ONE-SIDED ALMOST SPECIFICATION AND INTRINSIC ERGODICITY

VAUGHN CLIMENHAGA AND RONNIE PAVLOV

Abstract. Shift spaces with the specification property are intrinsically ergodic, i.e. they have a unique measure of maximal entropy. This can fail for shifts with the weaker *almost specification* property. We define a new property called *one-sided* almost specification, which lies in between specification and almost specification, and prove that it guarantees intrinsic ergodicity 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 $g = 1$ implies one-sided almost specification (with $g = 1$), and hence uniqueness. On the other hand, the second author showed recently that almost specification with $g = 4$ does not imply uniqueness. This leaves open the question of whether almost specification implies intrinsic ergodicity when $g = 2$ or $g = 3$.

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** [15].

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.

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

Date: May 17, 2016.

2010 *Mathematics Subject Classification.* Primary: 37B10; Secondary: 37B40, 37D35.

Key words and phrases. Measure of maximal entropy, specification, subshift.

V.C. is partially supported by NSF grant DMS-1362838. R.P. is partially supported by NSF grant DMS-1500685.

question is uniqueness. Intrinsic ergodicity for mixing subshifts of finite type was proved by Parry [8] using Perron–Frobenius theory; a different proof for shift spaces 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 are represented by **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** [7, 9, 10, 11, 13], 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 Section 2 for a formal definition. Almost specification is enough to establish various results in large deviation theory [11] and multifractal analysis [10, 13], but it was unknown for some time whether almost specification also implies intrinsic ergodicity. This question was answered in the negative by (independently) [7] and [9]; in fact, [9, Theorem 1.2] shows that intrinsic ergodicity may fail even with the constant mistake function $g(n) = 4$.

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 $g = 1$ and do satisfy intrinsic ergodicity. In fact, β-shifts satisfy a slightly stronger property; when one wishes to concatenate some words $w^{(1)}$, …, $w^{(n)}$ together, it suffices to make the permitted number of changes to the words $w^{(1)}, \dots, w^{(n-1)}$, 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 shift space 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 $g = 1$, and deduce the following.

Corollary 1.2. *If X is a shift space 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 [9] that $g \equiv 4$ is not enough to guarantee uniqueness, we are led to the following open problem.

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 exists a subshift X with LAS with $g(n) = 1+2\log_2\log_2 n$ 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 [9] to create examples satisfying LAS with $g = C \log\log N$ and multiple MMEs for arbitrarily small C. So, the mistake functions for which we do not know whether LAS implies uniqueness of MME are those functions which are unbounded, but have growth rate $o(\log_2\log_2 n)$.

The example with multiple MMEs from Theorem 1.4 has exactly two ergodic MMEs, whose supports are disjoint. It is actually possible for a shift space 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 an example of 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 [4] that uses a certain non-uniform specification condition (superficially quite different from almost specification) to model X with a countable state Markov shift. A key intermediate step is the proof of a property called *entropy minimality* for the subshift X (see Definition 2.11). The following result shows that there is no unbounded g for which irreducibility and LAS with mistake function g implies entropy minimality. This means that if any such g implies intrinsic ergodicity, it could not be proved via our methods.

Theorem 1.7. *For every unbounded $g(n)$, there is an irreducible subshift X satisfying LAS with mistake function $g(n)$ and which is not entropy minimal.*

The conditions in [4] are mild variants of conditions introduced in [5], where they were used to prove that not only are β-shifts intrinsically ergodic, but so are their subshift factors. Although the conditions from [5] are well-behaved under passing to factors, their behaviour upon passing to products is not immediately clear. On the other hand, as we will prove in the following result, the hypotheses of Theorem 1.1 are stable under both products and factors.

Proposition 1.8. *If X, Y are shift spaces satisfying LAS with bounded g, then so is their product $X \times Y$. Similarly, if X is a shift space satisfying LAS with bounded g, and Z is a symbolic factor of X, then Z satisfies LAS with bounded g.*

Remark 1.9. Since β-shifts satisfy LAS with $g = 1$, Proposition 1.8 implies that any shift created by taking iterated products/factors of β-shifts also has 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 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 unique MME. It is quite possible that the methods of [5] or [4] 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 $\rho(x, y) = 2^{-\min\{|k| : x(k) \neq y(,)\}}$.

A **word** over A is an element of $A^* := \bigcup_{n=0}^{\infty} A^n$. 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 **shift space over** A (or **subshift**) 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

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

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 so that $[u] \cap \sigma^{-n}[v] \neq \varnothing$, or equivalently if there exists a word w so that $uwv \in \mathcal{L}(X)$.

2.2. **Thermodynamic formalism for shift spaces.** We recall some of the main elements of thermodynamic formalism as they appear in the context of shift spaces; further details and results can be found in [14].

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 $\inf_n \frac{1}{n} \ln |\mathcal{D}_n|$. 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 §4.5 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** [14, Theorem 8.6] relates the two kinds of entropy: for every subshift X, we have $h(X) = \sup\{h(\mu) : \mu \in \mathcal{M}(X)\}$.

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 [14, Theorem 8.6].

Lemma 2.10. *Let X be a shift space 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 $\nu_k = \frac{1}{|\mathcal{D}_{n_k}|}\sum_{v\in\mathcal{D}_{n_k}} \delta_{x_v}$ and $\mu_k = \frac{1}{n_k}\sum_{i=0}^{n_k-1} \sigma_*^i \nu_k$. Then any weak* limit point μ of the sequence μ_k is a σ-invariant measure with $h(\mu) \geq \liminf \frac{1}{n_k} \ln |\mathcal{D}_{n_k}|$. In particular, if $\liminf \frac{1}{n_k} \ln |D_{n_k}| = h(X)$, then any weak* limit point of the sequence μ_k is an MME for X.*

Our main goal is to prove that certain shift spaces are intrinsically ergodic. A key intermediate step in our approach will be to establish the property of **entropy minimality**, first defined in [6].

Definition 2.11. A subshift X is **entropy minimal** if every nonempty 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 iff 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 iff 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 has a fully supported MME that 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.[1] More recently, various weakenings of the specification property have been introduced. The version we will focus on was introduced by Pfister and Sullivan [11] as the "g-almost product property"; we will follow the convention of [13] 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** $d(v, w) := |\{i \ : \ v(i) \neq w(i)\}|$. We will frequently use the associated **Hamming balls**

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

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

- $\frac{g(n)}{n} \to 0$
- For any words $w^{(1)}$, $w^{(2)}$, …, $w^{(k)} \in \mathcal{L}(X)$, there exist words $w'^{(1)}$, $w'^{(2)}$, …, $w'^{(k)} \in \mathcal{L}(X)$ so that $d(w'^{(i)}, w^{(i)}) \leq g(|w^{(i)}|)$ for $1 \leq i \leq k$, and the concatenation $w'^{(1)}w'^{(2)} \dots w'^{(k)}$ 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

- $\frac{g(n)}{n} \to 0$
- For any words $w^{(1)}, w^{(2)} \in \mathcal{L}(X)$, there exists a word $w'^{(1)}$ in $\mathcal{L}(X)$ where $d(w'^{(1)}, w^{(1)}) \leq g(|w^{(1)}|)$ and $w'^{(1)}w^{(2)} \in \mathcal{L}(X)$.

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.*

[1] 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].

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 $g(n) = \tau$. To see this, choose any $w^{(1)}, w^{(2)} \in \mathcal{L}(X)$. If $|w^{(1)}| \leq \tau$, then choose any word $w'^{(1)}$ with length $|w^{(1)}|$ which can precede $w^{(2)}$ in a point of X; then trivially $d(w^{(1)}, w'^{(1)}) \leq \tau$ and $w'^{(1)}w^{(2)} \in \mathcal{L}(X)$. 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 $d(vu, w^{(1)}) \leq \tau$. In both cases, we found $w'^{(1)}$ with $d(w^{(1)}, w'^{(1)}) \leq \tau$ and $w'^{(1)}w^{(2)} \in \mathcal{L}(X)$, so X has LAS with $g = \tau$. □

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 nondecreasing. There are many subshifts known to satisfy AS and/or LAS; for instance, any β-shift has LAS with gap function $g(n) = 1$ (see [10]); since β-shifts only have specification for a Lebesgue-null set of values of β, 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, 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 $g(n)$.*

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 $w^{(1)}$, $w^{(2)}$, …, $w^{(k)} \in \mathcal{L}(X)$, there exist words $w'^{(1)}$, $w'^{(2)}$, …, $w'^{(k-1)} \in \mathcal{L}(X)$ so that $d(w'^{(i)}, w^{(i)}) \leq g(|w^{(i)}|)$ for $1 \leq i \leq k-1$, and $w'^{(1)}w'^{(2)} \dots w'^{(k-1)}w^{(k)}$ is in $\mathcal{L}(X)$. (Here, note that $w'^{(k)}$ was not defined, i.e. $w^{(k)}$ 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 $w^{(1)}$, $w^{(2)}$, …, $w^{(k+1)} \in \mathcal{L}(X)$, first use LAS to define $w'^{(k)}$ with $d(w'^{(k)}, w^{(k)}) \leq g(|w^{(k)}|)$ so that $w'^{(k)}w^{(k+1)} \in \mathcal{L}(X)$, and then apply the inductive hypothesis to the k words $w^{(1)}, w^{(2)}, \dots, w^{(k-1)}, w'^{(k)}w^{(k+1)}$ to yield the desired $w'^{(i)}$ for $1 \leq i \leq k-1$. □

Almost specification has been used in the literature to study statistical behavior such as large deviations [10] and multifractal properties [11, 13]. In [5], the first author and D.J. 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 [9] 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 [4, 5]. 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")

[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 [5] (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 [5]. Rather, we use the following conditions on $\mathcal{G}$ that were introduced in [4], 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$_\mathrm{a}$] 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$_\mathrm{b}$] 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 [4] that conditions **[I]**–**[III$_\mathrm{b}$]** guarantee intrinsic ergodicity, as well as strong statistical properties for the unique MME, which we describe next. We observe that although [4, Theorem 1.1] is stated using a stronger version of **[III$_\mathrm{a}$]**–**[III$_\mathrm{b}$]** (in which no assumption on $x([a,\ell])$ is given in **[III$_\mathrm{b}$]**), the condition that we use here is sufficient; the result we need is stated as Theorem 2.22 below, and proved in §3.4 using [4, Theorems 3.1(B) and 3.2].

2.4. **Statistical properties.** The unique MMEs that [4, 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 $\mu[w] \leq Q_1 e^{-|w|h(X)}$ 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 every pair of Hölder continuous functions $\psi_1, \psi_2 \colon X \to \mathbb{R}$ has $K(\psi_1, \psi_2) > 0$ such that

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

Definition 2.21. 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 $\frac{1}{\sqrt{n}} S_n \psi = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi \circ \sigma^k$ converges in distribution to a normal distribution $\mathcal{N}(0, \sigma_\psi)$ for some $\sigma_\psi \geq 0$; that is, if

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

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-a.e. $x \in X$. 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 [4]; see §3.4 for details.

Theorem 2.22. *Let X be a shift space on a finite alphabet and suppose there are $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)$ satisfying* **[I]**, **[II]**, **[III$_\mathrm{a}$]**, *and* **[III$_\mathrm{b}$]**. *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 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 X_O 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 3.5). 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$_\text{b}$]** and hence allow us to apply the results on [4]. 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, consider $X = \{x \in \{0,1\} \ : \ |\{n \ : \ x(n) = 1\}| \leq 1\}$, the set of 0-1 sequences with at most one 1. This system is clearly not irreducible, but satisfies LAS with $g = 1$; 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 $\mathcal{L}(X)$.

The following theorem shows that we can remove any such words from an LAS system and gain irreducibility "without loss of generality."

Theorem 3.1. *Suppose X has left almost specification with $g = m$. Let O be the set of words $w \in \mathcal{L}(X)$ for which no point of X contains $m+1$ disjoint occurrences of w. Define X_O to be the subshift of points of X containing no word from O. Then*

(i) *X and X_O have the same sets of periodic points;*
(ii) *X and X_O have the same simplices of invariant measures;*
(iii) *X_O is irreducible;*
(iv) *X_O has left almost specification with $g = m$.*

Proof. Any point $x \in X \setminus X_O$ contains a word from O, which only appears finitely many times in x by definition. Therefore, x is not periodic, and we've proven (i).

Take μ to be any invariant measure on X. If $\mu([w]) > 0$ for some $w \in O$, then by Poincaré recurrence there would exist a point of X which returns to $[w]$ infinitely often under σ, contradicting the definition of O. Then clearly μ is a measure on X_O as well, and so $\mathcal{M}(X) \subseteq \mathcal{M}(X_O)$. The reverse containment is obvious, and so (ii) is proved.

To prove (iii), first consider any $u, v \in \mathcal{L}(X_O)$. Then by definition of X_O, there exist words $\overline{u}, \overline{v} \in \mathcal{L}(X)$ which contain $m+1$ disjoint occurrences of u and v respectively; denote by N the sum of their lengths. For any $t \in \mathbb{N}$, use left almost specification of X to construct a word

$$y^{(t)} := u^{(1)} v^{(1)} \dots u^{(2t(mt|A|^t+1))} v^{(2t(mt|A|^t+1))} \in \mathcal{L}(X),$$

where $u^{(i)} \in B_m(\overline{u})$ and $v^{(i)} \in B_m(\overline{v})$ for each $1 \leq i \leq 2t(mt|A|^t + 1)$. (In fact $v^{(2t(mt|A|^t+1))}$ can be taken to be $\overline{v}$, but that won't be necessary for our proof.) Since $y^{(t)} \in \mathcal{L}(X)$, it does not contain $m+1$ disjoint occurrences of any word from O. Since there are trivially less than or equal to $|A|^t$ words of length t in O, there are at most $mt|A|^t$ letters of $y^{(t)}$ which are part of an occurrence of an t-letter word from O within $y^{(t)}$. Therefore, since $y^{(t)}$ has length $2tN(mt|A|^t + 1)$, there exists a subword of $y^{(t)}$ with length $2tN - 1$ which contains no t-letter words from O, which must itself contain a subword $z^{(t)}$ of length tN of the form $u^{(j+1)}v^{(j+1)} \dots u^{(j+t)}v^{(j+t)}$ for some j.

Now we construct a point $x \in X_O$ as a 'weak limit' of the words $z^{(t)}$. More precisely, taking t to be even for simplicity, we consider $x^{(t)} \in X$ such that $x^{(t)}([-\frac{t}{2}N, \frac{t}{2}N-1]) = z^{(t)}$, and then pass to a subsequence $t_\ell \to \infty$ such that $x^{(t_\ell)}([i,j])$ is eventually constant for every $i < j$. Writing $x(i) = \lim_\ell x^{(t_\ell)}(i)$, we arrive at a point $x = \dots U^{(-1)}V^{(-1)}U^{(0)}V^{(0)}U^{(1)}V^{(1)} \dots$ where $U^{(i)} \in B_m(\overline{u})$ and $V^{(i)} \in B_m(\overline{v})$ for every $i \in \mathbb{Z}$. Since $x^{(t)} \in X$ and X is closed, we have $x \in X$. Furthermore, since each $z^{(t)}$ was devoid of t-letter words from O, we have $x \in X_O$. Therefore, its subword $U^{(0)}V^{(0)}$ is in $\mathcal{L}(X_O)$. Since $U^{(0)} \in B_m(\overline{u})$, it contains an occurrence of u, and since $V^{(0)} \in B_m(\overline{v})$, it contains an occurrence of v. Then, $U^{(0)}V^{(0)}$ contains a subword starting with u and ending with v, which is in $\mathcal{L}(X_O)$, proving the desired irreducibility of X_O.

Finally, we must prove (iv). Consider any words $v, w \in \mathcal{L}(X_O)$. We will first inductively create a sequence $y^{(n)}$ of words in $\mathcal{L}(X)$ where for every n, $y^{(n)}$ contains n disjoint subwords $v^{(1)}w, \dots, v^{(n)}w$ with $v^{(i)} \in B_m(v)$ for each i. Then we will mimic the proof of (iii) to construct an element of X_O that contains a subword $v'w$ with $v' \in B_m(v)$.

For $n = 1$, we can use LAS to create a word $v^{(1)}w \in \mathcal{L}(X)$ with $v^{(1)} \in B_m(v)$, and we define $y^{(1)} = v^{(1)}w$. Proceeding inductively, suppose that $y^{(n)} \in \mathcal{L}(X)$ has already been defined. Let i_n be the maximum number of changes required to concatenate a word from X_O on the left of $y^{(n)}$, i.e.

$$i_n = \max_{u \in \mathcal{L}(X_O)} \min\{i \ : \ \exists u' \in B_i(u) \text{ s.t. } u'y^{(n)} \in \mathcal{L}(X)\}.$$

(Clearly, $i_n \leq m$ by definition of LAS.) Then, take $u^{(n)} \in \mathcal{L}(X_O)$ which achieves this maximum, i.e. $u^{(n)}$ requires i_n changes to be concatenated to the left of $y^{(n)}$. By irreducibility of X_O, there exists $t^{(n)} \in \mathcal{L}(X_O)$ so that $wt^{(n)}u^{(n)} \in \mathcal{L}(X_O)$. Then, by definition of i_n, there exists $w't'^{(n)}u'^{(n)} \in B_{i_n}(wt^{(n)}u^{(n)})$ (where the lengths of w', $t'^{(n)}$, $u'^{(n)}$ are those of w, $t^{(n)}$, $u^{(n)}$ respectively) such that $w't'^{(n)}u'^{(n)}y^{(n)} \in \mathcal{L}(X)$. However, this implies that $u'^{(n)}y^{(n)} \in \mathcal{L}(X)$, and by definition of $u^{(n)}$, $d(u^{(n)}, u'^{(n)}) \geq i_n$. Therefore, $w' = w$ and $t'^{(n)} = t^{(n)}$. We therefore have a word $wt^{(n)}u'^{(n)}y^{(n)} \in \mathcal{L}(X)$, and by LAS, we can define $v^{(n+1)} \in B_m(v)$ so that $v^{(n+1)}wt^{(n)}u'^{(n)}y^{(n)} \in \mathcal{L}(X)$. Define $y^{(n+1)} := v^{(n+1)}wt^{(n)}u'^{(n)}y^{(n)}$. By the inductive hypothesis, $y^{(n)}$ already contained disjoint subwords $v^{(1)}w, \dots, v^{(n)}w$ as desired, and so

$y^{(n+1)}$ contains all of those along with its prefix $v^{(n+1)}w$, which is disjoint from all of them. This completes the inductive definition of the sequence $y^{(n)}$.

Now, for every t, consider the word $z^{(t)} := y^{((t+1)(mt|A|^t+1))}$. This word is in $\mathcal{L}(X)$, and contains $(t+1)(mt|A|^t+1)$ disjoint words of the form $v^{(i)}w$ with $v^{(i)} \in B_m(v)$. Since $z^{(t)} \in \mathcal{L}(X)$, it does not contain $m+1$ disjoint occurrences of any word from O. As before, at most $mt|A|^t$ letters of $z^{(t)}$ are part of an occurrence of an t-letter word from O within $z^{(t)}$. Therefore, there exists a subword of $z^{(t)}$, call it $x^{(t)}$, which contains t disjoint words $v^{(j+1)}w, \dots, v^{(j+t)}w$ and contains no t-letter words from O. Then, we may take a weak limit of the words $x^{(t)}$ to arrive at a point $x = \dots v'w \dots$, where $v' \in B_m(v)$. As above, $x \in X_O$, and so its subword $v'w$ is in $\mathcal{L}(X_O)$. Since v, w were arbitrary, we've proved LAS for X_O, which verifies (iv) and completes the proof of Theorem 3.1. □

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 that will be important. Recall from Lemma 2.6 that $|\mathcal{L}_n(X)| \geq e^{nh(X)}$ for every n. In general the definition of $h(X)$ gives the upper bound $|\mathcal{L}_n(X)| \leq C_n e^{nh(X)}$ for some subexponential C_n. An important part of the uniqueness proof in [3, 5] 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.

Lemma 3.2. *If X is a subshift with almost specification (or left almost specification) with constant $g(n) = m$, then for every n,*

$$|\mathcal{L}_n(X)| \leq |A|^{2m} n^{2m} e^{nh(X)}. \tag{3.1}$$

Proof. Suppose that X is a subshift satisfying AS with $g = m$ (by Lemma 2.16 this includes the LAS case). Use a greedy algorithm to 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

$$|S_n| \geq |\mathcal{L}_n(X)| \cdot |A|^{-2m} n^{-2m}. \tag{3.2}$$

For any $k \in \mathbb{N}$, almost specification lets us define $f\colon (S_n)^k \to \mathcal{L}_{kn}(X)$ by $f(w^{(1)}, \dots, w^{(k)}) = w'^{(1)} \dots w'^{(k-1)} w'^{(k)} \in \mathcal{L}_{kn}(X)$ where $w'^{(i)} \in B_m(w^{(i)})$ for $1 \leq i \leq k$. Then the map f is clearly injective, since for any $s \neq s'$ in S, the Hamming balls $B_m(s)$ and $B_m(s')$ are disjoint. Using (3.2), this gives the bound

$$|\mathcal{L}_{kn}(X)| \geq |S_n|^k \geq |\mathcal{L}_n(X)|^k |A|^{-2mk} n^{-2mk},$$
$$\frac{1}{kn} \ln |\mathcal{L}_{kn}(X)| \geq \frac{1}{n} \ln |\mathcal{L}_n(X)| - \frac{2m}{n} \ln(|A|n).$$

Sending $k \to \infty$ gives $\frac{1}{n} \ln |\mathcal{L}_n(X)| \leq h(X) + \frac{2m}{n} \ln(|A|n)$; then multiplying by n and taking exponentials gives (3.1).

□

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 3.3. *Suppose that X is an entropy minimal subshift and that there exist $C, t > 0$ so that $|\mathcal{L}_n(X)| \leq Cn^t e^{nh(X)}$ for every n. Then, for every $w \in \mathcal{L}(X)$, there exist $\epsilon > 0$ and a set $S \subset \mathbb{N}$ with positive upper density for which $|\mathcal{L}_s(X) \cap \mathcal{L}(X)w| > \epsilon s^{-1/2} e^{sh(X)}$ for all $s \in S$.*

Proof. Suppose that X, C, and t are as in the theorem. For all n, Lemma 2.6 gives $|\mathcal{L}_n(X)| \geq e^{nh(X)}$, while for $n > C$, we have $|\mathcal{L}_n(X)| < n^{t+1} e^{nh(X)}$. Therefore, there must exist $q \in \mathbb{N}$ so that there are infinitely many n with $|\mathcal{L}_n(X)| \geq n^{(q-1)/2} e^{nh(X)}$, but only finitely many n with $|\mathcal{L}_n(X)| > n^{q/2} e^{nh(X)}$. Define N to be the max of the latter finite set, and define (n_k) a sequence from the former set.

Now we recall the construction of an MME from Lemma 2.10. For every k, define the measure $\nu_k = \frac{1}{|\mathcal{L}_{n_k}(X)|} \sum_{v \in \mathcal{L}_{n_k}(X)} \delta_{x_v}$, where for each $v \in \mathcal{L}(X)$, x_v is an arbitrary member of $[v]$. Then, define μ_k as $\frac{1}{n_k} \sum_{i=0}^{n_k - 1} \sigma_*^i \nu_k$, and let μ be any weak* limit point of the μ_k. Without loss of generality, we assume that $\mu_k \to \mu$ (pass to a subsequence of n_k if necessary). Then μ is shift-invariant, and is an MME for X.

Since X is entropy minimal, $\mu([w]) > 0$. Fix $\alpha > 0$ such that $5\alpha < \mu([w])$; then by passing to a further subsequence of n_k if necessary, we assume without loss of generality that for every k we have

$$\mu_k([w]) > 5\alpha \text{ and } \alpha n_k \geq |w| + N. \tag{3.3}$$

Now we fix a word $w \in \mathcal{L}(X)$, and obtain a lower bound on how many times w appears in words $v \in \mathcal{L}_{n_k}(X)$. Given $v \in \mathcal{L}(X)$, let

$$\mathrm{Sub}(w, v) := \{i \in [|w|, |v|] : v([i - |w|, i]) = w\}$$

be the set of indices that mark the end of an instance where w appears as a subword of v. For every k, we have

$$\begin{aligned} 5\alpha < \mu_k([w]) = \frac{1}{n_k} \sum_{i=0}^{n_k - 1} \nu_k(\sigma^{-i}[w]) &= \frac{1}{n_k |\mathcal{L}_{n_k}(X)|} \sum_{v \in \mathcal{L}_{n_k}(X)} \sum_{i=0}^{n_k - 1} \chi_{[w]}(\sigma^i x_v) \\ &\leq \frac{1}{n_k |\mathcal{L}_{n_k}(X)|} \sum_{v \in \mathcal{L}_{n_k}(X)} (|\mathrm{Sub}(w, v)| + |w|). \end{aligned}$$

Observing that (3.3) gives $|w| \leq \alpha n_k$, we conclude that

$$\frac{1}{|\mathcal{L}_{n_k}(X)|} \sum_{v \in \mathcal{L}_{n_k}(X)} |\mathrm{Sub}(w, v)| \geq 4\alpha n_k. \tag{3.4}$$

Now given $i \in \mathbb{N}$, let

$$\mathcal{H}(w,i) = \{v \in \mathcal{L}(X) : i \in \mathrm{Sub}(w,v)\}$$

be the set of words in which an appearance of w ends at position i. Let

$$S_k' := \{i \in [|w|, n_k] : |\mathcal{H}_{n_k}(w,i)| \geq \alpha |\mathcal{L}_{n_k}(X)|\}$$

be the set of indices for which at least α of the words of length n_k contain w as a subword ending at index i. Then we have

$$\begin{aligned}\frac{1}{|\mathcal{L}_{n_k}(X)|} \sum_{v \in \mathcal{L}_{n_k}(X)} |\mathrm{Sub}(v,w)| &= \sum_{i=|w|}^{n_k} \frac{|\mathcal{H}_{n_k}(w,i)|}{|\mathcal{L}_{n_k}(X)|} \\ &\leq |S_k'| + (n_k - |S_k'|)\alpha \leq |S_k'| + \alpha n_k,\end{aligned}$$

which together with (3.4) gives $|S_k'| \geq 3\alpha n_k$. Let $S_k = S_k' \cap [\alpha n_k, (1-\alpha)n_k]$, then we have

$$|S_k| \geq \alpha n_k \text{ for all } k \in \mathbb{N}. \tag{3.5}$$

Let $S = \bigcup_k S_k$. We claim that S satisfies the conclusion of the theorem. Note that (3.5) immediately implies that S has positive upper density (at least α). We must estimate $|\mathcal{L}_s(X) \cap \mathcal{L}(X)w|$ from below for $s \in S$.

Let $i \in S_k$ for some k. By the definition of S_k and of the subsequence n_k,

$$|\mathcal{H}_{n_k}(w,i)| \geq \alpha \mathcal{L}_{n_k}(X) \geq \alpha n_k^{(q-1)/2} e^{n_k h(X)}. \tag{3.6}$$

On the other hand, if $v \in \mathcal{H}_{n_k}(w,i)$, then $v([1,i]) \in \mathcal{L}_i(X) \cap \mathcal{L}(X)w$, so

$$|\mathcal{H}_{n_k}(w,i)| \leq |\mathcal{L}_i(X) \cap \mathcal{L}(X)w| \cdot |\mathcal{L}_{n_k - i}(X)|. \tag{3.7}$$

Since $S_k \subset [\alpha n_k, (1-\alpha)n_k]$, we have $n_k - i \geq \alpha n_k \geq N$, so (3.3) gives

$$|\mathcal{L}_{n_k - i}(X)| \leq (n_k - i)^{q/2} e^{(n_k - i)h(X)} \leq n_k^{q/2} e^{(n_k - i)h(X)}.$$

Together with (3.6) and (3.7), this gives

$$\alpha n_k^{q/2} n_k^{-1/2} e^{n_k h(X)} \leq |\mathcal{H}_{n_k}(w,i)| \leq |\mathcal{L}_i(X) \cap \mathcal{L}(X)w| \cdot n_k^{q/2} e^{(n_k - i)h(X)},$$

and dividing through by $n_k^{q/2} e^{n_k h(X)}$ gives

$$\alpha n_k^{-1/2} e^{ih(X)} \leq |\mathcal{L}_i(X) \cap \mathcal{L}(X)w|,$$

which proves Theorem 3.3. □

Remark 3.4. In fact Theorem 3.3 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 $|\mathcal{L}_s(X) \cap \mathcal{L}(X)w| > \epsilon f(n)^{-1/k} e^{sh(X)}$ as a conclusion. In particular, we note that we could bound $|\mathcal{L}_s(X) \cap \mathcal{L}(X)w|$ from below by some constant multiple of $n^{-\beta} e^{sh(X)}$ for any desired $\beta > 0$; for our purposes, $\beta = 1/2$ will suffice.

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

Theorem 3.5. *If X is an irreducible subshift with left almost specification with constant $g(n) = m$, then X is entropy minimal.*

Proof. Take X a subshift as in the theorem. By Proposition 2.8 of [12], X contains a subshift Y with $h(Y) = h(X)$ for which Y is entropy minimal. 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 3.2, there exist $C, t > 0$ such that for all n, we have $|\mathcal{L}_n(X)| \leq Cn^t e^{nh(X)}$. Since $Y \subseteq X$, the same bound holds for $|\mathcal{L}_n(Y)|$, and since $h(Y) = h(X)$, we can apply Theorem 3.3 to Y, obtaining $S \subset \mathbb{N}$ with positive upper density and $\epsilon > 0$ so that

$$M_s := |\mathcal{L}_s(Y) \cap \mathcal{L}(Y)v| > \epsilon s^{-1/2} e^{sh(Y)} = \epsilon s^{-1/2} e^{sh(X)} \tag{3.8}$$

for all $s \in S$, where v is the 'maximally changing word' from above.

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. Without loss of generality, we assume that all unequal elements of S are separated by distance at least N; this can be done by passing to a subset. We choose $t \in \mathbb{N}$ a parameter to be defined later, and define $s_1 < s_2 < \ldots < s_t$ to be the smallest t elements of S. We also make the notation $M = \sum_{i=1}^t s_i$.

We will make many words in $\mathcal{L}_n(X)$ by using left almost specification to almost concatenate words in various $\mathcal{L}_{s_i}(Y)$ with copies of w' 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 $[1, t]$ and words $w^{(i,j)} \in \mathcal{L}_{s_i}(Y) \cap \mathcal{L}(Y)v$ 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 $u^{(\ell)}$ 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 $u^{(tk)}$, 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 $z'^{(tk-1)}$ with length less than or equal to N. Then

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

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

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

$$f(\boldsymbol{\pi}, \mathbf{w}) = U^{(1)}z'^{(1)}U^{(2)}z'^{(2)} \dots U^{(tk-1)}z'^{(tk-1)}u^{(tk)} \in \mathcal{L}_{k(M+tN)}(X). \tag{3.9}$$

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

We claim that f is injective. Suppose $(\boldsymbol{\pi}_1, \mathbf{w}_1) \neq (\boldsymbol{\pi}_2, \mathbf{w}_2)$, that $(\boldsymbol{\pi}_1, \mathbf{w}_1)$ induces $u_1^{(1)}, \dots, u_1^{(tk)}$ as in (b) 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 $1 \leq \ell \leq tk$. 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 $z'^{(\ell+1)}U^{(\ell+1)} \dots z'^{(tk-1)}u^{(tk)}$. Then, at the next step, the unequal words $u_1^{(\ell)}$ and $u_2^{(\ell)}$ are extended to the left by $N - |z'^{(\ell+1)}|$ units, which yields unequal words $u_1'^{(\ell)}$ and $u_2'^{(\ell)}$, and then left almost specification is used to find $U_1^{(\ell)}$ and $U_2^{(\ell)}$ which can be appended to the left of $z'^{(\ell+1)}U^{(\ell+1)} \dots z'^{(tk-1)}u^{(tk)}$. However, as noted above, the only changes made are inside the terminal copies of v within $u_1'^{(\ell)}$ and $u_2'^{(\ell)}$, 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 $|u_1^{(\ell)}| > |u_2^{(\ell)}|$. By assumption on S, $|u_2^{(\ell)} - u_1^{(\ell)}| \geq N$. 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 $z'^{(\ell+1)}U^{(\ell+1)} \dots z'^{(tk-1)}u^{(tk)}$.

Then, in the construction of $f(\boldsymbol{\pi}_2, \mathbf{w}_2)$, once $U_2^{(\ell)}$ is appended to the left of $z'^{(\ell+1)}U^{(\ell+1)} \dots z'^{(tk-1)}u^{(tk)}$, somewhere in the N letters immediately to the left an occurrence of u will be added at the next step (inside some $z_2'^{(\ell)} \in B_m(z)$). However, the corresponding locations in $U_1^{(\ell)}z'^{(\ell+1)}U^{(\ell+1)} \dots z'^{(tk-1)}u^{(tk)}$ are part of $U_1^{(\ell)} \in \mathcal{L}(Y)$ since $|u_1^{(\ell)}| \geq |u_2^{(\ell)}| + N$, 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 $(t! \prod_{i=1}^t M_{s_i})^k$, where we recall that M_{s_i} was defined in (3.8) and satisfies $M_{s_i} \geq \epsilon s_i^{-1/2} e^{s_i h(X)}$. Thus

$$\prod_{i=1}^t M_{s_i} \geq \prod_{i=1}^t \epsilon s_i^{-1/2} e^{s_i h(X)} = e^{Mh(X)} \prod_{i=1}^t \epsilon s_i^{-1/2} = e^{(M+tN)h(X)} \prod_{i=1}^t \beta s_i^{-1/2}$$

for $\beta = \epsilon e^{-Nh(X)}$; this gives the bound

$$|\mathcal{L}_{k(M+tN)}(X)| \geq e^{k(M+tN)h(X)} \left(t! \prod_{i=1}^t \beta s_i^{-1/2} \right)^k. \tag{3.10}$$

We now recall that S has positive upper density, and so there exist $D > 0$ and a sequence $t_n \to \infty$ so that for all n, $s_{t_n} < Dt_n$. For any such n,

$$\prod_{i=1}^{t_n} s_i^{-1/2} \geq s_{t_n}^{-t_n/2} \geq D^{-t_n/2}(t_n)^{-t_n/2}. \tag{3.11}$$

Combining (3.10) and (3.11), we see that for any k and n,

$$|\mathcal{L}_{k(M+t_nN)}(X)| \geq e^{k(M+t_nN)h(X)} \left((t_n)!(\beta/\sqrt{D})^{t_n} t_n^{-t_n/2} \right)^k. \tag{3.12}$$

The reader may check that $(t_n)!(\beta/\sqrt{D})^{t_n} t_n^{-t_n/2} \to \infty$ as $n \to \infty$, so there exists n for which this quantity is greater than 2; fix such an n for the remainder of the proof. Now, we take logarithms in (3.12), divide by $k(M + t_nN)$, and let $k \to \infty$, which implies that $h(X) \geq h(Y) + \frac{\log 2}{M+t_nN}$, 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. $\square$

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 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.

$$i = \max_{u \in \mathcal{L}(X), v \in \mathcal{L}(X)} \min\{j \ : \ \exists u' \in B_j(u) \text{ for which } u'v \in \mathcal{L}(X)\}. \tag{3.13}$$

We have $i \leq m$ from the LAS property. Choose words $u \in \mathcal{L}(X)$ and $v \in \mathcal{L}(X)$ achieving this maximum, i.e. for $j < i$, there exists no $u' \in B_j(u)$ for which $u'v \in \mathcal{L}(X)$. 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 [4] 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 the LAS property. We prove it with $\tau = |u|$ by showing that for any $w^{(1)}, w^{(2)} \in \mathcal{G}$, there exists u' with $|u'| = |u|$ for which $w^{(1)}u'w^{(2)} \in \mathcal{G}$. Given any such $w^{(1)}, w^{(2)}$, by definition we can write $w^{(1)} = vt^{(1)}$, $w^{(2)} = vt^{(2)}$, and $vt^{(1)}u, vt^{(2)}u \in \mathcal{L}(X)$. Then, by LAS and the definition of i, there exists $v't'^{(1)}u'vt^{(2)}u \in \mathcal{L}(X)$ where $v't'^{(1)}u' \in B_i(vt^{(1)}u)$ and $v', t'^{(1)}, u'$ have the same lengths as v, $t^{(1)}$, u. However, since $u'v \in \mathcal{L}(X)$, by (3.13) we have $d(u', u) \geq i$. This implies that $v' = v$ and $t'^{(1)} = t^{(1)}$. We then have that $vt^{(1)}u'vt^{(2)}u \in \mathcal{L}(X)$, implying that $vt^{(1)}u'vt^{(2)} = w^{(1)}u'w^{(2)} \in \mathcal{G}$ and proving **[I]**.

Now we prove **[II]** by showing that $h(\mathcal{C}) < h(X)$. First, we note that by Theorem 3.5, 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

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

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 $h(\mathcal{C}') \leq h(X_u)$, 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, 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've 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

$$|\mathcal{B}_n| \leq |(\mathcal{C}')_n| + \sum_{i=0}^{n-|u|} |(\mathcal{C}^p)_i| \cdot |(\mathcal{C}')_{n-i}|. \tag{3.14}$$

Choose $\epsilon > 0$ so that $h(\mathcal{C}^p), h(\mathcal{C}') < h(X) - 2\epsilon$. Then there is K such that $|\mathcal{C}_n^p|, |\mathcal{C}_n'| \leq Ke^{n(h(X)-\epsilon)}$ for every $n \in \mathbb{N}$. Combining this with (3.14) yields

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

Taking logs, dividing by n, and letting $n \to \infty$ yields $h(\mathcal{B}) \leq h(X) - \epsilon < h(X)$. 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 $x \in X$ and $i \leq j \leq k \leq \ell$ are such that $x([i,k]), x([j,\ell]) \in \mathcal{G}$ and $k - j \geq L$. Then v is a prefix of $x([j,\ell])$; it is also a prefix of $x([j,k])$ since $k - j \geq L = |v|$. 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 $a \leq i$ 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

$$\gcd\{|w| + |u| : w \in \mathcal{G}\} = 1. \tag{3.15}$$

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 (3.15) and completes the proof of Theorem 1.1, modulo the derivation in the next section of Theorem 2.22 from the results in [4].

3.4. **Proof of Theorem 2.22.** Theorem 2.22 is a consequence of results in [4]. Since it is not stated in exactly this form there, we explain the necessary steps. First note that [4] 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 [4, Theorem 3.2].

The main result in that paper is [4, 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 $[\mathbf{III}_{\mathrm{a}}]$ and $[\mathbf{III}_{\mathrm{b}}]$. 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 [4] allows for a period d in the Bernoulli property and exponential decay of correlations;
- the definition of periodic orbit measures in [4] 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 $[\mathbf{III}_{\mathrm{a}}]$ and $[\mathbf{III}_{\mathrm{b}}]$ 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 [4, §3], which establish [4, Theorem 1.1]. In [4, Theorem 3.2], it is shown that **[I]**, **[II]**, $[\mathbf{III}_{\mathrm{a}}]$, and $[\mathbf{III}_{\mathrm{b}}]$ 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, [4, Theorem 3.2] guarantees two further conditions **[II′]** and $[\mathbf{III}^*]$, which we do not state here but which are used in [4, Theorem 3.1] to give the conclusions of [4, 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 $[\mathbf{III}_{\mathrm{a}}]$ and $[\mathbf{III}_{\mathrm{b}}]$ can replace **[III]** in [4, Theorem 1.1], which deals with the first two issues raised above.

It only remains to address the issue concerning periodic orbit measures. By [4, Lemma 4.5] applied to $\mathcal{F}$, there are $Q > 0$ and $N \in \mathbb{N}$ such that for sufficiently large n 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 $n \geq M + N$ 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

$$|\operatorname{Per}_n(X)| \geq |\mathcal{F}_n| \geq |\mathcal{F}_j| \geq Qe^{-(M+N)h(X)}e^{nh(X)}, \tag{3.16}$$

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 $\mu_n = \frac{1}{|\operatorname{Per}_n(X)|}\sum_{x\in\operatorname{Per}_n(X)}\delta_x$ 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 Lemma 2.10. By (3.16), clearly $\liminf \frac{1}{n_k}\ln|\operatorname{Per}_{n_k}(X)| = h(X)$, 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 $g = 1$ implies either left or right almost specification.

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

Proof. Suppose that X is an irreducible subshift satisfying almost specification with $g(n) = 1$, and for a contradiction assume that it has neither LAS with $g = 1$ nor RAS with $g = 1$. 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 $vwU \in \mathcal{L}(X)$. Then, use almost specification to create a word $u'yV' \in \mathcal{L}(X)$, where $u' \in B_1(u)$, $y \in B_1(vwU)$, and $V' \in B_1(V)$. Clearly either the initial v or terminal U from vwU is unchanged in y. If v is unchanged, then $u'v \in \mathcal{L}(X)$, a contradiction. If U is unchanged, then $UV' \in \mathcal{L}(X)$, a contradiction. Either way, our assumption must have been false, and so the lemma is proved. □

We also need the following result, which we state without proof, as the proof is completely analogous to that of Theorem 3.1.

Theorem 4.2. *For any X with almost specification with $g = m$, define by O the set of words $w \in \mathcal{L}(X)$ for which no point of X contains $m+1$ disjoint occurrences of w. Define X_O to be the subshift of points of X containing no word from O. Then*

(i) *X and X_O have the same sets of periodic points;*
(ii) *X and X_O have the same simplices of invariant measures;*
(iii) *X_O is irreducible;*

(iv) X_O *has almost specification with* $g = m$.

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 $g = 1$ or RAS with $g = 1$. In the former case, the proof is complete by applying Theorem 1.1.

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

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

Then $\rho(X)$ is a subshift, and since X was irreducible and had RAS with $g = 1$, it's clear that $\rho(X)$ is irreducible and has LAS with $g = 1$. 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 even $g = O(\log\log n)$ can coexist with multiple MMEs. We will use the following technical lemma from [9].

Lemma 4.3. ([9], Lemma 4.4) *For every alphabet* A *and positive integer* n, *there exists a set* $U_n \subseteq A^n$ *such that* $|U_n| \leq \frac{16}{n^2}|A|^n$ *and* U_n *is* 2*-spanning with respect to the Hamming metric.*

To prove Theorem 1.4, we build a shift over the alphabet

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

where $N > 2^{17} + 4$. We begin by using Lemma 4.3 to fix, for every n, a set $U_n \subseteq \{1, \dots, N\}^n$ which is 2-spanning in $\{1, \dots, N\}^n$ and with $|U_n| \leq \frac{16}{n^2}N^n$. We will use U_n to define certain collections $T^+ \subset \{1, \dots, N\}^*$ and $T^- \subset \{-1, \dots, -N\}^*$, 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, define $k = \lfloor \log_2 \log_2 n \rfloor$; then $2^{2^k} \leq n < 2^{2^{k+1}}$. For any n, we then define T_n^+ to be the set

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

and define $T^+ = \bigcup_n T_n^+$. We first note that the collection T^+ is suffix-closed, meaning that $w \in T^+$, $w = uv \implies v \in T^+$; 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 $\{1, \dots, N\}^n$; given a word in $\{1, \dots, N\}^n$, one only needs to change the last 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

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

For $k \geq 4$, $2^{k+2} - 6k - 4 > 1.125 \cdot 2^{k+1}$, and so for $n \geq 2^{16}$,

$$|T_n^+| \leq \frac{N^n}{N \cdot n^{1.125}}. \tag{4.1}$$

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 suffix-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 (4.1). 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 first $2^{2^{k+1}} - 2^{2^k} - 1 > k$ letters of words in $T^+_{2^{2^{k+1}}-1}$ are completely unconstrained. In other words, every word in $\{1, \dots, N\}^*$ is the prefix 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 prefix of some word in T^-, and so $\{-N, \dots, -1\}^{\mathbb{Z}} \subset X$. Trivially, this implies that $h(X) \geq \log N$. We will now bound $|\mathcal{L}_n(X)|$ from above similarly to [9] 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 suffix-closed, and that every word of constant sign is the prefix of a word in T. Therefore, we can rephrase by saying that $w^{(1)}, \dots, w^{(k-1)} \in T$, and $w^{(k)}$ has constant sign. Defining $n_i = |w^{(i)}|$ for $1 \leq i \leq k$, we see that

$$\begin{aligned} |\mathcal{L}_n(X)| &\leq \sum_{k=1}^{n} \sum_{\sum_{i=1}^k n_i = n} 2N^{n_k} \prod_{i=1}^{k-1} |T_{n_i}| \\ &\leq 2N^n \sum_{k=1}^{n} \sum_{\sum_{i=1}^k n_i = n} \prod_{i=1}^{k-1} \frac{|T_{n_i}|}{N^{n_i}} \leq 2N^n \sum_{k=1}^{n} \left(\sum_{j=1}^{\infty} \frac{|T_j|}{N^j} \right)^{k-1}. \end{aligned} \tag{4.2}$$

By (4.1), we have

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

Since $N > 2^{17}+4$, we can define $\alpha := \frac{2^{17}+4}{N} < 1$, and then (4.2) shows that $|\mathcal{L}_n(X)| < 2N^n \frac{1}{1-\alpha}$. Taking logs, dividing by n, and letting $n \to \infty$ shows that $h(X) \leq \log N$, 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.

It only remains to show that X has LAS with $g(n) = 1 + 2\log_2 \log_2 n$. To see this, consider any $v \in \mathcal{L}_m(X)$ and $w \in \mathcal{L}_n(X)$. 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, $w^{(n)}$ is a prefix of a word t in T, and all $w^{(j)}$ for $j < n$ are in T (recall that T is suffix-closed). Similarly, $v^{(j)}$ is in T for $j < m$. Since T_m is $(1+2\lfloor \log_2 \log_2 m \rfloor)$-spanning in $\{-N, \dots, -1\}^m \cup \{1, \dots, N\}^m$ and $v^{(m)}$ has constant sign, there exists $u^{(m)} \in B_{1+2\lfloor \log_2 \log_2 m \rfloor}(v^{(m)})$ which is also in T. Since it is a concatenation of words in T, $v^{(1)} \dots v^{(m-1)} u^{(m)} w^{(1)} \dots w^{(n-1)} t \in \mathcal{L}(X)$. Then its subword $v^{(1)} \dots v^{(m-1)} u^{(m)} w^{(1)} \dots w^{(n-1)} w^{(n)}$ must also be in $\mathcal{L}(X)$. Finally, since $d(v^{(1)} \dots v^{(m-1)} u^{(m)}, v) = d(u^{(m)}, v^{(m)}) \leq 1 + 2\lfloor \log_2 \log_2 m \rfloor \leq 1 + 2\log_2 \log_2 m$, we have shown that X has LAS with $g(n) = 1 + 2\log_2 \log_2 n$.

4.3. **Proof of Theorem 1.7.** We will show that no matter how slowly an unbounded mistake function g grows, LAS with mistake function g does not imply entropy minimality.

Consider any unbounded $g(n)$, and assume without loss of generality that $g = o(n)$ (this is because 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 $g(n)$.

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 $v' = 0^{|v|} \in B_{|v|}(v)$. Since $w \in \mathcal{L}(X)$, there exists some point $x \in X$ containing w. Then, if we create x' by changing the $|v|$ letters before an occurrence of w to all 0s, clearly $x' \in X$ by the rules defining X, and x' contains $v'w$, 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 without loss of generality assume that they have the same length n. As above, it's clear that $x^{(v)} = 0^\infty v 0^\infty$ and $x^{(w)} = 0^\infty w 0^\infty$ are both in X. Since g is unbounded, there exists N so that $g(N) \geq 2n$. Then, consider the point $x^{(v,w)} = 0^\infty v 0^N w 0^\infty$. Any subword of $x^{(v,w)}$ which does not contain letters of both v and w is a subword of either $x^{(v)}$ or $x^{(w)}$, and so satisfies the rules defining X. Any other subword w of $x^{(v,w)}$ contains letters of both v and w, and so has length greater than N. Since the number of 1 symbols in $x^{(v,w)}$ is less than or equal to $2n$, the same is true of w, and we know that

$g(N) \geq 2n$. Therefore, all subwords of $x^{(v,w)}$ satisfy the rules defining X, and so $x^{(v,w)} \in X$, proving irreducibility of X.

We also claim that $h(X) = 0$. To see this, note that since $g = o(n)$, the limiting frequency of 0 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.

4.4. **Proof of Proposition 1.8.** 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 $g = m$ and $g = n$ 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 $u'^{(1)} \in B_m(u^{(1)})$, $v'^{(1)} \in B_n(v^{(1)})$ so that $u'^{(1)}u^{(2)} \in \mathcal{L}(X)$ and $v'^{(1)}v^{(2)} \in \mathcal{L}(Y)$. Then, if we define $w'^{(1)}$ to have first and second coordinates $u'^{(1)}$ and $v'^{(1)}$, we get $w'^{(1)}w^{(2)} \in \mathcal{L}(X \times Y)$, and $w'^{(1)} \in B_{m+n}(w^{(1)})$. Therefore, $X \times Y$ has LAS with $g = m + n$.

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

4.5. **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

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

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

$$|\mathcal{D}_n^{(kn)}| \geq e^{n(h(\mathcal{D})-2\epsilon)} \tag{4.3}$$

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

We start by proving (4.3). 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

$$e^{nh(\mathcal{D})} \leq |\mathcal{D}_n| \leq e^{n(h(\mathcal{D})+\epsilon)} \tag{4.4}$$

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

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

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

$$|\mathcal{D}_{(kn)}^{(kn)}| \leq |\mathcal{D}_n^{(kn)}|^k. \tag{4.5}$$

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 (4.4) and (4.5). Dividing by $e^{2kn(h(\mathcal{D})+\epsilon)}$ and taking a kth root gives (4.3) for all $k > N/n$.

Now with n, ϵ fixed, we have proved that for all $k > N(\epsilon)/n$ there are at least $e^{n(h(\mathcal{D})-2\epsilon)}$ words of length n that can be extended by kn letters on each side to form a word in $\mathcal{D}$. Thus there are at least this many words of length n that have such an extension for infinitely many values of k, and all such words are in $\mathcal{L}(X(\mathcal{D}))$. That is, we have proved that $\mathcal{C}_K := \bigcup_{k \geq K} \mathcal{D}_n^{(kn)}$ has $|\mathcal{C}_K| \geq e^{n(h(\mathcal{D})-2\epsilon)}$ for every $K \in \mathbb{N}$, and moreover we have

$$\bigcap_{K \in \mathbb{N}} \mathcal{C}_K = \bigcap_{K \in \mathbb{N}} \bigcup_{k \geq K} \mathcal{D}_n^{(kn)} \subset \mathcal{L}_n(X(\mathcal{D})).$$

Since $\mathcal{C}_1 \supset \mathcal{C}_2 \supset \mathcal{C}_3 \cdots$ and all $\mathcal{C}_i$ are finite, we have

$$|\mathcal{L}_n(X(\mathcal{D})) \geq \left| \bigcap_{K \in \mathbb{N}} \mathcal{C}_K \right| \geq \inf_K |\mathcal{C}_K| \geq e^{n(h(\mathcal{D})-2\epsilon)},$$

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

References

[1] Anne Bertrand. Specification, synchronisation, average length. In *Coding theory and applications (Cachan, 1986)*, volume 311 of *Lecture Notes in Comput. Sci.*, pages 86–95. Springer, Berlin, 1988.

[2] Rufus Bowen. Periodic points and measures for Axiom A diffeomorphisms. *Trans. Amer. Math. Soc.*, 154:377–397, 1971.

[3] Rufus Bowen. Some systems with unique equilibrium states. *Math. Systems Theory*, 8(3):193–202, 1974/75.

[4] Vaughn Climenhaga. Specification and towers in shift spaces. Submitted. arXiv:1502.00931v2, 2015.

[5] Vaughn Climenhaga and Daniel J. Thompson. Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors. *Israel J. Math.*, 192(2):785–817, 2012.

[6] E. M. Coven and J. Smítal. Entropy-minimality. *Acta Math. Univ. Comenian. (N.S.)*, 62(1):117–121, 1993.

[7] Marcin Kulczycki, Dominik Kwietniak, and Piotr Oprocha. On almost specification and average shadowing properties. *Fund. Math.*, 224(3):241–278, 2014.

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

[9] Ronnie Pavlov. On intrinsic ergodicity and weakenings of the specification property. *Adv. Math.*, 295:250–270, 2016.

[10] Charles-Edouard Pfister and Wayne G. Sullivan. Large deviations estimates for dynamical systems without the specification property. Applications to the β-shifts. *Nonlinearity*, 18(1):237–261, 2005.

[11] Charles-Edouard Pfister and Wayne G. Sullivan. On the topological entropy of saturated sets. *Ergodic Theory Dynam. Systems*, 27(3):929–956, 2007.

[12] Anthony N. Quas and Paul B. Trow. Subshifts of multi-dimensional shifts of finite type. *Ergodic Theory Dynam. Systems*, 20(3):859–874, 2000.

[13] Daniel J. Thompson. Irregular sets, the β-transformation and the almost specification property. *Trans. Amer. Math. Soc.*, 364(10):5395–5414, 2012.

[14] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.

[15] B. Weiss. Intrinsically ergodic systems. *Bull. Amer. Math. Soc.*, 76:1266–1269, 1970.

Vaughn Climenhaga, Department of Mathematics, University of Houston, 4800 Calhoun St., Houston, TX 77204

E-mail address: `climenha@math.uh.edu`

URL: `http://www.math.uh.edu/~climenha/`

Ronnie Pavlov, Department of Mathematics, University of Denver, 2280 S. Vine St., Denver, CO 80208

E-mail address: `rpavlov@du.edu`

URL: `www.math.du.edu/~rpavlov/`