \leq g(|u|)$ and so $v'w \in \mathcal{L}(X)$, completing the proof of LAS.

Now we show that X is irreducible. Consider any $v, w \in \mathcal{L}(X)$ and let $n \geq \max\{|v|, |w|\}$. Without loss of generality, assume that they have the same length n. Since g is unbounded, there exists N so that $g(N) \geq 2n$. Then let $u = v0^N w$. Any subword u' of u which does not contain letters of both v and w is a subword of either $v0^N$ or $0^N w$, and so has a number of 1s less than or equal to $g(|u'|)$ exactly as argued above. Any other subword u'' of u contains letters of both v and w and so has length greater than N. The number of 1s in u'' is then less than or equal to $2n \leq g(N) \leq g(|u''|)$. We have then shown that $u \in \mathcal{L}(X)$, proving irreducibility of X.

We also claim that $h(X) = 0$. To see this, note that since $g(n)/n \to 0$, the limiting frequency of 1 symbols in every point of x is zero. Therefore, an invariant measure μ on X must have $\mu([1]) = 0$ and so the only such measure is the delta measure on the sequence of all 0s. This measure clearly has entropy 0 and so, by the variational principle, $h(X) = 0$. Then X is not entropy minimal, since it contains the proper subshift containing only the point of all 0s, which also has entropy 0.

Acknowledgements. We are grateful to the anonymous referee for a careful reading and many helpful comments, including suggestions that led to the current proofs (and formulations) of Theorems 3.1 and 4.2. V.C. is partially supported by NSF grant DMS-1362838. R.P. is partially supported by NSF grant DMS-1500685.

References

[**1**] A. Bertrand. Specification, synchronisation, average length. *Coding Theory and Applications (Cachan, 1986) (Lecture Notes in Computer Science, 311)*. Springer, Berlin, 1988, pp. 86–95.
[**2**] R. Bowen. Periodic points and measures for Axiom *A* diffeomorphisms. *Trans. Amer. Math. Soc.* **154** (1971), 377–397.
[**3**] R. Bowen. Some systems with unique equilibrium states. *Math. Systems Theory* **8**(3) (1974–1975), 193–202.
[**4**] J. Buzzi. Specification on the interval. *Trans. Amer. Math. Soc.* **349**(7) (1997), 2737–2754.
[**5**] V. Climenhaga. Specification and towers in shift spaces. *Preprint*, 2015, arXiv:1502.00931v2, submitted.
[**6**] V. Climenhaga and D. J. Thompson. Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors. *Israel J. Math.* **192**(2) (2012), 785–817.
[**7**] E. M. Coven and J. Smítal. Entropy-minimality. *Acta Math. Univ. Comenian. (N.S.)* **62**(1) (1993), 117–121.
[**8**] M. Kulczycki, D. Kwietniak and P. Oprocha. On almost specification and average shadowing properties. *Fund. Math.* **224**(3) (2014), 241–278.
[**9**] D. Kwietniak, P. Oprocha and M. Rams. On entropy of dynamical systems with almost specification. *Israel J. Math.* **213**(1) (2016), 475–503.
[**10**] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–66.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press

[**11**] R. Pavlov. On intrinsic ergodicity and weakenings of the specification property. *Adv. Math.* **295** (2016), 250–270.
[**12**] R. Pavlov. On non-uniform specification and uniqueness of the equilibrium state in expansive systems. *Preprint*, 2017, arXiv:1702.02870v3, submitted.
[**13**] C.-E. Pfister and W. G. Sullivan. Large deviations estimates for dynamical systems without the specification property. Applications to the β-shifts. *Nonlinearity* **18**(1) (2005), 237–261.
[**14**] C.-E. Pfister and W. G. Sullivan. On the topological entropy of saturated sets. *Ergod. Th. & Dynam. Sys.* **27**(3) (2007), 929–956.
[**15**] A. N. Quas and P. B. Trow. Subshifts of multi-dimensional shifts of finite type. *Ergod. Th. & Dynam. Sys.* **20**(3) (2000), 859–874.
[**16**] J. Schmeling. Symbolic dynamics for β-shifts and self-normal numbers. *Ergod. Th. & Dynam. Sys.* **17**(3) (1997), 675–694.
[**17**] B. Stanley. Bounded density shifts. *Ergod. Th. & Dynam. Sys.* **33**(6) (2013), 1891–1928.
[**18**] D. J. Thompson. Irregular sets, the β-transformation and the almost specification property. *Trans. Amer. Math. Soc.* **364**(10) (2012), 5395–5414.
[**19**] P. Walters. *An Introduction to Ergodic Theory (Graduate Texts in Mathematics, 79)*. Springer, New York, 1982.
[**20**] B. Weiss. Intrinsically ergodic systems. *Bull. Amer. Math. Soc.* **76** (1970), 1266–1269.
[**21**] X. Wu, P. Oprocha and G. Chen. On various definitions of shadowing with average error in tracing. *Nonlinearity* **29**(7) (2016), 1942–1972.

https://doi.org/10.1017/etds.2017.135 Published online by Cambridge University Press