Abstract. We show that shift spaces with a non-uniform specification property admit a tower with exponential tails for the unique equilibrium state. In particular, we obtain various strong statistical properties including the Bernoulli property, exponential decay of correlations, the central limit theorem, and analyticity of pressure; these are the first such results using any version of the specification property.

1. Introduction

Consider a continuous map \( f: X \to X \) on a compact metric space. An equilibrium state for a potential function \( \varphi: X \to \mathbb{R} \) is an \( f \)-invariant measure maximising the quantity \( h_\mu(f) + \int \varphi \, d\mu \).

Existence, uniqueness, and statistical properties of equilibrium states have consequences for many areas of dynamics and geometry, including SRB (physical) measures for smooth maps [Rue76]; entropy rigidity for geodesic flow [Kat82]; large deviations [Kif90]; distribution of geodesics in negative [Bow72] and non-positive curvature [Kni98]; multifractal analysis [BSS02]; the Weil–Petersson metric [McM08]; Teichmüller flow [BG11]; representation theory [BCLS13]; phase transitions and quasicrystals [BL13]; and diffusion along periodic surfaces [AHS14].

In this paper we consider the case where \( X \) is a shift space and \( f \) is the shift map \( \sigma \). The most complete results are known when \( X \) is a mixing subshift of finite type: in this case every Hölder continuous potential has a unique equilibrium state, and this measure has strong statistical properties (Bernoulli property, exponential decay of correlations, central limit theorem) [Bow75]; moreover, the topological pressure function is analytic [PP90].

A weaker criterion for uniqueness is given by the specification condition [Bow74], which amounts to asking that \( X \) be ‘uniformly mixing’. Uniqueness results using non-uniform specification conditions have been proved by the author and D.J. Thompson [CT12] [CT13], but so far nothing has been known about the stronger statistical properties using any version of specification. It is known that these stronger statistical properties hold for systems on which a certain ‘tower’ can be built [You98] [You99]. Our main result is the following theorem, which uses a non-uniform specification condition to establish uniqueness and statistical properties by building a tower.

\begin{flushright}
Date: February 4, 2015.

The author is partially supported by NSF grant DMS-1362838.
\end{flushright}
Theorem 1.1. Let $X$ be a shift space (either one- or two-sided) on a finite alphabet and let $\varphi : X \to \mathbb{R}$ be Hölder continuous. If the non-uniform specification conditions [I], [II], and [III] defined in §2.5 are satisfied for some $C^p, G, C^s \subset L(X)$, then the following are true.

1. $(X, \varphi)$ has a unique equilibrium state $\mu$.
2. $\mu$ has the Gibbs property (2.5) with respect to $G$.
3. $\mu$ is the limiting distribution of $\varphi$-weighted periodic orbits.
4. There is $d \in \mathbb{N}$ such that $(X, \sigma^d, \mu)$ has the Bernoulli property and has exponential decay of correlations for Hölder observables.
5. $(X, \sigma, \mu)$ satisfies the central limit theorem for Hölder observables $\psi$, with variance $0$ if and only if $\psi$ is cohomologous to a constant.
6. Given any Hölder continuous $\psi : X \to \mathbb{R}$, there is $\varepsilon > 0$ such that the pressure function $t \mapsto P(\varphi + t\psi)$ is real analytic on $(-\varepsilon, \varepsilon)$.

Remark 1.2. The conditions of Theorem 1.1 are satisfied when $(X, \sigma)$ has the classical specification property; we emphasise that conclusions (4)–(6) are new even in this setting.

We give a rough description here of the non-uniform specification conditions [I] [II] and [III] deferring precise definitions to §2. These conditions are mild variants of criteria introduced in [CT12, CT13], and are given in terms of the language $L$ of the shift $X$, which contains all finite words appearing in some element of $X$. The classical specification property asks for some $\tau \in \mathbb{N}$ such that any two words $v, w \in L$ can be joined by a connecting word $u$ with length $\leq \tau$, where ‘joined’ means that $vuw \in L$. Condition [I] weakens this by restricting to some collection $G \subset L$ of ‘good’ words.

Condition [II] asks that $G$ be ‘thermodynamically large’, in the sense that (nearly) every word in $L$ can be written as $w^p w u^s$, where $w \in G$ and $u^p, u^s$ come from subsets $C^p, C^s \subset L$ that have small topological pressure (relative to the whole system). Finally, [III] is a mildly technical requirement that intersections and unions of good words are themselves good.

The proof of Theorem 1.1 is carried out via two intermediate theorems, which we summarise here (precise statements are in §3.1).

- **Theorem 3.1**: If we strengthen [I] to require that $\tau = 0$, then we can build a tower and use the machinery of countable state topological Markov chains [Sar99, Sar01] together with Young’s results from [You99] to establish properties (4)–(6).

- **Theorem 3.2**: If [I] [III] hold for $C^p, G, C^s$, then there are $E^p, F, E^s$ satisfying the conditions of Theorem 3.1. In particular, $F$ is ‘freely concatenable’ and $E^p, E^s$ are ‘thermodynamically small’.

---

1. Ruelle [Rue92] proved a Perron–Frobenius theorem using specification, but did not establish spectral gap or a rate of convergence.

2. Without the upper bound $\tau$ this is just the definition of topological transitivity.

3. Conclusions (1)–(3) can be established using [CT13]. Uniqueness is enough to show that the pressure function is $C^1$, but analyticity requires more.
The collection $\mathcal{F}$ is produced by adapting an argument of Bertrand for existence of a synchronising word in a shift with specification [Ber88]. The bulk of the proof of Theorem 3.2 is devoted to using [II] and [III] to show that $\mathcal{F}$ can be chosen to satisfy the ‘largeness’ condition [II] which is necessary to guarantee that equilibrium states lift to the tower and that the tower has exponential tails.

Remark 1.3. The previous paragraph illustrates a general theme: even when it is relatively clear how to build a tower, it is usually a non-trivial problem to verify that equilibrium states lift to the tower and that the tower’s tails decay exponentially [Kel89, PZ07, PSZ08, PS08, BT09, IT10]. One goal of the present approach is to give a set of more readily verifiable conditions that can establish liftability and exponential tails.

Remark 1.4. The non-uniform specification properties from [CTT12, CT13] can be extended to the non-symbolic setting, where along with a non-uniform expansivity property they once again yield uniqueness [CT14, CT15]; this has applications to smooth systems [CFT15]. Although the results given here are for symbolic systems, they should admit a similar generalisation (see §7.1 for a discussion).

In §2 we give the definitions and background used in the hypotheses and conclusion of Theorem 1.1. Several subsidiary results and applications are given in §3. We remark that among these subsidiary results are mild strengthenings of the Birkhoff ergodic theorem (Theorem 3.17) and the Shannon–McMillan–Breiman theorem (Theorem 3.18) that hold quite generally, not just in the setting of this paper. The proof of Theorem 1.1 is given in §§4–6. In §7 we give some further remarks on the main results, potential applications, and the relationship to previous work.

2. Definitions

2.1. Shift spaces. We start by recalling basic definitions from symbolic dynamics; see [LM95] for further details. Let $A$ be a finite set (the alphabet) and let $\sigma : A^\mathbb{Z} \to A^\mathbb{Z}$ be defined by $\sigma(x)_k = x_{k+1}$. Define $\sigma : A^{\mathbb{N}_0} \to A^{\mathbb{N}_0}$ similarly. Equip $A^\mathbb{Z}$ and $A^{\mathbb{N}_0}$ with the metric $d(x, y) = 2^{-n(x, y)}$, where $n(x, y) = \min\{|k| \mid x_k \neq y_k\}$.

Let $X$ be a shift space on $A$; that is, a closed $\sigma$-invariant subset of $A^\mathbb{Z}$ or $A^{\mathbb{N}_0}$. A word is a finite sequence of symbols from $A$ (we allow the empty word, which has no symbols), and the language of $X$, which we denote $\mathcal{L}$, is the set of words that appear in some element of $X$, so $\mathcal{L} = \bigcup_{n\geq 0} \mathcal{L}_n$, where

$$\mathcal{L}_n = \mathcal{L}_n(X) = \{w \in A^n \mid \text{there is } x \in X \text{ such that } x_1 \cdots x_n = w\}.$$  

We write $A^* = \bigcup_{n\geq 0} A^n$ for the collection of all finite concatenations of elements of $A$, so $\mathcal{L}(X) \subset A^*$, with equality if and only if $X$ is the full shift.

---

4All of our results apply to both one- and two-sided shift spaces.
When we need to work with an indexed collection of words, we will write the indices as superscripts; thus \( w^1, w^2 \) represent two different words, while \( w_1, w_2 \) represent the first and second symbols in the word \( w \).

Juxtaposition denotes concatenation and will be used liberally throughout the paper both for words and for collections of words; for example, given \( w \in A^* \), we will have occasion to refer to the following sets, or similar ones:

\[
w \mathcal{L} \cap \mathcal{L} = \text{the set of all words in } \mathcal{L} \text{ that begin with } w,
\]

\[
\mathcal{L} \setminus \mathcal{L}w\mathcal{L} = \text{the set of all words in } \mathcal{L} \text{ that do not contain } w \text{ as a subword}.
\]

Given a word \( w \in A^* \), we write \( |w| \) for the length of \( w \); given \( 1 \leq i \leq j \leq |w| \), we write \( w_{[i,j]} = w_i \cdots w_j \). When convenient, we use the notation \( w_{[i+1,j]} = w_{[i+1,j]} \), and similarly for \( w_{(i,j)} \) and \( w_{(i,j)} \). We will use the same notation for subwords of an infinite sequence \( x \in X \).

Given a collection of words \( D \subset \mathcal{L} \), we write

\[
D_n = D \cap \mathcal{L}_n = \{ w \in D \mid |w| = n \},
\]

\[
D \leq n = \{ w \in D \mid |w| \leq n \}.
\]

We recall several important classes of shift spaces.

- \( X \) has **specification** \(^5\) if there is \( \tau \in \mathbb{N} \) such that for all \( v, w \in \mathcal{L} \) there is \( u \in \mathcal{L} \) with \( |u| \leq \tau \) such that \( vuw \in \mathcal{L} \).
- \( X \) is **synchronised** if there is a word \( z \in \mathcal{L} \) such that if \( uz \in \mathcal{L} \) and \( zv \in \mathcal{L} \), then \( uzv \in \mathcal{L} \).
- \( X \) is **coded** if there is a collection of **generators** \( G \subset \mathcal{L} \) such that \( w \in \mathcal{L} \) if and only if \( uwv \in G^* \) for some \( u, v \in A^* \). Equivalently, \( X \) can be represented as the closure of a uniformly continuous image of a countable-state irreducible topological Markov chain \([FF92]\).

It is well-known that for transitive shift spaces,

\[
\text{specification} \Rightarrow \text{synchronised} \Rightarrow \text{coded}.
\]

The first implication was shown in \([Bers88]\), and inspires an important part of the proof of our main result, given below as Proposition 3.7. To prove that synchronised shifts are coded, it suffices to take \( G = \{ zw \mid zwz \in \mathcal{L} \} \), where \( z \) is a synchronising word. We will follow a similar strategy to construct a collection of words that can be freely concatenated.

**Remark 2.1.** There is a close connection between coded systems and tower constructions, which can be most easily seen using the representation in terms of countable-state topological Markov chains. We discuss this in \( \S 7.4 \).

---

\(^5\)There are many non-equivalent versions of the specification property in the literature. This definition is specialised for the symbolic setting and is slightly weaker than the most common one, which requires \( |u| = \tau \).
2.2. Thermodynamic formalism. We recall some basic definitions from thermodynamic formalism and describe the notation that we will use. See [Wal82, CT12, CT13] for further details.

Let $X$ be a shift space on a finite alphabet and $\varphi : X \to \mathbb{R}$ a continuous function, called a potential. Given $\varphi$, we define a function $\hat{\varphi} : L \to \mathbb{R}$ by

$$\hat{\varphi}(w) = \sup_{x \in [w]} S_{|w|\varphi}(x),$$

where $[w] = \{x \in X \mid x_{[0,|w|]} = w\}$ is the cylinder defined by $w$, and $S_{n}\varphi(x) = \sum_{j=0}^{n-1} \varphi(\sigma^{j}x)$ is the $n$th Birkhoff sum. Given a collection of words $D \subset L$, we write

$$\Lambda_n(D, \varphi) = \sum_{w \in D_n} e^{\hat{\varphi}(w)},$$

$$P(D, \varphi) = \lim_{n \to \infty} \frac{1}{n} \log \Lambda_n(D, \varphi).$$

The first quantity is the partition sum of $\varphi$ on $D$, and the second is the pressure of $\varphi$ on $D$. When $D = L$ this gives the standard definition of topological pressure, and we write $P(\varphi) = P(L, \varphi)$. When $\varphi = 0$ we write $h(D) = P(D, 0)$ for the entropy of $D$.

We will frequently use the following consequence of (2.2):

$$P(C \cup D, \varphi) = \max\{P(C, \varphi), P(D, \varphi)\} \text{ for every } C, D \subset L.$$

The variational principle [Wal82, Theorem 9.10] states that $P(\varphi)$ is the supremum of the quantities $h(\mu) + \int \varphi \, d\mu$ taken over all $\sigma$-invariant Borel probability measures, where $h(\mu)$ is the measure-theoretic entropy. The supremum is achieved at equilibrium states. Because $\sigma$ is expansive, equilibrium states exist for all continuous $\varphi$, and so the real questions here are uniqueness and statistical properties. These require some regularity condition on $\varphi$. Given $\beta > 0$, we consider the following set of Hölder continuous functions:

$$C_{\beta}(X) = \{\varphi : X \to \mathbb{R} \mid \text{there exists } |\varphi|_{\beta} \in \mathbb{R} \text{ such that } |\varphi(x) - \varphi(y)| \leq |\varphi|_{\beta} e^{-\beta n} \text{ whenever } x_k = y_k \text{ for all } |k| \leq n\}.$$

We write $C^{h}(X) = \bigcup_{\beta > 0} C_{\beta}(X)$ for the set of all Hölder functions.

Uniqueness of the equilibrium state was shown in [Bow71] for the case when $\varphi \in C^{h}(X)$ and $(X, \sigma)$ has specification. Uniqueness in the case of non-uniform specification (defined below) was shown in [CT13]. The purpose of this paper is to establish statistical properties in these settings.

2.3. Non-uniform specification. We are interested in shift spaces where the specification property does not necessarily hold on the entire shift, but does hold on a suitable collection of subwords. That is, we are interested in shifts for which there is a collection of “good” words $G \subset L$ with the following specification property.

[I] There is $\tau \in \mathbb{N}$ such that for all $v, w \in G$ we have $vL_{\leq \tau} w \cap G \neq \emptyset$. 


This condition means that for every \( v, w \in G \) there is \( u \in L \) with \( |u| \leq \tau \) and \( uvw \in G \). An important special case is when \( \tau = 0 \) and so words in \( G \) can be freely concatenated. We state this as its own condition.

[**] Given any \( v, w \in G \) we have \( vw \in G \).

In Theorem 3.2 we show that [I] can be strengthened to [I₀] while retaining the properties listed below. This is the key to carrying out a tower construction (Theorem 3.1).

Now we fix a Hölder continuous\(^6\) potential function \( \varphi : X \to \mathbb{R} \) and give a condition guaranteeing that \( G \) is ‘large enough’ for thermodynamic purposes. Given \( C^p, C^s \subset L \), which we refer to as the prefix collection and suffix collection, we write

\[
C^pGC^s = \{uvw \mid u \in C^p, v \in G, w \in C^s\},
\]

\[
\mathcal{C} := C^p \cup C^s \cup (L \setminus C^pGC^s).
\]

Note that \( \mathcal{C} \) depends not only on \( C^p, C^s \) but also on \( G \). We think of \( \mathcal{C} \) as the collection of ‘obstructions to specification’, and will require it to satisfy the following condition.

[**] \( P(\mathcal{C}, \varphi) < P(\varphi) \).

As formulated, this is a condition on the triple \( (C^p, G, C^s) \). We will sometimes use the phrase “\( G \) satisfies [II]”, which means that there are \( C^p, C^s \subset L \) such that the triple \( (C^p, G, C^s) \) satisfies [II]

We need one more condition in order to state our main result. This condition is somewhat technical, but is reasonable to expect from potential applications; see §7.1 for a discussion.

[**] If \( u, v, w \in L \) are such that \( uvw \in L \) and \( uv, vw \in G \), then \( uvw \in G \) and \( v \in G \).

This condition is illustrated in Figure 2.1. Conditions [I] [II] and [III] are all we need for the statement of the main result (Theorem 1.1). A weaker version of [III] is described in §3.1 and used in Theorems 3.1 and 3.2. Some similar conditions appearing in other related work are described in §3.2.

2.4. **Statistical properties.** We recall several good statistical properties that a measure can have; Theorem 1.1 says that these are all satisfied for the unique equilibrium state produced by [I] [II] [III].

---

\(^6\)The definitions can be made for any continuous potential, but we need a regularity condition for the results. The uniqueness result in [CT13] requires \( \varphi \) to have the Bowen property on \( G \), which is weaker than Hölder continuity. Something in between these conditions would be needed for the present results; roughly speaking, \( |\varphi(x) - \varphi(y)| \) should decay exponentially in the ‘number of good times’ before \( x, y \) disagree. For simplicity of exposition we only consider the case where \( \varphi \) is Hölder.
2.4.1. Gibbs property. Given $G \subset L$ and $\varphi: X \to \mathbb{R}$, we say that $\mu$ has the Gibbs property for $\varphi$ with respect to $G$ if there is $Q_1 > 0$ such that

\begin{equation}
\text{for every } w \in L, \text{ we have } \mu[w] \leq Q_1 e^{-|w| P(\varphi) + \hat{\varphi}(w)}; \text{ and}
\end{equation}

\begin{equation}
\text{for every } w \in G, \text{ we have } \mu[w] \geq Q_1^{-1} e^{-|w| P(\varphi) + \hat{\varphi}(w)}.
\end{equation}

Note that the lower bound is only required to hold on $G$, while the upper bound holds on all of $L$. This weakened version of the classical Gibbs property was introduced in [CT12, CT13].

2.4.2. Periodic orbits. For $n \in \mathbb{N}$, let $\text{Per}_n = \{ x \in X | \sigma^n x = x \}$ be the set of $n$-periodic points. Note that $\text{Per}_n$ is finite. Let

\begin{equation}
\mu_n = \frac{1}{\sum_{k=1}^{n} \sum_{x \in \text{Per}_k} e^{S_k \varphi(x)}} \sum_{k=1}^{n} \sum_{x \in \text{Per}_k} e^{S_k \varphi(x)} \delta_x
\end{equation}

be the $\varphi$-weighted periodic orbit measure corresponding to periodic orbits of length at most $n$. Say that $\mu$ is the limiting distribution of $\varphi$-weighted periodic orbits if $\mu_n$ converges to $\mu$ in the weak* topology.

2.4.3. Bernoulli property. Given a state space $S$ and a probability vector $p = (p_a)_{a \in S}$, the Bernoulli scheme with probability vector $p$ is $(S^\mathbb{Z}, \sigma, \mu_p)$, where $\sigma$ is the left shift map and $\mu_p[w] = \prod_{i=1}^{|w|} p_{a_i}$ for every $w \in S^\mathbb{Z}$. The measure $\mu$ is said to have the Bernoulli property if $(X, \sigma, \mu)$ is measure-theoretically isomorphic to a Bernoulli scheme.

2.4.4. Decay of correlations. Let $(X, \sigma, \mu)$ be a shift space with an invariant measure $\mu$. Given $\psi_1, \psi_2: X \to \mathbb{R}$, consider the correlation functions

$$\text{Cor}_n^\mu(\psi_1, \psi_2) = \int (\psi_1 \circ \sigma^n) \psi_2 \, d\mu - \int \psi_1 \, d\mu \int \psi_2 \, d\mu.$$ 

We say that the system has exponential decay of correlations for observables in $C_\beta(X)$ if there is $\theta \in (0, 1)$ such that for every $\psi_1, \psi_2 \in C_\beta(X)$ there is $K(\psi_1, \psi_2) > 0$ such that

\begin{equation}
|\text{Cor}_n^\mu(\psi_1, \psi_2)| \leq K(\psi_1, \psi_2) \theta^n \text{ for every } n \in \mathbb{Z}.
\end{equation}

\text{We will use } Q_1, Q_2, \ldots \text{ to denote 'global constants' that are referred to throughout the paper. We will use } K \text{ or } C \text{ for 'local constants' that appear only within the proof of a given lemma or proposition, and are not used for more than one or two paragraphs.}
2.4.5. **Central limit theorem.** Given \((X, \sigma, \mu)\) as above and \(\psi : X \to \mathbb{R}\), we say that the **central limit theorem** holds for \(\psi\) if 

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

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

Say that \(\psi\) is **cohomologous to a constant** if there are \(u : X \to \mathbb{R}\) and \(c \in \mathbb{R}\) such that 

\[
\psi(x) = u(x) - u(\sigma x) + c \quad \text{for } \mu\text{-a.e. } x \in X.
\]

One generally expects that in the central limit theorem the variance \(\sigma^2 \psi\) is 0 if and only if \(\psi\) is cohomologous to a constant. This will hold in our setting as well.

### 3. Results

3.1. **Subsidiary results.** As described above, we will prove Theorem 1.1 by producing \(F \subset \mathcal{L}\) satisfying \([I_0]\) and \([II]\). We will also require \(F\) to satisfy another property that is related to \([III]\). First we note that \([III]\) can be reformulated as follows (see Figure 3.1).

**[III]** If \(x \in X\) and \(i \leq j < k \leq \ell \in \mathbb{Z}\) are such that \(x_{[i,k)}, x_{[j,\ell)} \in \mathcal{G}\), then 

\(x_{[i,\ell)}, x_{[j,k)} \in \mathcal{G}\).

We will produce \(F \subset \mathcal{L}\) with the following similar (but weaker) property.

**[III]** If \(x \in X\) and \(i \leq j < k \leq \ell \in \mathbb{Z}\) are such that \(x_{[i,k)}, x_{[j,\ell)} \in \mathcal{F}\), and there are \(a < j\) and \(b > \ell\) such that \(x_{[a,j)}, x_{[\ell,b)} \in \mathcal{F}\), then \(x_{[j,k)} \in \mathcal{F}\).

Both conditions are illustrated in Figure 3.1. Note that although \(i, j, k, \ell\) must appear in the order shown, \(a\) can be to either side of \(i\) (or equal to it), so long as \(a < j\), and similarly there is no constraint on \(b\) and \(\ell\).

![Figure 3.1. Conditions [III] and [III*]](image)

Theorem 1.1 is a consequence of the following two results.

**Theorem 3.1.** Let \(X\) be a shift space (one- or two-sided) on a finite alphabet, and \(\varphi : X \to \mathbb{R}\) a Hölder continuous potential function. Suppose there are collections \(\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s \subset \mathcal{L}(X)\) satisfying \([I_0]\), \([II]\), and \([III]\). Then the following are true.

1. \((X, \varphi)\) has a unique equilibrium state \(\mu\).
2. \(\mu\) has the Gibbs property \((2.5)\) with respect to \(\mathcal{F}\).
3. \(\mu\) is the limiting distribution of \(\varphi\)-weighted periodic orbits.
(4) With \( d = \gcd\{|w| \mid w \in \mathcal{F}\} \), \((X, \sigma^d, \mu)\) has the Bernoulli property and has exponential decay of correlations for Hölder observables.

(5) \((X, \sigma, \mu)\) satisfies the central limit theorem for Hölder observables \(\psi\), with variance 0 if and only if \(\psi\) is cohomologous to a constant.

(6) Given any Hölder continuous \(\psi: X \to \mathbb{R}\), there is \(\varepsilon > 0\) such that the map \(t \mapsto \mathbb{P} (\phi + t\psi)\) is real analytic on \((-\varepsilon, \varepsilon)\).

**Theorem 3.2.** Let \(X\) be a shift space (one- or two-sided) on a finite alphabet and let \(\varphi: X \to \mathbb{R}\) be Hölder continuous. Suppose there are \(C^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)\) satisfying [I], [II], and [III]. Then there are \(\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s \subset \mathcal{L}\) satisfying [I0], [II], and [III∗], and a measure \(\mu\) has the Gibbs property for \(\mathcal{G}\) if and only if it has the Gibbs property for \(\mathcal{F}\).

**Remark 3.3.** Although Theorem 3.2 guarantees that \(P(\mathcal{E}, \varphi) < P(\varphi)\) (condition [II]), we may have \(P(\mathcal{E}, \varphi) > P(\mathcal{C}, \varphi)\). This happens already for SFTs; in §7.2 we show that for every \(\delta > 0\) there is a mixing SFT \(X\) with \(h(X) = \log 2\) and with the property that if \(\mathcal{F} \subset \mathcal{L}(X)\) satisfies [I0], then for every choice of \(\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}\) we have \(h(\mathcal{E}) \geq h(X) - \delta\). This is despite the fact that \(\mathcal{L}\) itself has specification and so we can take \(\mathcal{G} = \mathcal{L}\) and \(\mathcal{C} = \emptyset\) in Theorem 3.2.

Theorem 3.1 is proved in [5] by assembling various well-known pieces of machinery (in particular, Sarig’s work on countable state topological Markov chains and Young’s work on towers). Broadly speaking, the plan is as follows.

(1) Embed a countable state topological Markov chain \(\Sigma\) into \(X\) via a shift-commuting map \(\pi: \Sigma \to X\). A state of \(\Sigma\) will correspond to a pair \((w, i)\), where \(w\) is an ‘irreducible’ word from \(\mathcal{F}\) and \(i \in [1, |w|]\) is an integer marking how many symbols of \(w\) we have seen.

(2) Pull back the potential \(\varphi\) to \(\Phi: \Sigma \to \mathbb{R}\) and use [II] to show that \(\Phi\) is strongly positive recurrent [Sar01].

(3) Use [III∗] to show that \(\pi\) is 1-1, and [II] to show that every equilibrium state of \((X, \varphi)\) gives full weight to \(\pi(X)\); this step in particular requires careful work and does not follow from pre-existing theories.

(4) Deduce existence of a unique equilibrium state from [Sar99], the Gibbs property and equidistribution of periodic points following the approach in [CT12, CT13], analyticity from [Sar01], the Bernoulli property from [Sar11], and other statistical properties from [You99].

**Remark 3.4.** It is an interesting open question to determine which conclusions of Theorem 3.1 can be established without Condition [III∗], particularly since in the proof of Theorem 3.2 we must exert some effort to establish this condition (see §6.3). Even without Condition [III∗], conclusions [1]–[3] of Theorem 3.1 still hold (see Theorem 7.4), but the situation with conclusions [4]–[6] is unclear.

Theorem 3.2 is proved in [6] and contains the most important new ideas in this paper. The construction of \(\mathcal{F}\) relies on the following definition, which is illustrated in Figure 3.2.
Definition 3.5. Given \( \mathcal{G} \subset \mathcal{L} \) satisfying [I], we say that \((r, c, s)\) is a synchronising triple for \( \mathcal{G} \) if \( r, s \in \mathcal{G} \), \( c \in \mathcal{L}_{\leq \tau} \), and given any \( r' \in \mathcal{L} \cap \mathcal{G} \) and \( s' \in s \mathcal{L} \cap \mathcal{G} \), we have \( r'c s' \in \mathcal{G} \). If \((r, c, s)\) is a synchronising triple for \( \mathcal{G} \), write \( \mathcal{B}^{r,s} = \mathcal{L} \cap s \mathcal{L} \cap \mathcal{G} \), and \( \mathcal{F}^{r,c,s} = c \mathcal{B}^{r,s} = \{cw \mid w \in \mathcal{B}^{r,s}\} \).

![Figure 3.2](image)

Figure 3.2. A synchronising triple \((r, c, s)\) and the collections \( \mathcal{B}^{r,s}, \mathcal{F}^{r,c,s} \) it generates.

Remark 3.6. We think of \( c \) as the ‘connecting’ word and \( \mathcal{B}^{r,s} \) as the set of ‘bridges’ between occurrences of \( c \). Note that every \( w \in \mathcal{B}^{r,s} \) must begin with \( s \) and end with \( r \), but the subwords \( s \) and \( r \) may overlap; we are allowed to have \( |w| < |s| + |r| \).

Proposition 3.7. Every \( \mathcal{G} \subset \mathcal{L} \) satisfying [I] has a synchronising triple. Moreover, if \((r, c, s)\) is any synchronising triple for \( \mathcal{G} \), then \( \mathcal{F}^{r,c,s} \) satisfies [I0]. Finally, a \( \sigma \)-invariant measure \( \mu \) is Gibbs for \( \varphi \) with respect to \( \mathcal{G} \) if and only if it is Gibbs with respect to \( \mathcal{F}^{r,c,s} \).

In [6], we prove Proposition 3.7 and show that if \( \mathcal{C}^{p}, \mathcal{G}, \mathcal{C}^{s} \) also satisfy [II] and [III] then there is a synchronising triple \((r, c, s)\) for \( \mathcal{G} \) such that the corresponding \( \mathcal{F}^{r,c,s} \) satisfies [II] (for some \( \mathcal{E}^{p}, \mathcal{E}^{s} \)) and [III].

Remark 3.8. Every word in \( \mathcal{F} = \mathcal{F}^{r,c,s} \) can be extended to a word in \( \mathcal{G} \) in a uniform number of symbols, and vice versa. On the one hand, given any \( w \in \mathcal{F} \), we have \( w = cv \) for some \( v \in s \mathcal{L} \cap \mathcal{G} \), hence \( rw = rcw \in \mathcal{G} \). On the other hand, if \( w \in \mathcal{G} \) then there are \( u, v \in \mathcal{L}_{\leq \tau} \) such that \( suwvr \in \mathcal{B}^{r,s} = s \mathcal{L} \cap \mathcal{L} \cap \mathcal{G} \), and hence \( csuwvr \in \mathcal{F} \).

3.2. Connection to previous results. Condition [I] is a version of a property first introduced in [CT12, CT13]; the following strengthening of [I] implies the condition used there.

[I] There is \( \tau \in \mathbb{N} \) such that for all \( v, w \in \mathcal{G} \) we have \( v \mathcal{L}_{\tau} w \cap \mathcal{G} \neq \emptyset \).

Similarly, [II] is a version of the following property from [CT13], which appeared in [CT12] in the case \( \varphi = 0 \).

[II] There are \( \mathcal{C}^{p}, \mathcal{C}^{s} \subset \mathcal{L} \) such that \( \mathcal{L} \subset \mathcal{C}^{p} \mathcal{G} \mathcal{C}^{s} \), and \( P(\mathcal{C}^{p} \cup \mathcal{C}^{s}, \varphi) < P(\varphi) \).

In [CT12], these two conditions were used to prove existence of a unique measure of maximal entropy, subject to a condition on the collections

\[
\mathcal{G}^{M} = \mathcal{L} \cap (\mathcal{C}_{\leq M}^{p} \mathcal{G} \mathcal{C}_{\leq M}^{s})
= \{uww \in \mathcal{L} \mid u \in \mathcal{C}^{p}, v \in \mathcal{G}, w \in \mathcal{C}^{s}, |u| \leq M, |w| \leq M\},
\]
given by the following ‘extendability’ requirement.

\[ E \] For every \( M \geq 0 \) there is \( T = T(M) \in \mathbb{N} \) such that every \( w \in \mathcal{G}^M \) has \( (L \leq T)w(L \leq T) \cap \mathcal{G} \neq \emptyset \).

In other words, every word in \( \mathcal{G}^M \) can be extended to a word in \( \mathcal{G} \) by appending no more than \( T \) symbols to either end. If \( \mathcal{G} \) satisfies \([I]\) and the collections \( \mathcal{G}^M \) from (3.1) satisfy \([E]\), then the following is true.

\[ [I^M] \quad \text{Every } \mathcal{G}^M \text{ satisfies } [I] \quad \text{(the gluing time } \tau \text{ may depend on } M). \]

It was shown in [CT13, Theorem C] that if \( \varphi \) is Hölder, then \([I^M]\) and \([II']\) imply existence of a unique equilibrium state for \( \varphi \).

Remark 3.9. In fact the uniqueness result in [CT13] are stated in the case when \( \mathcal{G}^M \) satisfies \([I']\), but this can be replaced by \([I]\) in both [CT12, CT13], as shown in [CTY13, Appendix A]; the proof requires only minor modifications. A more subtle point is that the results on factors from [CT12] also go through with \([I']\) replaced by \([I]\); this is discussed in the next section.

Remark 3.10. Formally there is little difference between \([II]\) and \([II']\): if \( C^p, C^s \) satisfy \([II]\), then one can put \( \hat{C}^p = C^p \cup (L \setminus C^pGC^s) \) and obtain \([II']\) for \( \hat{C}^p, \mathcal{G}, C^s \). However, this enlarges the collections \( \mathcal{G}^M \) and in particular may cause \([E]\) and \([I']\) to fail.

Despite Remark 3.10, an examination of the proof of [CT13, Theorem C] reveals that the result continues to hold if \([II']\) is replaced by \([II]\). Indeed, the only place where \([II']\) is used in the proof is in the partition sum estimates in [CT13, §5.1]; Lemma 4.1 of the present paper establishes some of these using the weaker condition \([II]\) and the others extend in a completely analogous manner. Together with Remark 3.9 this gives the following result.

**Theorem 3.11.** Let \( X \) be a shift space on a finite alphabet and \( \varphi : X \to \mathbb{R} \) a Hölder continuous potential. Suppose \( C^p, \mathcal{G}, C^s \subseteq L(X) \) are such that

(1) \( \mathcal{G}^M \) satisfies \([I]\) for every \( M \in \mathbb{N} \);

(2) \( \hat{C} := C^p \cup C^s \cup (L \setminus C^pGC^s) \) satisfies \( P(C, \varphi) < P(\varphi) \).

Then \( (X, \varphi) \) has a unique equilibrium state \( \mu \), which satisfies the Gibbs property (2.5) with respect to every \( \mathcal{G}^M \). If \( \mathcal{G} \) satisfies \([L_0]\) then \( \mu \) is the limiting distribution of \( \varphi \)-weighted periodic orbits.

### 3.3. Factors and examples.

The original motivation for the non-uniform specification property introduced in [CT12] was to show that every subshift factor of a \( \beta \)-shift has a unique measure of maximal entropy, since for \( \varphi = 0 \), conditions \([I']\) \([II']\) and \([E]\) are well-behaved under passing to factors. We will find that the same is true is \([I]\) \([II]\) \([III]\) and the following condition on \( \mathcal{G}, L. \)

---

8The same result holds with Hölder continuity replaced by the weaker Bowen property on \( \mathcal{G} \) [CT13, Definition 2.2].

9The constant \( Q_1 \) in (2.5) is allowed to depend on \( M \).

10This last assertion holds because \([L_0]\) implies the (Per)-specification condition from [CT13, Definition 2.1].
For every \( w \in \mathcal{L} \) there are \( u, v \in \mathcal{L} \) such that \( uwv \in \mathcal{G} \).

Condition \([E^*]\) is similar in form to \([E]\) but has no uniformity assumption.

Following the idea in \([CT14]\), we define the entropy of obstructions to specification as

\[
(3.2) \quad h_{\text{spec}}^\bot(X) = \inf \{ h(C) | C^p, \mathcal{G}, C^s \subset \mathcal{L}(X) \text{ satisfy } [I], [III], \text{ and } [E^*] \}.
\]

Then we have the following result, which strengthens \([CT12, Corollary 2.3 \text{ and Theorem D}]\).

**Theorem 3.12.** Let \((X, \sigma)\) be a shift space on a finite alphabet.

1. Let \((\tilde{X}, \tilde{\sigma})\) be a subshift factor of \((X, \sigma)\) such that \(h_{\text{spec}}^\bot(X) < h(\tilde{X})\).

Then \(\tilde{X}\) has a unique measure of maximal entropy, which is the limiting distribution of periodic orbits, has the Bernoulli property and exponential decay of correlations for some iterate of \(\tilde{\sigma}\), and satisfies the central limit theorem. In this case for any H"older \(\psi: Y \to \mathbb{R}\) there is \(\varepsilon > 0\) such that \(t \mapsto P(t\psi)\) is real analytic on \((-\varepsilon, \varepsilon)\).

2. If \(h_{\text{spec}}^\bot(X) = 0\), then the conclusion of the previous part applies to every subshift factor of \((X, \sigma)\).

**Proof.** This is a consequence of Theorem 1.1 and the following two propositions, which are proved in \([CT12, Section 7.3]\).

**Proposition 3.13.** Let \((\tilde{X}, \tilde{\sigma})\) be a shift factor of \((X, \sigma)\), and denote the corresponding languages by \(\tilde{\mathcal{L}}\) and \(\mathcal{L}\). Given \(C^p, \mathcal{G}, C^s \subset \mathcal{L}\) such that:

1. if \(\mathcal{G}\) satisfies \([I]\), then \(\tilde{\mathcal{G}}\) satisfies \([I]\);
2. if \(\mathcal{G}\) satisfies \([III]\), then \(\tilde{\mathcal{G}}\) satisfies \([III]\);
3. \(h(\tilde{C}) \leq h(C)\);
4. if \(\mathcal{L}\) satisfies \([E^*]\), then so does \(\tilde{\mathcal{L}}\).

**Proposition 3.14.** Let \((X, \sigma)\) be a shift space with \(\mathcal{G} \subset \mathcal{L}(X)\) satisfying \([I]\) and \([E^*]\). Then either \(h(X) > 0\) or \(X\) is a single periodic orbit.

Here we remark that the proof of Proposition 3.13 is essentially a reiteration of the proof of \([CT12, Proposition 2.2]\), with the difference that \([I]\) replaces \([I']\) in (1); \([III]\) in (2) is new; \(\mathcal{C}\) replaces \(C^p \cup C^s\) in (3); \([E^*]\) replaces \([E]\) in (4). A more subtle issue arises in Proposition 3.14 whose proof requires some new ideas compared to the corresponding result for \([I]\) \([CT12, Proposition 2.4]\); in particular, it benefits from the introduction of synchronising triples and from Lemma 6.7. Full details are in \([CT12, Section 7.3]\). \(\square\)

**Remark 3.15.** Although the most obvious way to get \(h_{\text{spec}}^\bot(X) = 0\) is to have \(h(C) = 0\) for some \(C^p, \mathcal{G}, C^s\), we expect that there are examples where \(h_{\text{spec}}^\bot(X) = 0\) but the infimum is not achieved; a natural class of candidate examples is given by shift spaces coding transitive piecewise monotonic

\[\text{This differs slightly from } [CT14] \text{ in that we use } [I], [III], \text{ and } [E^*] \text{ instead of } [I']\]
transformations of the interval, whose structure has been described by Hofbauer \cite{Hof79, Hof81}.

The motivating classes of examples in \cite{CT12} were $\beta$-shifts and $S$-gap shifts (see \cite{CT12} for definitions and properties). Theorem 3.12 applies to both these classes of shifts, strengthening the results of \cite{CT12}.

Both $\beta$-shifts and $S$-gap shifts admit natural descriptions in terms of a countable graph, which allows them to be studied directly without use of the non-uniform specification techniques described here. Indeed, equilibrium states for $\beta$-shifts were well understood already by Hofbauer \cite{Hof78} and Walters \cite{Wal78}, and the same technology could be applied to the $S$-gap shifts, although this does not seem to have been done in the literature.

Thus since the work in \cite{CT12, CT13} it has been natural to ask if there are examples of systems that can be studied using non-uniform specification but do not come from a countable graph. This was one motivation for the present paper. In some sense, then, Theorem 1.1 (and in particular Theorem 3.2) can be considered as a negative result, because it tells us that every shift with the non-uniform specification properties [I]–[III] can in fact be described by a countable graph as in §5.2 (at least up to some thermodynamically negligible part), so that there are no new examples; this is discussed further in §7.4.

On the other hand, if a system is defined in some manner that does not make this Markov structure explicit, then it may be difficult to find the graph that does the job. This is particularly true in the non-symbolic setting, where the task of building a suitable tower can be quite difficult. This leads to another motivation for the present paper, namely that a non-symbolic version of Theorem 1.1 may eventually be useful in studying smooth systems, particularly since the results of \cite{CT12, CT13} have already been generalised to and applied in this setting \cite{CT14, CT15, CFT15}. See §7.1 for more.

Before moving on to the proofs, we observe that since Theorem 3.12 only deals with measures of maximal entropy, one may ask what can be said about equilibrium states for non-zero potentials on the factors ($\tilde{X}, \tilde{\sigma}$). Following \cite{IRRL12}, say that a potential $\varphi: X \to \mathbb{R}$ is \textbf{hyperbolic} if

\begin{equation}
\sup_{x \in X} \lim_{n \to \infty} \frac{1}{n} S_n \varphi(x) < P(\varphi).
\end{equation}

If $\varphi$ is hyperbolic and $h_{\text{spec}}(X) = 0$, then a simple computation shows that there are $C^p, G, C^s$ satisfying [I]–[III] for $\varphi$; in particular, we can apply Theorem 1.1 as long as $\varphi$ is Hölder.

SFTs have the property that every Hölder potential is hyperbolic. The same is true for a broad class of non-uniformly expanding interval maps \cite{LRL14}, as well as for $\beta$-shifts \cite[Proposition 3.1]{CT13} and $S$-gap shifts \cite[(5.1)]{CTY13}. The proofs of this result for $\beta$-shifts and for $S$-gap shifts rely heavily on the structure of the shift and in particular do not pass to their factors. This motivates the following question.
Question 3.16. Is there an axiomatic condition (perhaps some form of non-uniform specification) guaranteeing that every Hölder potential on a shift space $X$ is hyperbolic? Is there such a condition that is preserved under passing to factors? In particular, does every subshift factor of a $\beta$-shift or an $S$-gap shift have the property that every Hölder potential is hyperbolic?

A positive answer to this question would allow a version of Theorem 3.12 to be extended to equilibrium states for all Hölder potentials, not just measures of maximal entropy.

3.4. Ergodic theorems. In the proof of Theorem 3.1, we will need mild strengthenings of the Birkhoff and Shannon–McMillan–Breiman ergodic theorems. Since these are general results that hold beyond the setting of this paper, we state them here.

Recall that the usual version of the Birkhoff ergodic theorem [Pet89, Theorem 2.2.3] can be stated as follows: if $(X, T, \mu)$ is an ergodic measure-preserving transformation and $f: X \rightarrow \mathbb{R}$ is an $L^1$ function, then for $\mu$-a.e. $x \in X$ and every $\epsilon > 0$ there is $N = N(x, \epsilon)$ such that for all $n \geq N$ we have

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) - \int f \, d\mu \right| < \epsilon.$$  

The following theorem is proved in §4.3.

**Theorem 3.17.** If $(X, T, \mu)$ is an invertible ergodic measure-preserving transformation and $f: X \rightarrow \mathbb{R}$ is an $L^1$ function, then for $\mu$-a.e. $x \in X$ and every $\epsilon > 0$ there is $N = N(x, \epsilon)$ such that for all $n \geq N$ and all $\ell \in [0, n]$ we have

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^{-\ell} x) - \int f \, d\mu \right| < \epsilon.$$

(3.4)

Given $(X, T, \mu)$ as above and a countable (or finite) measurable partition $\alpha$ of $X$, write $\alpha(x)$ for the partition element containing $x$, and for each $i < j \in \mathbb{Z}$, write $\alpha_i^j = \bigvee_{k=i}^{i-1} T^{-k} \alpha$. Recall that

$$H_\mu(\alpha) := \sum_{A \in \alpha} -\mu(A) \log \mu(A),$$

$$h_\mu(\alpha, T) := \lim_{n \to \infty} \frac{1}{n} H_\mu(\alpha_0^n).$$

The Shannon–McMillan–Breiman theorem in its usual form [Pet89, Theorem 6.2.3] states that if $\alpha$ has $H_\mu(\alpha) < \infty$, then for $\mu$-a.e. $x \in X$ and every $\epsilon > 0$ there is $N = N(x, \epsilon)$ such that for all $n \geq N$ we have

$$\left| \frac{1}{n} \log \mu(\alpha_0^n(x)) - h_\mu(\alpha, T) \right| < \epsilon.$$  

The following theorem is proved in §4.3.
Theorem 3.18. If \((X,T,\mu)\) is an invertible ergodic measure-preserving transformation and \(\alpha\) is a countable measurable partition with \(H_\mu(\alpha) < \infty\), then for \(\mu\)-a.e. \(x \in X\) and every \(\varepsilon > 0\) there is \(N = N(x,\varepsilon)\) such that for all \(n \geq N\) and all \(\ell \in [0, n]\) we have

\[
\left| -\frac{1}{n} \log \mu(\alpha_{-\ell}^\ell(x)) - h_\mu(\alpha, T) \right| < \varepsilon.
\]

4. Preparation for the proofs

4.1. Bounded distortion from Hölder continuity. Given \(\beta > 0\) and \(\varphi \in C_\beta(X)\), we see that for every \(w \in \mathcal{L}_n\) and every \(x, y \in [w]\), (2.4) yields

\[
|\varphi(\sigma^k x) - \varphi(\sigma^k y)| \leq |\varphi| \beta e^{-\beta \min(k, n-k)} \text{ for all } 0 \leq k < n,
\]

so that in particular

\[
|S_n \varphi(x) - S_n \varphi(y)| \leq 2 |\varphi| \beta \sum_{j=0}^\infty e^{-\beta j} =: |\varphi|_d < \infty.
\]

This can be thought of as a bounded distortion condition, and we think of \(|\varphi|_d\) as the bound on distortion of \(S_n \varphi\) within an \(n\)-cylinder.

It follows from (4.1) that for every \(v, w \in \mathcal{L}\) such that \(vw \in \mathcal{L}\), we have

\[
\hat{\varphi}(v) + \hat{\varphi}(w) - |\varphi|_d \leq \hat{\varphi}(vw) \leq \hat{\varphi}(v) + \hat{\varphi}(w).
\]

(The upper bound is immediate from (2.1).)

4.2. Counting estimates. In the proofs of both Theorems 3.1 and 3.2, we will need various estimates on partition sums over \(\mathcal{L}\) and over \(\mathcal{G}\). We start with the general observation that given \(\mathcal{C}, \mathcal{D} \subset \mathcal{L}\) and \(m, n \in \mathbb{N}\), we have

\[
\Lambda_{m+n}(\mathcal{L} \cap (\mathcal{C}_m \mathcal{D}_n), \varphi) = \sum_{u \in \mathcal{C}_m} \sum_{v \in \mathcal{D}_n} e^{\hat{\varphi}(uv)} \leq \sum_{u \in \mathcal{C}_m} \sum_{v \in \mathcal{D}_n} e^{\hat{\varphi}(u)} e^{\hat{\varphi}(v)} = \Lambda_m(\mathcal{C}, \varphi) \Lambda_n(\mathcal{D}, \varphi);
\]

this will be used in several places.

The following are similar to estimates appearing in [CT12, Lemmas 5.1–5.4] and [CT13, Section 5]. The chief difference here is that we may have \(\mathcal{L} \setminus \mathcal{C} \mathcal{G} \mathcal{C}^* \neq \emptyset\), but because the pressure of this collection is controlled, we get the same results.

Lemma 4.1. Let \(X\) be a shift space on a finite alphabet and \(\varphi \in C_\beta(X)\) for some \(\beta > 0\). Let \(\mathcal{G} \subset \mathcal{L}(X)\) be such that \([I]\) and \([II]\) hold. Then there is \(Q_2 > 0\) such that for every \(n\) we have

\[
eq n P(\varphi) \leq \Lambda_n(\varphi) \leq Q_2 e^{nP(\varphi)}.
\]

Furthermore, there is \(Q_3 > 0\) and \(N \in \mathbb{N}\) such that for every sufficiently large \(n\) there is \(j \in (n-N, n]\) with

\[
\Lambda_j(\mathcal{G}, \varphi) \geq Q_3 e^{jP(\varphi)}.
\]
Proof. For the first inequality in (4.4), we observe that \( L_{kn} \subset L_n L_n \cdots L_n \) (\( k \) times), and so by iterating (4.3) we get

\[
\Lambda_{kn}(\varphi) \leq \Lambda_n(\varphi)^k,
\]

which yields

\[
\frac{1}{kn} \log \Lambda_{kn}(\varphi) \leq \frac{1}{n} \log \Lambda_n(\varphi).
\]

Sending \( k \to \infty \) gives the first half of (4.4). Next we prove

\[
\Lambda_n(G, \varphi) \leq C e^{nP(\varphi)}
\]
using [I] and (4.1), then use this together with [II] to prove the second half of (4.4).

By [I] there is a map \( \pi: G_m \times G_n \to G \) given by \( \pi(v, w) = v u w \), where \( u \in L \) depends on \( v, w \) but always satisfies \( |u| \leq \tau \) and \( u \). Iterating and abusing notation slightly gives a map \( \hat{\pi}: (G_n)^k \to G \) of the form \( \hat{\pi}(v_1, \ldots, v_k) = v_1 u_1 v_2 \cdots u_{k-1} v_k \). Truncating this image to the first \( kn \) symbols gives a map \( \hat{\pi}: (G_n)^k \to L_{nk} \). Because we delete at most \( k\tau \) symbols to go from \( \pi \) to \( \hat{\pi} \), we have \( \#\hat{\pi}^{-1}(w) \leq (#A + 1)^k \tau \). (We need \( #A + 1 \) instead of \( #A \) since the number of deleted symbols is at most \( k\tau \), rather than exactly \( k\tau \).)

Furthermore, (4.2) yields

\[
\hat{\varphi}(v_1 u_1 \cdots u_{k-1} v_k) \geq \hat{\varphi}(v_1) + \cdots + \hat{\varphi}(v_k) - k (\tau \|\varphi\| + |\varphi|_d),
\]
and since truncation deletes at most \( k\tau \) symbols, we have

\[
\hat{\varphi}(\hat{\pi}(v_1, \ldots, v_k)) \geq \left( \sum_{i=1}^{k} \hat{\varphi}(v^i) \right) - Ck.
\]

It follows that \( \Lambda_{kn}(\varphi) \geq (\#A + 1)^{-k\tau} e^{-C'k} \Lambda_n(G, \varphi)^k \), hence

\[
\frac{1}{kn} \log \Lambda_{kn}(\varphi) \geq \frac{1}{n} \log \Lambda_n(G, \varphi) - \frac{1}{n} (C' + \tau \log(\#A + 1)).
\]

Sending \( k \to \infty \) gives (4.7).

By Condition [II] there is \( \varepsilon > 0 \) and \( K > 0 \) such that

\[
\Lambda_n(C^p \cup C^s, \varphi) \leq K e^{nP(\varphi) - \varepsilon},
\]

(4.9)

\[
\Lambda_n(L \setminus C^p G C^s, \varphi) \leq K e^{nP(\varphi) - \varepsilon}
\]

for all \( n \). From (4.9) we get \( \Lambda_n(\varphi) \leq \Lambda_n(C^p G C^s, \varphi) + K e^{nP(\varphi) - \varepsilon} \), so it suffices to prove the upper bound in (4.4) for \( \Lambda_n(C^p G C^s, \varphi) \).

Write \( a_j = \Lambda_j(G, \varphi) e^{-P(\varphi)} \), and observe that \( a_j \leq Q_2 \) by (4.4). Since every word \( x \in (C^p G C^s)_n \) can be decomposed as \( x = u v w \) where \( u \in C^p \),
SPECIFICATION AND TOWERS IN SHIFT SPACES

$v \in \mathcal{G}$, and $w \in \mathcal{C}^s$, we have

$$\Lambda_n((\mathcal{C}^p \mathcal{G}^s, \varphi) \leq \sum_{i+j+k=n} \Lambda_i(\mathcal{C}^p, \varphi) \Lambda_j(\mathcal{G}, \varphi) \Lambda_k(\mathcal{C}^s, \varphi) \leq K^2 \sum_{i+j+k=n} e^{i(P(\varphi)-\epsilon)} a_j e^{j(P(\varphi)-\epsilon)} e^{k(P(\varphi)-\epsilon)} = K^2 e^{nP(\varphi)} \sum_{i+j+k=n} a_j e^{-(i+k)\epsilon} = K^2 e^{nP(\varphi)} \sum_{m=0}^n a_{n-m}(m+1)e^{-m\epsilon}.$$ (4.10)

Because $a_{n-m} \leq Q_2$ and $\sum_{m=0}^{\infty} (m+1)e^{-m\epsilon} < \infty$, this establishes the second half of (4.4).

Finally, we use (4.9) and (4.10) to show (4.5). Note that it suffices to produce $j \in (n-N,n]$ with $a_j \geq Q_3$. Using (4.9), (4.10) and the first half of (4.4), we have

$$e^{nP(\varphi)} - Ke^{nP(\varphi)-\epsilon} \leq \Lambda_n((\mathcal{C}^p \mathcal{G}^s, \varphi) \leq K^2 e^{nP(\varphi)} \sum_{m=0}^n a_{n-m}(m+1)e^{-m\epsilon},$$

which yields

$$\frac{1}{2} \leq 1 - Ke^{-n\epsilon} \leq K^2 \sum_{m=0}^n a_{n-m}(m+1)e^{-m\epsilon}$$

whenever $n \geq n_0$, where $n_0$ is chosen such that $Ke^{-n_0\epsilon} < \frac{1}{2}$. Thus

$$\frac{1}{2} K^{-2} \leq \sum_{m=0}^{N-1} a_{n-m}(m+1)e^{-m\epsilon} + \sum_{m=0}^{\infty} Q_2(m+1)e^{-m\epsilon},$$

using the inequality $a_{n-m} \leq Q_2$. Now let $N$ be large enough that $K' := \frac{1}{2} K^{-2} - Q_2 \sum_{m=0}^{N-1} (m+1)e^{-m\epsilon} > 0$. Then

$$\sum_{m=0}^{N-1} a_{n-m}(m+1)e^{-m\epsilon} \geq K',$$

and since $(m+1)e^{-m\epsilon} \leq 1/\epsilon$ for every $m \geq 0$, we may put $Q_3 := \epsilon K'/N$ and conclude that there is $0 \leq m < N$ such that $a_{n-m} \geq Q_3$, which completes the proof of Lemma 4.1.

The result from Lemma 4.1 can be used to prove the following important bound. Given $v \in \mathcal{L}_k$ and $1 \leq i \leq n-k$, consider the set

$$(4.11) \quad \mathcal{H}_n(v,i) = \{ w \in \mathcal{L}_n \mid w_{[i,i+k)} = v \}$$

of words where $v$ appears starting in index $i$, but the entries of $w$ before $i$ and after $i+k$ are free to vary (this is the finite-length analogue of a cylinder set). We will mostly be interested in the case when $v \in \mathcal{G}$. In this case we have the following non-stationary version of the Gibbs property (for
the measure-theoretic equivalent, see [CT13, §5.2], [CTY13 Appendix A], and §5.9.

**Proposition 4.2.** There is $Q_4 > 0$ such that for every $v \in G_k$ and $1 \leq i \leq i + k \leq n$, we have

\[
Q_4^{-1} \leq \frac{\Lambda_n(H_n(v, i))}{e^{(n-k)P(\varphi) + \hat{\varphi}(v)}} \leq Q_4.
\]

**Proof.** For the upper bound, observe that (4.3) gives

\[
\Lambda_n(H_n(v, i), \varphi) \leq \Lambda_i(L, \varphi)e^{\hat{\varphi}(v)}\Lambda_{n-(i+k)}(L, \varphi),
\]

and using (4.4) gives $\Lambda_n(\varphi) \leq (Q_2)^2e^{(n-k)P(\varphi)}e^{\hat{\varphi}(v)}$.

![Figure 4.1. Estimating $\Lambda_n(H_n(v, i))$.](image)

For the lower bound, we use the usual specification argument, illustrated in Figure 4.1. By Lemma 4.1, there are $p, q \in \mathbb{N}$ such that

\[
\Lambda_p(G, \varphi) \geq Q_3e^{pP(\varphi)} \quad \text{and} \quad p \in [i, i + N],
\]

\[
\Lambda_q(G, \varphi) \geq Q_3e^{qP(\varphi)} \quad \text{and} \quad i + k + q \in [n, n + N].
\]

Let $w^1 \in G_p$ and $w^2 \in G_q$ be arbitrary. Then by [I] there are $u^1, u^2 \in L$ with $|u^i| \leq \tau$ such that $w^1u^1vu^2w^2 \in G$. Note that $|w^1u^1| \in [i, i + N + \tau]$, and so by truncating at most $\tau + N$ symbols from the beginning and end of $w^1u^1vu^2w^2$, we obtain a word $T(w^1, w^2) \in H_n(v, i)$ with the property that the first $i - |u^1|$ symbols of $T(w^1, w^2)$ match the last $i - |u^1|$ symbols of $w^1$, and similarly for the end of $T(w^1, w^2)$ and the beginning of $w^2$.

This defines a map $T : G_p \times G_q \to H_n(v, i)$. Note that

\[
\hat{\varphi}(T(w^1, w^2)) \geq \hat{\varphi}(w^1u^1vu^2w^2) - (2N + 2\tau)\|\varphi\|
\]

\[
\geq \hat{\varphi}(w^1) + \hat{\varphi}(v) + \hat{\varphi}(w^2) - (2N + 4\tau)\|\varphi\| - 2|\varphi|_d,
\]

Moreover, since the act of truncation removes at most $2(N + \tau)$ symbols from $w^1u^1vu^2w^2$, we see that each word in $H_n(v, i)$ has at most $2(N + \tau)$ preimages under the map $T$. This yields the estimate

\[
\Lambda_n(H_n(v, i)) \geq \frac{1}{2(N + \tau)} \sum_{w^1 \in G_p} \sum_{w^2 \in G_q} e^{\hat{\varphi}(T(w^1, w^2))}
\]

\[
\geq C \sum_{w^1 \in G_p} \sum_{w^2 \in G_q} e^{\hat{\varphi}(w^1)}e^{\hat{\varphi}(v)}e^{\hat{\varphi}(w^2)}
\]

\[
\geq C(Q_3)^2e^{\hat{\varphi}(v)}e^{-(n-k)P(\varphi)},
\]
that for all $n$
Thus for every $m$ we have $|\sum_{k=0}^{n-1} f(T^k x)|$ for all $n \leq k \leq n$.

Let $f$ be such that $\tilde{f}_n(x) = f(T^{k(n)} x)$. Note that $|k(n)|$ is non-decreasing. If $|k(n)| \to \infty$ then we have

$$\left| \frac{\tilde{f}_n(x)}{n} \right| = \left| \frac{f(T^{k(n)} x)}{k(n)} \right| \frac{|k(n)|}{n} \to 0$$

since $|k(n)| \leq n$ and $\frac{f(T^{k(n)} x)}{k(n)} \to 0$ (using the fact that $|k(n)| \to \infty$). If on the other hand the sequence $|k(n)|$ is bounded, say by $k'$, then we have

$$\left| \frac{\tilde{f}_n(x)}{n} \right| \leq \left| \frac{\tilde{f}_{k'}(x)}{n} \right| \to 0.$$
Applying Lemma 4.4 to $T$ and $T^{-1}$, for $\mu$-a.e. $x \in X$ there is $\delta > 0$ (depending on $x$) such that $\delta\|f\|_1 < \varepsilon/3$ and such that for every $m, n \in \mathbb{N}$ with $m \leq \delta n$, we have

\begin{equation}
\frac{1}{n} \sum_{k=1}^{m} f(T^k x) > \frac{\varepsilon}{3}, \quad \frac{1}{n} \sum_{k=1}^{m} f(T^{-k} x) > \frac{\varepsilon}{3}.
\end{equation}

Let $N' = N(x, \varepsilon)/\delta$, and suppose we have $n \geq N'$ and $\ell \in [0, n]$. We want to estimate $\sum_{k=-\ell}^{-\ell+n-1} f(T^k x) = \sum_{k=1}^{\ell} f(T^{-k} x) + \sum_{k=0}^{n-\ell-1} f(T^k x)$. There are three cases to consider: $\ell \in [0, N]$, $\ell \in (N, n-N)$, and $\ell \in [n-N, n]$.

In the second case we can apply (4.15) to get

\[ \left| -\ell \sum_{k=1}^{n-\ell-1} f(T^k x) - n \int f \, d\mu \right| \leq \ell \sum_{k=1}^{n-\ell-1} f(T^{-k} x) - \ell \int f \, d\mu \]

\[ + \sum_{k=0}^{n-\ell-1} f(T^k x) - (n - \ell) \int f \, d\mu \]

\[ \leq \ell \varepsilon/3 + (n - \ell) \delta n \leq n \varepsilon. \]

In the first and third cases we must use one of the inequalities from (4.15) and one from (4.16); for example, if $0 \leq \ell \leq N \leq \delta N' \leq \delta n$, then we have

\[ \left| -\ell \sum_{k=1}^{n-\ell-1} f(T^k x) - n \int f \, d\mu \right| \]

\[ \leq \ell \sum_{k=1}^{n-\ell-1} f(T^{-k} x) \]

\[ + \ell \|f\|_1 + (n - \ell) \varepsilon \]

\[ < n \varepsilon/3 + \delta n \|f\|_1 + n \varepsilon \leq n \varepsilon. \]

The case $\ell \in [n - N, n]$ is analogous. This proves Theorem 3.17.

4.3.3. Proof of Theorem 3.18. Once again let $(X, T, \mu)$ be an invertible ergodic measure-preserving transformation, and let $\alpha$ be a countable measurable partition of $X$ with $H_\mu(\alpha) < \infty$. We want to prove that for $\mu$-a.e. $x \in X$ and every $\varepsilon > 0$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ and all $\ell \in [0, n]$ we have

\begin{equation}
-\frac{1}{n} \log \mu(\alpha^{-\ell+n}(x)) - h_\mu(\alpha, T) < \varepsilon.
\end{equation}

We show that this follows from Theorem 3.17 using the standard argument for proving the Shannon–McMillan–Breiman theorem from the Birkhoff ergodic theorem. We follow the presentation in [Pet89] Theorem 6.2.3.

Let $f_n(x) = -\log \left( \frac{\mu(\alpha^{-n}(x))}{\mu(\alpha_{-n}^{n}(x))} \right)$, and $f^* = \sup_{n \geq 1} f_n$. Then $f^* \in L^1$ [Pet89 Corollary 6.2.2], and as shown in the proof of [Pet89] Theorem 6.2.3, we have $f_n \to f$ both pointwise a.e. and in $L^1$, where $f$ is an $L^1$ function such that $\int f \, d\mu = h(\alpha, T)$. 
Let $I_n(x) = -\log \mu(\alpha_n(x))$, so that $f_n = I_n - I_{n-1} \circ T$. We see that

$$I_n = f_n + I_{n-1} \circ T = f_n + f_{n-1} \circ T + I_{n-2} \circ T = \cdots = \sum_{k=0}^{n-1} f_{n-k} \circ T^k.$$  

Thus (4.18)

$$\frac{1}{n} I_n(x) = \frac{1}{n} S_n f(x) + \frac{1}{n} \sum_{k=0}^{n-1} (f_{n-k} - f)(T^k x).$$

Note that $-\log \mu(\alpha_{-\ell+n}(x)) = I_n(T^{-\ell} x)$, so (4.17) can be rewritten as

$$\frac{1}{n} I_n(T^{-\ell} x) - h_\mu(\alpha, T) < \varepsilon.$$  

We can use (4.18) to get

$$\frac{1}{n} I_n(T^{-\ell} x) = \frac{1}{n} S_n f(T^{-\ell} x) + \frac{1}{n} \sum_{k=0}^{n-1} (f_{n-k} - f)(T^k \ell x).$$

By Theorem 3.17 for $\mu$-a.e. $x \in X$ and every $\varepsilon > 0$ there is $N$ such that for all $n \geq N$ we have

$$\left| \frac{1}{n} S_n f(T^{-\ell} x) - h_\mu(\alpha, T) \right| = \left| \frac{1}{n} S_n f(T^{-\ell} x) - \int f \, d\mu \right| \leq \frac{\varepsilon}{2}.$$  

Thus to prove (4.19) it suffices to show that there is $N'$ such that for all $n \geq N'$ and $0 \leq \ell \leq n$ we have

$$\frac{1}{n} \sum_{k=0}^{n-1} (f_{n-k} - f)(T^k \ell x) < \frac{\varepsilon}{2}.$$  

Given $m \in \mathbb{N}$, let $F_m = \sup_{k \geq m} |f_k - f|$. Then

$$\frac{1}{n} \sum_{k=0}^{n-1} (f_{n-k} - f)(T^k \ell x) = \frac{1}{n} \sum_{k=0}^{n-m} (f_{n-k} - f)(T^k \ell x)$$

$$+ \frac{1}{n} \sum_{k=n-m+1}^{n-1} (f_{n-k} - f)(T^k \ell x).$$

(4.22)

Because $|f_{n-k} - f| \leq f^* + f \in L^1$ for all $k$, we can control the second term by applying Lemma 1.3 to $g := |f_{n-k} - f|$, observing that

$$\frac{1}{n} \sum_{k=n-m+1}^{n-1} (f_{n-k} - f)(T^k \ell x) \leq \frac{m}{n} \hat{g}_n(x) \to 0$$

as $n \to \infty$ since $m$ is fixed. For the first term in (4.22), we observe that

$$\frac{1}{n} \sum_{k=0}^{n-m} (f_{n-k} - f)(T^k \ell x) \leq \frac{1}{n} \sum_{k=0}^{n-m} F_m(T^k \ell x),$$

and since $0 \leq F_m \leq f^* + f \in L^1$, we can apply the strengthened ergodic theorem from the previous section, showing that for $\mu$-a.e. $x$, this is bounded
above by \(2 \int F_m \, du\) for all sufficiently large \(n\). By the dominated convergence theorem and the fact that \(F_m \to 0\) pointwise a.e., this can be made arbitrarily small by taking \(m\) sufficiently large.

5. **Proof of Theorem 3.1**

In this section we assume that the conditions of Theorem 3.1 are satisfied, so that \(F \subset L\) satisfies \([\text{I}_0]\), \([\text{II}]\), and \([\text{III}^*]\). Without loss of generality we assume that \(X\) is a two-sided shift space; if it is one-sided then we pass to the natural extension and define \(\varphi\) to depend only on non-negative coordinates. We will prove the conclusions of Theorem 3.1 for both the one-sided and two-sided shifts (see (5.5)).

5.1. **More counting estimates.** Let \(d\) be the gcd of the lengths \(\{|w| \mid w \in F\}\). Replacing \(\sigma\) with \(\sigma^d\), we will assume without loss of generality that \(d = 1\). Then we have the following estimate, which strengthens (4.5).

**Lemma 5.1.** There is \(Q_5 > 0\) such that \(\Lambda_n(F, \varphi) \geq Q_5 e^{nP(\varphi)}\) for all sufficiently large \(n\).

**Proof.** Since \(\gcd\{|w| \mid w \in F\} = 1\), there is \(m \in \mathbb{N}\) such that for every \(n \geq m\) we have \(n = \sum_{i=1}^{k} a_i |w^i|\) for some \(a_i \in \mathbb{N}\) and \(w^i \in F\). Write \((w^i)^{a_i}\) for the word \(w^i\) repeated \(a_i\) times and note that by \([\text{I}_0]\) we have \(w := (w^1)^{a_1} \cdots (w^k)^{a_k} \in F\), and \(|w| = n\). Thus \(F_n\) is non-empty for every \(n \geq m\).

Now by (4.5) there are \(n_0, N \in \mathbb{N}\) and \(Q_3 > 0\) such that for every \(n \geq n_0 + m\) there is \(j \in [n-m-N, n-m]\) with \(\lambda_j(F, \varphi) \geq Q_3 e^{jP(\varphi)}\). Note that \(n-j \in [m, m+N]\), so by the previous paragraph there is \(w \in F_{n-j}\). Now we can use \([\text{I}_0]\) to get

\[
\Lambda_n(F, \varphi) \geq \sum_{v \in F_j} e^{\hat{\varphi}(v)} \geq e^{-||\varphi||(n-j)} \sum_{v \in F_j} e^{\hat{\varphi}(v)} \geq e^{-||\varphi||(m+N)} \lambda_j(F, \varphi)
\]

\[
\geq e^{-||\varphi||(m+N)} Q_3 e^{jP(\varphi)} \geq e^{-||\varphi||(m+N)} Q_3 e^{nP(\varphi)} e^{-(m+N)P(\varphi)},
\]

which completes the proof of Lemma 5.1.

5.2. **Relating \(X\) to a countable state topological Markov chain.**

With \(X\) and \(F\) as above, we consider the following set of ‘irreducible’ words in \(F\):

\[
I = F \setminus \left( \bigcup_{n \geq 2} F^n \right)
\]

(5.1)

\[
= \{ w \in F \mid w \neq v^1 \cdots v^k \text{ for any } v^1, \ldots, v^k \in F, k \geq 2 \}.
\]

Note that \(F = I^* = \bigcup_{n \geq 1} I^n\), so that every word in \(F\) can be written as a concatenation of words in \(I\). We will say something about the uniqueness of this decomposition below: it requires Condition \([\text{III}^*]\).
We want to model \((X, \sigma)\) using a countable state topological Markov chain. The idea is to consider sequences in \(X\) as concatenations of elements of \(I\). Indeed, given any sequence \(\{w^k\}_{k \in \mathbb{Z}} \in I^\mathbb{Z}\), we can produce an element \(x \in A^\mathbb{Z}\) by concatenating the words \(w^k\); by Condition \([I_0]\), we in fact get \(x \in X\). This relates \((X, \sigma)\) to the full shift on the alphabet \(I\). The relationship is not a conjugacy because shifting one symbol in \(I^\mathbb{Z}\) corresponds to shifting some variable number of symbols in \(X\) (depending on the length of \(w^0\)); we also must deal with the fact that not every element of \(X\) can be obtained from a sequence in \(I^\mathbb{Z}\).

The second issue will be dealt with later; for now we deal with the first. We consider the set \(A_I \subset I \times \mathbb{N}\) defined by
\[
A_I = \{(w, k) \mid w \in I \text{ and } 1 \leq k < |w|\},
\]
and think of \((w, k)\) as representing the state “we are currently in the word \(w\), and have seen the first \(k\) symbols of \(w\)”. We adjoin to \(A_I\) one more symbol \(a\), which will denote a ‘reset state’, to get \(A'_I = A_I \cup \{a\}\). Think of \(a\) as representing the state “we have just completed an element of \(I\) and are ready to start a new one”.

![Figure 5.1. A single loop in the graph representing \(\Sigma\).](image)

Now we let \(\Sigma \subset (A'_I)^\mathbb{Z}\) be the topological Markov chain on the alphabet \(A'_I\) given by allowing the following transitions:
\[
\begin{align*}
(a, (w, 1)) & \mapsto w_1, \\
((w, k), (w, k + 1)) & \mapsto w_{k+1}, \\
((w, |w| - 1), a) & \mapsto w_{|w|},
\end{align*}
\]
Thus the graph representing \(\Sigma\) is a union of loops that intersect only at the vertex \(a\), which is the only vertex of the graph whose in or out degree is greater than 1. Each loop corresponds to an element \(w \in I\), and the length of the loop is the length of \(w\). Figure 5.1 shows a typical such loop corresponding to a word \(w \in \mathcal{F}\) of length 5; by labelling the edges as in the figure we obtain a two-block code \(\pi : \Sigma \rightarrow X\) given by
\[
\begin{align*}
(a, (w, 1)) & \mapsto w_1, \\
((w, k), (w, k + 1)) & \mapsto w_{k+1}, \\
((w, |w| - 1), a) & \mapsto w_{|w|},
\end{align*}
\]
where \( w_k \) denotes the \( k \)th symbol of the word \( w \).

Let \( T : \Sigma \to \Sigma \) be the shift map and note that \( \pi \) commutes with \( T \) and \( \sigma \). We will denote a typical element of \( \Sigma \) as \( z = \{ z_j \}_{j \in \mathbb{Z}} \), where each \( z_j \) is either \( a \) or \((w,k)\) for some \( w \in I, 1 \leq k < |w| \).

Let \( X^+ \) and \( \Sigma^+ \) be the one-sided versions of \( X \) and \( \Sigma \) respectively; that is, define \( p : A^\mathbb{Z} \to A^{\mathbb{R} \cup \{0\}} \) by \( p(\cdots x_{-1} x_0 x_1 \cdots) = x_0 x_1 \cdots \), let \( X^+ = p(X) \), and then similarly define \( \hat{p} : (A'_I)^\mathbb{Z} \to (A'_I)^{\mathbb{R} \cup \{0\}} \) and put \( \Sigma^+ = \hat{p}(\Sigma) \). We write \( \pi^+ : \Sigma^+ \to X^+ \) for the two-block code described above. Now we have the following commutative diagrams.

\[
\begin{array}{ccc}
\Sigma & \xrightarrow{\pi} & X \\
\downarrow{\hat{p}} & & \downarrow{p} \\
\Sigma^+ & \xrightarrow{\pi^+} & X^+
\end{array}
\quad \quad
\begin{array}{ccc}
\mathcal{M}_T(\Sigma) & \xrightarrow{\pi^+} & \mathcal{M}_\sigma(X) \\
\downarrow{\hat{p}_*} & & \downarrow{p_*} \\
\mathcal{M}_T(\Sigma^+) & \xrightarrow{\pi^+} & \mathcal{M}_\sigma(X^+)
\end{array}
\]

The following facts regarding (5.5) are either immediate or well-known.

- The maps \( \pi, \pi^+, p, \hat{p} \) commute with the shifts \( T, \sigma \).
- The fact that \( \pi(\Sigma) \subset X \) is a consequence of [10].
- \( p, \hat{p} \) are not 1-1, but \( p_*, \hat{p}_* \) are [CT12, Proposition 2.1].

We can now outline the remainder of the proof of Theorem 3.1.

- §5.3–5.4: Condition [III] will guarantee that \( \pi \) is 1-1.
- §5.5: The potential \( \varphi : X \to \mathbb{R} \) induces a Hölder potential \( \Phi : \Sigma \to \mathbb{R} \) given by \( \Phi = \varphi \circ \pi \). This is cohomologous to a Hölder potential \( \Phi^+ : \Sigma^+ \to \mathbb{R} \) via a standard procedure [12] and by the estimates from Lemma 5.1, \( \Phi^+ \) is \textbf{positive recurrent} in the sense of Sarig [Sar99].
- §5.6: All of the above applies to \( C^0 \)-small Hölder perturbations of \( \varphi \), so \( \Phi^+ \) is in fact \textbf{strongly positive recurrent} and [Sar01] gives analyticity of the pressure function.
- §5.7: Although \( \pi \) and \( \pi^+ \) are not surjective in general, the positive recurrence condition can be used to show that every equilibrium state on \( X \) gives full weight to \( \pi(\Sigma) \).
- §5.8: Sarig’s results [Sar99] show that there is a unique equilibrium state \( m \) for \( (\Sigma^+, T, \Phi^+) \), and that the log Jacobian of \( m \) is given by \( \Phi^+ \), and hence is Hölder [13]. Moreover, \( m \) is Bernoulli by [Sar11].
- §5.9: The measure \( \mu = \pi_* m \) is Gibbs for \( \varphi \) on \( F \), by Proposition 4.2 and standard constructions. Distribution properties for periodic orbits follow from results in [CT13] as stated in Theorem 7.4.
- §5.10: Since \( \Sigma^+ \) can be related to a Young tower [You99], regularity of the log Jacobian will allow us to deduce exponential decay of correlations and the central limit theorem for \( \Sigma^+ \).

\[12\] This procedure uses the Markov structure of \( \Sigma \) and hence cannot be used directly to get a potential \( \varphi^+ : X^+ \to \mathbb{R} \).

\[13\] Uniqueness can also be deduced from Theorem 7.4 which uses [CT13], but this does not give the needed results on the log Jacobian.
• [5.11] The isomorphism between equilibrium states on $X^+, X, \Sigma, \Sigma^+$
lets us use exponential decay on $\Sigma^+$ to obtain exponential decay on $X^+$ and $X$.

• [5.12] The CLT for $\Sigma^+$ implies the CLT for $\Sigma$ via the standard
cohomological reduction to one-sided observables. Then the CLT
for $\Sigma, \Sigma^+$ gives the CLT for $X, X^+$.

In what follows we carry out the above steps in detail.

5.3. $\mathcal{F}$-marking sets. First we set up some terminology that will help us
relate $\Sigma$ and $X$. Say that $J \subset \mathbb{Z}$ is $\mathcal{F}$-marking for $x \in X$ if $x_{[i,j]} \in \mathcal{F}$ for
all $i, j \in J$ with $i < j$. The set $J$ may be finite or infinite. Say that $i, j \in J$
are consecutive if $i < j$ for all $k$ between $i$ and $j$.

Lemma 5.2. $J \subset \mathbb{Z}$ is $\mathcal{F}$-marking for $x$ if and only if $x_{[i,j]} \in \mathcal{F}$ for all
consecutive $i < j \in J$. In particular, the following are equivalent.

1. $J$ is $\mathcal{F}$-marking.
2. There are $a_k \to -\infty$ and $b_k \to \infty$ such that $J \cap [a_k, b_k]$ is $\mathcal{F}$-marking
for every $k$.
3. $J \cap [a, b]$ is $\mathcal{F}$-marking for every $a < b \in \mathbb{Z}$.

Proof. One direction of the first claim is obvious; the other follows from
Condition [16] The equivalence of the three statements follows from the
first claim.

Say that $J \subset \mathbb{Z}$ is bi-infinite if $J \cap [0, \infty)$ and $J \cap (-\infty, 0]$ are both
infinite. Say that $J \subset \mathbb{Z}$ is maximally $\mathcal{F}$-marking for $x$ if there is no $\mathcal{F}$-
marking set $J' \subset \mathbb{Z}$ with $J' \supsetneq J$. Recall that $I$ is the collection of irreducible
elements of $\mathcal{F}$.

Lemma 5.3. A bi-infinite set $J \subset \mathbb{Z}$ is maximally $\mathcal{F}$-marking for $x$ if and
only if $x_{[i,j]} \in I$ for all consecutive $i < j \in J$.

Proof. Let $J$ be maximally $\mathcal{F}$-marking. Then $x_{[i,j]} \in \mathcal{F}$ for all consecutive
$i < j \in J$. Suppose for a contradiction that there are consecutive $i < j \in J$
for which $x_{[i,j]} \not\in I$; then there is $k \in (i, j)$ such that $x_{[i,k]}, x_{[k,j]} \in \mathcal{F}$. By
Lemma 5.2 this implies that $J \cup \{k\}$ is $\mathcal{F}$-marking, contradicting maximality.

For the other direction, suppose $J$ is not maximal; then there is $k \in \mathbb{Z} \setminus J$
such that $J \cup \{k\}$ is $\mathcal{F}$-marking. Let $i < k < j$ be such that $i < j \in J$ are
consecutive (here we use that $J$ is bi-infinite). Then $x_{[i,j]} = x_{[i,k]}x_{[k,j]} \in \mathcal{F}I$, so
$x_{[i,j]} \not\in I$.

Lemma 5.4. Given $z \in \Sigma$, the set $J(z) := \{j \mid z_j = a\}$ is bi-infinite and
maximally $\mathcal{F}$-marking for $\pi(z) \in X$.

Proof. To see that $J(z)$ is bi-infinite, pick any $j \not\in J(z)$; then $z_j = (w, k)$ for
some $w \in I$ and $1 \leq k < |w|$, and the transition rules (5.3) guarantee that
$z_{j-k} = z_{j+|w|-k} = a$.

By Lemma 5.3 it suffices to check that $x_{[i,j]} \in I$ whenever $i < j \in J(z)$
are consecutive. In this case we have $j > i + 1$ since $a \to a$ is not a legal
transition in $\Sigma$, and so $x_{i+1} = (w, 1)$ for some $w \in I$. We see from (5.3) that $x_{[i,j)} = w \in I$.

The following is a sort of converse to Lemma 5.4.

**Lemma 5.5.** If $J \subset \mathbb{Z}$ is bi-infinite and maximally $F$-marking for $x \in X$, then there is exactly one $z \in \Sigma$ such that $\pi(z) = x$ and $J(z) = J$.

**Remark 5.6.** Before proving Lemma 5.5, we observe that it does not yet prove injectivity of $\pi$, since it is a priori possible that some $x \in X$ has multiple maximally $F$-marking sets. In the next section, we prove injectivity by showing uniqueness of the maximal $F$-marking set.

**Proof of Lemma 5.5.** Suppose $z \in \Sigma$ has $\pi(z) = x$ and $J(z) = J$. Then $z_j = a$ for all $j \in J$, and given consecutive $i < j \in J$ we have $w := x_{[i,j)} \in I$. By the definition of $\Sigma$ we have $x_{i+1} = (v, 1)$ for some $v \in I$. We must have $|v| = j - i = |w|$ since the next occurrence of $a$ in $z$ is at $z_j$. Moreover, for each $i \leq k < j$, we have $\pi(z)_k = v_{k-i+1}$, and so $\pi(z) = x$ implies $v = w$. The above argument shows that $z$ is uniquely determined by $x$ and $J$. Moreover, defining $z$ by $z_j = a$ for all $j \in J$ and $z_{i+1} = (x_{[i,j)}, 1)$ for all consecutive pairs $i < j \in J$ gives an element $z \in \Sigma$ with $\pi(z) = x$ and $J(z) = J$, establishing existence. □

Every bi-infinite $F$-marking set $J$ is contained in a bi-infinite maximal $F$-marking set: this follows by taking every pair of consecutive $i < j \in J$ and decomposing $x_{[i,j)}$ as a concatenation of elements of $I$. Adding the indices marking these decompositions to $J$ gives a maximal $F$-marking set $J' \supset J$. The following lemma is a consequence of this observation together with Lemmas 5.4 and 5.5.

**Lemma 5.7.** Given $x \in X$, we have $x \in \pi(\Sigma)$ if and only if there is a bi-infinite $F$-marking set $J \subset \mathbb{Z}$ for $x$. There is a 1-1 correspondence between elements of $\pi^{-1}(x) \subset \Sigma$ and bi-infinite maximal $F$-marking sets for $x$.

In the next section we show that every bi-infinite $F$-marking set is contained in a unique bi-infinite maximal $F$-marking set, which will show that $\pi$ is 1-1. In §5.7, we show that if $\mu$ is any equilibrium state for $(X, \phi)$, then $\mu$-a.e. $x \in X$ has a bi-infinite maximal $F$-marking set, and hence $\mu(\pi(\Sigma)) = 1$.

5.4. **Injectivity of $\pi$.** In light of the previous section, to show injectivity of $\pi$ it suffices to show that if $x \in X$ has a bi-infinite $F$-marking set, then it has a unique maximal such set. We accomplish this by showing that arbitrary unions of bi-infinite $F$-marking sets are still bi-infinite and $F$-marking (Lemma 5.9 below).

The following result will be important both here and later on, when we prove that $\pi(\Sigma)$ has full weight for any equilibrium state. This is the only place in the proof of Theorem 3.1 where we use [III*].
Lemma 5.8. Let \( \{J_\lambda\}_{\lambda \in \Lambda} \) be any collection of sets \( J_\lambda \subset \mathbb{Z} \) such that each \( J_\lambda \) is \( \mathcal{F} \)-marking for \( x \). Let \( r < s \in \mathbb{Z} \) be such that \( r > \min J_\lambda \) and \( s < \max J_\lambda \) for all \( \lambda \). Then \( (\bigcup \lambda \) \( J_\lambda \) \( ) \cap [r, s] \) is \( \mathcal{F} \)-marking for \( x \).

Proof. Pick \( j, k \in \bigcup \lambda \) \( J_\lambda \) with \( r < j, k \leq s \). Let \( \lambda, \lambda' \) be such that \( j \in J_\lambda \) and \( k \in J_{\lambda'} \). Because \( \min J_{\lambda'} \leq a \) there is \( i \in J_{\lambda'} \) with \( i \leq j \); similarly, \( \max J_\lambda \geq b \) implies that there is \( \ell \in J_\lambda \) with \( \ell \geq k \). Thus \( i \leq j < k \leq \ell \) are such that \( x_{[i,k]}, x_{(j,\ell)} \in \mathcal{F} \). (See Figure 5.2)

Moreover, choosing \( a \in J_\lambda \) with \( a < r \) and \( b \in J_{\lambda'} \) with \( b > s \), we have \( x_{[a,j)}, x_{[k,b)} \in \mathcal{F} \), and it follows from Condition III\( ^* \) that \( x_{[j,k)} \in \mathcal{F} \). This holds for all \( j, k \in (\bigcup \lambda \) \( J_\lambda \) \( ) \cap [r, s] \), so we are done.

\[\begin{aligned}
&J_{\lambda'} \quad i \quad r \in \mathcal{F} \quad k \quad s \quad b \quad \ell \\
&J_\lambda \quad a \quad j \end{aligned}\]

Figure 5.2. The union of \( \mathcal{F} \)-marking sets is \( \mathcal{F} \)-marking.

When the sets \( J_\lambda \) in Lemma 5.8 are all bi-infinite, we can take \( a \to -\infty \) and \( b \to \infty \) and obtain the following result using Lemma 5.2

Lemma 5.9. Let \( \{J_\lambda\}_{\lambda \in \Lambda} \) be any collection of bi-infinite \( \mathcal{F} \)-marking sets \( J_\lambda \subset \mathbb{Z} \) for \( x \in X \). Then \( \bigcup \lambda \) \( J_\lambda \) is bi-infinite and \( \mathcal{F} \)-marking for \( x \).

Now we can prove injectivity of \( \pi \) as follows: given \( x \in \pi(\Sigma) \), let \( \{J_\lambda\}_{\lambda \in \Lambda} \) be the collection of all bi-infinite \( \mathcal{F} \)-marking sets for \( x \). This collection is non-empty by Lemma 5.4. Let \( J = \bigcup \lambda \) \( J_\lambda \). By Lemma 5.9, \( J \) is bi-infinite and \( \mathcal{F} \)-marking for \( x \). Moreover, if \( J' \) is any bi-infinite \( \mathcal{F} \)-marking set, we have \( J' \subset J \) by construction, so \( J \) is maximal, and it is the only bi-infinite maximal \( \mathcal{F} \)-marking set. This proves injectivity of \( \pi \).

5.5. Correspondence between \( (X, \sigma, \varphi) \) and \( (\Sigma^+, T, \Phi^+) \). From (2.4) we have \( |\varphi(x) - \varphi(y)| \leq |\varphi|_\beta e^{-\beta n} \) whenever \( x, y \in X \) have \( x_k = y_k \) for all \( |k| \leq n \). Define \( \Phi: \Sigma \to \mathbb{R} \) by \( \Phi = \varphi \circ \pi \); then for every \( z, z' \in \Sigma \) with \( z_k = z'_k \) for all \( |k| \leq n \), we have \( \pi(z)_k = \pi(z')_k \) for all \( |k| \leq n - 1 \) (since \( \pi \) is a 2-block code) and hence \( |\Phi(z) - \Phi(z')| \leq |\varphi|_\beta e^{-\beta n} \).

As in (2.2) we write \( C_\beta(\Sigma) \) for the set of all functions \( u: \Sigma \to \mathbb{R} \) such that there is \( |u|_\beta > 0 \) that makes the following hold for every \( n \geq 0 \).

\[|u(z) - u(z')| \leq |u|_\beta e^{-\beta n} \] whenever \( z_k = z'_k \) for all \( |k| \leq n \).

If (5.6) holds for all \( n \geq 1 \), but not necessarily for \( n = 0 \), we say that \( u \) is locally Hölder continuous [Sar99].

Again as in (2.2) we write \( C^h(\Sigma) = \bigcup_{\beta > 0} C_\beta(\Sigma) \). The first paragraph above shows that \( \Phi \in C^h(\Sigma) \).

Now we need the following result, which is taken almost verbatim from [Bow75] Lemma 1.6. We reproduce the proof here for completeness and
since we work in the countable state setting, which formally is not covered by [Bow75] (though the proof is the same).

**Lemma 5.10.** Given \( \psi \in C^h(\Sigma) \), there is a bounded function \( u \in C^h(\Sigma) \) such that the function \( \psi^+ := \psi - u \circ T \in C^h(\Sigma) \) only depends on non-negative coordinates; that is, \( \psi^+(z) = \psi^+(z') \) whenever \( z_k = z'_k \) for all \( k \geq 0 \). The maps \( \psi \mapsto u \) and \( \psi \mapsto \psi^+ \) are linear.

The map \( \psi^+ \) can be considered as a function \( \Sigma^+ \to \mathbb{R} \), and is Hölder continuous with the same constant and exponent as \( \psi^+: \Sigma \to \mathbb{R} \). Finally, for any \( z \in \Sigma \) we have

\[
S_n \psi^+(\hat{p}(z)) - S_n \psi(z) \leq 2\|u\|.
\]

**Proof.** For every symbol \( a \in A' \) choose \( y^a \in \Sigma \) with \( y^a_0 = a \). Define \( r: \Sigma \to \Sigma \) by

\[
r(z)_k = \begin{cases} 
z_k & \text{for } k \geq 0, 
y^a_k & \text{for } k \leq 0,
\end{cases}
\]

and let

\[
u(z) = \sum_{j=0}^{\infty} (\psi(T^jz) - \psi(T^jr(z))).
\]

Now \( T^jz \) and \( T^jr(z) \) agree in all indices from \(-j\) to \(+\infty\), so each term of the above sum is bounded in absolute value by \( |\psi|_\beta e^{-\beta j} \). This implies that \( u \) is defined, bounded, and continuous. Moreover, if \( z_i = z'_i \) for all \( |i| \leq n \), then for every \( j \in [0,n] \), we have

\[
|\psi(T^jz) - \psi(T^jz')| \leq |\psi|_\beta e^{-\beta(n-j)}
\]

and

\[
|\psi(T^jr(z)) - \psi(T^jr(z'))| \leq |\psi|_\beta e^{-\beta(n-j)}.
\]

It follows that

\[
|u(z) - u(z')| \leq \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} |\psi(T^jz) - \psi(T^jz')| + |\psi(T^jr(z)) - \psi(T^jr(z'))| + 2 |\psi|_\beta \sum_{j>\lfloor \frac{n}{2} \rfloor} e^{-\beta j}
\]

\[
\leq 2 |\psi|_\beta \left( \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} e^{-\beta(n-j)} + \sum_{j>\lfloor \frac{n}{2} \rfloor} e^{-\beta j} \right)
\]

\[
\leq 4 |\psi|_\beta \sum_{j>\lfloor \frac{n}{2} \rfloor} e^{-\beta j} \leq \frac{4 |\psi|_\beta}{1 - e^{-\beta |\frac{n}{2}|}},
\]
and hence \( u \in C^h \). Since \( C^h \) is a \( T \)-invariant vector space, it follows that \( \psi^+ := \psi - u + u \circ T \) is in \( C^h \) as well. Observe that

\[
u \circ T(z) = \sum_{j=0}^{\infty} (\psi(T^{j+1}z) - \psi(T^j r(Tz))),
\]

\[-u(z) = \sum_{j=0}^{\infty} (\psi(T^{j+1}r(z)) - \psi(T^{j+1}z)),
\]

so that \((u \circ T - u)(z) = \psi(r(z)) - \psi(z) + \sum_{j=0}^{\infty} (\psi(T^{j+1}r(z)) - \psi(T^j r(Tz))).\) In particular, we have

\[
\psi^+(z) = \psi(z) + (u \circ T - u)(z)
\]

\[
= \psi(r(z)) + \sum_{j=0}^{\infty} (\psi(T^{j+1}r(z)) - \psi(T^j r(Tz))).
\]

Since the last line depends only on \( r(z) \) and \( r(Tz) \), which depend only on \( z_k \) for \( k \geq 0 \), we have proved the claims in the first paragraph of Lemma 5.10.

For the second paragraph, we observe that if \( \gamma, |\psi^+|_\gamma > 0 \) are such that \((5.6)\) holds for all \( z, z' \in \Sigma \), then for all \( z, z' \in \Sigma^+ \) we can consider \( y = r(z), y' = r(z') \) as in the first part of the proof, and observe that if \( n \) is such that \( z_k = z'_k \) for all \( |k| \leq n \), then \( y_k = y'_k \) for all \( |k| \leq n \), and so

\[
|\psi^+(z) - \psi^+(z')| = |\psi^+(y) - \psi^+(y')| \leq |\psi^+| \gamma e^{-\gamma n},
\]

as required. Finally, given any \( z \in \Sigma \) we have

\[
S_n \psi^+(\hat{\rho}(z)) = (S_n \psi(z)) - u(z) + u(T^n z),
\]

which proves \((5.7)\). \( \square \)

Applying Lemma 5.10 to the function \( \Phi = \varphi \circ \pi: \Sigma \to \mathbb{R} \), we obtain a Hölder function \( u: \Sigma \to \mathbb{R} \) such that \( \Phi^+ = \Phi - u + u \circ T \) depends only on the non-negative coordinates, and thus may be considered as a Hölder continuous function on \( \Sigma^+ \). Following \cite{Sar99}, we write

\[
Z_n(\Phi^+, a) := \sum_{T^n \mathbf{z} = \mathbf{z}, \mathbf{z}_0 = 0} e^{S_n \Phi^+(\mathbf{z})},
\]

\[
Z^*_n(\Phi^+, a) := \sum_{T^n \mathbf{z} = \mathbf{z}, \mathbf{z}_0 = 0} e^{S_n \Phi^+(\mathbf{z})}.
\]

**Lemma 5.11.** There is \( Q_6 > 0 \) such that for every \( n \) we have

\[
e^{-Q_6} \Lambda_n(\mathcal{F}, \varphi) \leq Z_n(\Phi^+, a) \leq e^{Q_6} \Lambda_n(\mathcal{F}, \varphi),
\]

\[
e^{-Q_6} \Lambda_n(\mathcal{I}, \varphi) \leq Z^*_n(\Phi^+, a) \leq e^{Q_6} \Lambda_n(\mathcal{I}, \varphi).
\]
Proof. Given a word $w \in \mathcal{F} \subset \mathcal{L}$, let $w^1, \ldots, w^k \in I$ be such that $w = w^1 \cdots w^k$. Then let $\tau(w) \in \Sigma$ be the periodic point of $\Sigma$ that corresponds to starting at $a$ and following in sequence the loops labelled by $w^1, \ldots, w^k$. We see that

\begin{align}
Z_n(\Phi^+, a) &= \sum_{w \in \mathcal{F}} e^{S_n \Phi^+(\hat{p} \circ \tau(w))}, \\
Z_n^*(\Phi^+, a) &= \sum_{w \in I} e^{S_n \Phi^+(\hat{p} \circ \tau(w))}.
\end{align}

As in (4.1) we have

$$|S_n \varphi(x) - S_n \varphi(y)| \leq |\varphi|_d < \infty \text{ for all } x, y \in [w], w \in \mathcal{L}_n.$$ 

Together with (5.7), we see that for every $w \in \mathcal{F}_n$ we have

$$|S_n \varphi(x) - S_n \varphi(y)| \leq |\varphi|_d + 2 \|u\|.$$ 

Together with (5.10), this completes the proof of Lemma 5.11. □

By Lemmas 4.1 and 5.1 we have

$$Q_5 e^{nP(\varphi)} \leq \Lambda_n(\mathcal{F}, \varphi) \leq Q_2 e^{nP(\varphi)}$$

for all sufficiently large $n$, and so

$$Q_5 e^{-Q_6 e^{nP(\varphi)}} \leq Z_n(\Phi^+, a) \leq Q_2 e^{Q_6 e^{nP(\varphi)}}.$$ 

This implies that $P(\varphi)$ is equal to the Gurevich pressure

$$P_G(\Phi^+) := \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Phi^+, a),$$

and that $\Phi^+$ is positive recurrent (see [Sar99, Definition 2]).\footnote{In [Sar99], positive recurrence is the condition that there are some constants so that (5.11) holds; a different but equivalent definition is used in [Sar01].}

In the next section we show that $\Phi^+$ is in fact strongly positive recurrent.\footnote{This amounts to the condition that a certain discriminant is positive; see the proof of Lemma 5.13.}

5.6. Strong positive recurrence and analyticity. Note that of the conditions [I$_0$] [II] and [III$^*$], Condition [II] is the only one that depends on the potential $\varphi$. It follows from the definition of pressure that for every Hölder continuous $\psi: X \to \mathbb{R}$, we have

$$P(\mathcal{E}, \varphi + \psi) \leq P(\mathcal{E}, \varphi) + \|\psi\|,$$

$$P(\varphi + \psi) \geq P(\varphi) - \|\psi\|.$$ 

In particular, writing $\delta = P(\varphi) - P(\mathcal{E}, \varphi) > 0$, we see that if $\|\psi\| < \delta/2$, then $(X, \varphi + \psi)$ satisfies [I$_0$] [II] and [III$^*$] as well.

Lemma 5.12. Given any $\psi: X \to \mathbb{R}$ Hölder continuous, there is $\varepsilon > 0$ such that for every $t \in (-\varepsilon, \varepsilon)$, the potential function $\Phi^+ + t\psi$ is positive recurrent on $\Sigma^+$.\footnote{In [Sar99], positive recurrence is the condition that there are some constants so that (5.11) holds; a different but equivalent definition is used in [Sar01].}
Proof. This follows immediately from the discussion above; let \( \varepsilon > 0 \) be such that \( t \| \psi \| < \delta / 2 \), then Conditions \([I_0]\), \([II]\) and \([III^*]\) hold for \((X, \varphi + t \psi)\) for every \( t \in (-\varepsilon, \varepsilon) \), so we obtain a positive recurrent potential \( \Phi^+ + t \Psi^+ \). □

Lemma 5.13. \( \Phi^+ \) is strongly positive recurrent on \( \Sigma^+ \), in the sense of [Sar01, Definition 1].

Proof. Fix \( w \in I \) and let \( b = (w, 1) \in A'_1 \). The definition of strong positive recurrence involves the discriminant \( \Delta_b[\Phi^+] \) (see [Sar01, Definition 1]). Positive recurrence implies that \( \Delta_b[\Phi^+] \geq 0 \) (but not vice versa) [Sar01, Theorem 2]; \( \Phi^+ \) is strongly positive recurrent if \( \Delta_b[\Phi^+] > 0 \).

Let \( \psi = 1_{[w]} : X \rightarrow \mathbb{R} \) be the characteristic function of the cylinder \([w]\); then \( \Psi^+ : \Sigma^+ \rightarrow \mathbb{R} \) is the characteristic function of the cylinder \([ab]\). Suppose \( \Delta_b[\Phi^+] = 0 \); then as in the proof of [Sar01, Corollary 2], we have \( \Delta_b[\Phi^+ + t\Psi^+] = t \) for all \( t \in \mathbb{R} \).\(^{16}\) In particular, for \( t < 0 \) we have \( \Delta_b[\Phi^+ + t\Psi^+] < 0 \), contradicting the result of Lemma 5.12 that \( \Phi^+ + t\Psi^+ \) is positive recurrent for all \( |t| < \varepsilon \). Thus \( \Delta_b[\Phi^+] > 0 \), so \( \Phi^+ \) is strongly positive recurrent. □

We list two important consequences of strong positive recurrence.

Lemma 5.14. For any Hölder continuous \( \psi : X \rightarrow \mathbb{R} \), there is \( \varepsilon > 0 \) such that the map \( t \mapsto P(\varphi + t\psi) \) is real analytic on \((-\varepsilon, \varepsilon)\).

Proof. This follows from strong positive recurrence, [Sar01, Theorem 3], and the observation that for every Hölder \( \psi : X \rightarrow \mathbb{R} \), we have \( P_G(\Phi^+ + t\Psi^+) = P(\varphi + t\psi) \leq \infty \) whenever \( |t| < \varepsilon \). □

Lemma 5.15. \( \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\Phi^+, a) < P(\varphi) \) and \( P(I, \varphi) < P(\varphi) \).

Proof. The first claim follows from strong positive recurrence and [Sar01, Lemma 5]. The second follows from the first and Lemma 5.11. □

5.7. Liftability. Now we show that any equilibrium state \( \mu \) for \((X, \varphi)\) gives full weight to \( \pi(\Sigma) \subset X \). It suffices to consider the case when \( \mu \) is ergodic.

We use the criterion from Lemma 5.7 given \( x \in X \), we have \( x \in \pi(\Sigma) \) if and only if there is a bi-infinite \( \mathcal{F}\)-marking set \( J \subset \mathbb{Z} \) for \( x \).

We want to construct such a set as follows: given \( x \in X \), we consider the words \( x_{[-n,n]} \). Unless \( x_{[-n,n]} \in \mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s \) (which should be rare since this set has small pressure), there are \( a < b \in [-n, n] \) such that \( x_{[-n,a]} \in \mathcal{E}^p \), \( x_{[a,b]} \in \mathcal{F} \), and \( x_{[b,n]} \in \mathcal{E}^s \). Since \( \mathcal{F} = I^s \), the word \( x_{[a,b]} \) can be decomposed as a composition of words in \( I \); the positions where these words start and end gives a \( \mathcal{F}\)-marking set \( J \subset [a, b] \).

In order to use the sets \( J \) to produce a bi-infinite \( \mathcal{F}\)-marking set for \( x \), we need some information about which words \( x_{[i,j]} \) can be contained in a single element of \( I, \mathcal{E}^p, \) or \( \mathcal{E}^s \).

Let \( \mathcal{E}' = I \cup \mathcal{E} \subset \mathcal{L} \). By Lemma 5.15 and \([III]\), we have \( P(\mathcal{E}', \varphi) < P(\varphi) \).

\(^{16}\)This is because when \( T : \Sigma^+ \rightarrow \Sigma^+ \) is induced on the cylinder \( b \), \( \Psi^+ \) induces to the constant function \( 1 \), and \( p_b[\Phi^+ + t\Psi^+] = p_b[\Phi^+] \) (see [Sar01, Definition 1]).
Definition 5.16. We say that \( R \subset \mathbb{Z} \) is \( \mathcal{E}' \)-restricting for \( x \in X \) if for every \( i < j \in \mathbb{Z} \) with \( x_{[i,j]} \in \mathcal{E}' \), the interval \([i, j]\) contains at most one element of \( R \).

First we prove that to check \( x \in \pi(\Sigma) \) it suffices to check existence of a bi-infinite \( \mathcal{E}' \)-restricting set; then we show that \( \mu \)-a.e. \( x \in X \) has such a set.

**Lemma 5.17.** If \( x \in X \) has a bi-infinite \( \mathcal{E}' \)-restricting set \( R \subset \mathbb{Z} \), then it has a bi-infinite \( \mathcal{F} \)-marking set \( J \), and hence \( x \in \pi(\Sigma) \) by Lemma 5.7.

**Proof.** Enumerate \( R \) as \( R = \{r_n\}_{n \in \mathbb{Z}} \) where \( r_n \) is increasing. Given \( n \in \mathbb{N} \), note that \( x_{[r_n, r_n]} \in \mathcal{E}^p \mathcal{F} \mathcal{E}^s \) since otherwise we would have a word in \( \mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s \subset \mathcal{E} \) that crosses more than one index in \( R \). Thus we can apply the decomposition \( \mathcal{E}^p \mathcal{F} \mathcal{E}^s \) to \( x_{[r_n, r_n]} \) to get \( j' \leq j'' \in [r_n, r_n] \) such that

\[
 x_{[r_n, j']} \in \mathcal{E}^p, \quad x_{[j', j'']} \in \mathcal{F}, \quad x_{[j'', r_n]} \in \mathcal{E}^s.
\]

Now use the decomposition \( \mathcal{F} = I^* = \bigcup_{k \in \mathbb{N}} I^k \) to get an increasing sequence \( \{j^k_i\}_{i=1}^k \) such that \( j^1_i = j' \), \( j^k_i = j'' \), and \( x_{[j^k_i, j^k_{i+1}]} \in I \) for every \( 1 \leq i < k \).

Put \( j^0_0 = r_n \) and \( j^0_{k+1} = r_n \). Then \( x_{[j^0_0, j^0_{k+1}]} \in \mathcal{E}' \) for every \( 0 \leq i \leq k \). Let \( J_n = \{j^0_i\}_{i=1}^k \) and note that \( J_n \) is \( \mathcal{F} \)-marking for \( x \). Since \( R \) is \( \mathcal{E}' \)-restricting, for every \( \ell \in [-n, n] \) the interval \([r_{\ell}, r_{\ell+1}] \) contains at least one element of \( J_n \); in particular, we have \( \min J_n \leq r_{-(n-1)} \) and \( \max J_n \geq r_{n-1} \).

---

**Figure 5.3.** Constructing a bi-infinite \( \mathcal{F} \)-marking set.

Given \( N \in \mathbb{N} \) let \( J'_N = \bigcup_{n>N} J_n \). By the previous paragraph, \( \min J_n \leq r_{-N} \) and \( \max J_n \geq r_N \) for every \( n > N \), and every \( J_n \) is \( \mathcal{F} \)-marking for \( x \), so by Lemma 5.8 \( J'_N \cap [r_{-N}, r_N] \) is \( \mathcal{F} \)-marking for \( x \).

Let \( J = \bigcap_{N \in \mathbb{N}} J'_N \). Then \( J \cap [r_{-N}, r_N] \) is \( \mathcal{F} \)-marking for every \( N \), hence by Lemma 5.2 \( J \) is \( \mathcal{F} \)-marking for \( x \). It remains to show that \( J \) is bi-infinite.

Given \( \ell \in \mathbb{Z} \), there is \( i \in \bigcap_{\ell} [r_{\ell}, r_{\ell+1}] \) such that \( i \in J_n \) for infinitely many \( n \). Thus \( i \in J'_N \) for every \( N \), so \( i \in J \). It follows that \( J \) is bi-infinite. □

**Lemma 5.18.** Let \( \mu \) be any equilibrium state for \((X, \varphi)\). Then for \( \mu \)-a.e. \( x \in X \), there is \( n = n(x) \) such that for all \( k \geq n \) and all \( \ell \in [0, k] \) we have \( x_{[-\ell, k-\ell]} \notin \mathcal{E}' \).
Proof. Fix $\varepsilon > 0$ such that $P(\varphi) - 5\varepsilon > P(\mathcal{E}', \varphi)$. By Theorems 3.17 and 3.18 for $\mu$-a.e. $x \in X$ there is $N_x \in \mathbb{N}$ such that for all $n \geq N_x$ and $\ell \in [0, n]$ we have

$$
\mu[x_{-\ell, -\ell+n}] \leq e^{-nh(\mu)+n\varepsilon},
$$

(5.13)

$$
S_n\varphi(\sigma^{-\ell}x) \geq n\left(\int \varphi \, d\mu - \varepsilon \right),
$$

$n\varepsilon \geq |\varphi|_d$.

Let $A_n = \{x \in X \mid N_x \leq n\}$; then $\mu(A_n) \to 1$ as $n \to \infty$. If $k \geq n$ and $w \in \mathcal{L}_k$ is such that $\sigma^\ell[w] \cap A_n \neq \emptyset$ for some $0 \leq \ell \leq k$, then we can choose $x$ in the intersection, so that $x_{-\ell, -\ell+k} = w$ and $k \geq N_x$; then (5.13) gives

$$
\mu[w] \leq e^{-nh(\mu)+n\varepsilon},
$$

and

$$
\hat{\varphi}(w) \geq (S_n\varphi(\sigma^{-\ell}x)) - |\varphi|_d \geq n\left(\int \varphi \, d\mu - 2\varepsilon \right),
$$

so we get

$$
(5.14) \quad \mu[w]e^{-\hat{\varphi}(w)} \leq e^{-n(h(\mu)+\int \varphi \, d\mu - 3\varepsilon)} = e^{-nP(\varphi)-3\varepsilon} \leq e^{-nP(\mathcal{E}', \varphi)}e^{-2n\varepsilon}.
$$

Now we consider the sets $B_n = \bigcup_{k \geq n} \bigcup_{0 \leq \ell \leq k} \sigma^\ell[\mathcal{E}'_k]$; that is, $B_n$ is the set of points $x$ for which there is some $k \geq n$ with and $0 \leq \ell \leq k$ with $x_{-\ell, -\ell+k} \in \mathcal{E}'$. Now (5.14) gives

$$
\mu(A_n \cap B_n) \leq \sum_{k \geq n} \sum_{w \in \mathcal{E}'_k} \mu(\sigma^\ell[w] \cap A_n)
$$

$$
\leq \sum_{k \geq n} (k+1) \sum_{w \in \mathcal{E}'_k} e^{-kP(\mathcal{E}', \varphi)}e^{-2k\varepsilon}e^{\hat{\varphi}(w)}
$$

$$
= \sum_{k \geq n} (k+1)e^{-kP(\mathcal{E}', \varphi)}e^{-2k\varepsilon}\Lambda_k(\mathcal{E}', \varphi).
$$

Choose $C$ such that $\Lambda_k(\mathcal{E}', \varphi) \leq C e^{k(P(\mathcal{E}', \varphi)+\varepsilon)}$ for all $k$. Then

$$
\mu(A_n \cap B_n) \leq C \sum_{k \geq n} (k+1)e^{-k\varepsilon} \to 0 \text{ as } n \to \infty,
$$

and it follows that

$$
\mu(B_n) \leq \mu(X \setminus A_n) + \mu(A_n \cap B_n) \to 0 \text{ as } n \to \infty.
$$

In particular, we get $\mu(\cap_{n \in \mathbb{N}} B_n) = 0$, and so for $\mu$-a.e. $x \in X$ there is $n > 0$ such that $x \in B_n$, which proves Lemma 5.18.

\begin{lemma}
Let $\mu$ be any equilibrium state for $(X, \varphi)$. Then $\mu$-a.e. $x \in X$ has a bi-infinite $\mathcal{E}'$-restricting set. In particular, $\mu(\pi(\Sigma)) = 1$.
\end{lemma}

\begin{proof}
Let $E$ be the set of points satisfying the conclusion of Lemma 5.18, that is, for every $x \in E$ there is $n(x)$ such that for all $k \geq n(x)$ and all $\ell \in [0, k]$ we have $x_{-\ell, -k-\ell} \notin \mathcal{E}'$. By Lemma 5.18, we have $\mu(X \setminus E) = 0$, hence $\mu(\sigma^{-m}(X \setminus E)) = 0$ for every $m \in \mathbb{Z}$, and we conclude that $E' := \cap_{m \in \mathbb{Z}} \sigma^{-m}E$
has full $\mu$-measure. For every $x \in E'$ and every $m \in \mathbb{N}$ there is $n(m)$ such that for all $a \leq m \leq b$ with $b - a \geq n(m)$, we have $x_{(a,b)} \notin E'$.

Given $x \in E'$, define a bi-infinite sequence $r_j \in \mathbb{Z}$ by

1. $r_0 = 0$;
2. $r_{j+1} = r_j + n(r_j)$ for $j \geq 0$;
3. $r_{j-1} = r_j - n(r_j)$ for $j \leq 0$.

Let $R = \{r_j\}_{j \in \mathbb{Z}}$, and note that $R$ is bi-infinite. We claim that $R$ is $E'$-restricting for $x$. Note that by the construction of $R$, we have $r_{j+1} - r_j \geq \min(n(r_j), n(r_{j+1}))$ for every $j \in \mathbb{Z}$. Thus if $a < b \in \mathbb{Z}$ are such that $a \leq r_j$ and $b \geq r_{j+1}$, we either have $b - a \geq \ell(r_j)$ or $b - a \geq \ell(r_{j+1})$. By the definition of $n$, this is impossible, since $r_j, r_{j+1} \in [a, b]$. We conclude that $[a, b]$ can contain at most one member of $R$, hence $R$ is $E'$-restricting for $x$. \hfill \Box

### 5.8. Unique equilibrium state and regular log Jacobian

If $m$ is an ergodic $T$-invariant probability measure on $\Sigma^+$, then $\pi_\ast((\hat{p}_\ast)^{-1}m)$ is an ergodic $\sigma$-invariant probability measure on $X$ with the same entropy as $m$ (since $\pi$ is 1-1). It follows that $h(m) \leq h_{\text{top}}(X)$ for every $m$. Moreover, $\int \Phi^+ dm < \infty$ for every $m$ since $\Phi^+$ is bounded.

Because $\Phi^+$ is positive recurrent on $\Sigma^+$, [Sar99, Theorem 4] gives the existence of a $\sigma$-finite measure $\nu$ on $\Sigma^+$ and a function $h > 0$ such that $L^\ast \nu = e^{P(\varphi)}\nu$, $Lh = e^{P(\varphi)}h$, and $\nu(h) = 1$, where $L$ is the Ruelle–Perron–Frobenius operator associated to the potential $\Phi^+$. The measure $m$ defined by $dm = h d\nu$ is a $T$-invariant probability measure. By [Sar99, Theorem 7] and the remarks in the previous paragraph, it can be characterised as the only $T$-invariant probability measure such that

\begin{equation}
(5.15) \quad h(m) + \int \Phi^+ dm = P_G(\Phi^+) = P(\varphi).
\end{equation}

In particular, writing $\mu = \pi_\ast \hat{p}_\ast^{-1}m$, we see that $\mu$ is the unique equilibrium state for $(X, \varphi)$. (See also [BS03, Theorems 1.1 and 1.2].)

We also observe that by [Sar99, Theorem 6], we have

\begin{equation}
(5.16) \quad \Phi^+ = \log \frac{dm}{d\mu \circ T} + \psi \circ T - \psi + P(\varphi)
\end{equation}

for some locally Hölder continuous $\psi: \Sigma^+ \to \mathbb{R}$. (In fact we can take $\psi = \log h$.) This implies that $\log \frac{dm}{d\mu \circ T}$ is locally Hölder continuous, which we will need in \ref{5.10}.

### 5.9. The Gibbs property and periodic orbits

We recall the standard construction of an equilibrium state for $(X, \sigma, \varphi)$: for each $w \in \mathcal{L}$, let $x(w) \in \mathcal{L}$

\[ \text{Note that in the setting of [Sar99], some care must be taken to deal with the possibility that we may have } h(m) = \infty \text{ and/or } \int \psi dm = -\infty. \text{ Because } X \text{ has finite topological entropy and } \varphi \text{ is bounded, this is not a problem for us.} \]
$w$ be the point that maximises $S_{|w|} \varphi$; then consider the measures defined by

$$
\nu_n = \frac{1}{\Lambda_n(L, \varphi)} \sum_{w \in L_n} e^{\bar{\varphi}(w)} \delta_{x(w)}, \hspace{1cm} \mu_n = \frac{1}{n} \sum_{k=0}^{n-1} \sigma^k \nu_n,
$$

where $\delta_x$ is the point mass at $x$, and let $\mu$ be a weak* limit point of the sequence $\mu_n$. It is shown in [Wal82, Theorem 9.10] that $h(\mu) + \int \varphi \, d\mu = P(\varphi)$, and by the uniqueness result in the previous section we conclude that $\mu$ is the unique equilibrium state for $(X, \varphi, \mu)$. In particular, we have $\mu = \hat{\pi} \ast p_{s^{-1}} m$.

We can obtain the (non-uniform) Gibbs property for $\mu$ using Proposition 4.2: the following result mimics [CT13, Proposition 5.5 and Lemma 5.6], which are formally mildly weaker.

**Proposition 5.20.** $\mu$ has the Gibbs property (4.12) for $\varphi$ with respect to $F$.

**Proof.** Fix $n \in \mathbb{N}$ and $w \in L_n$. Then for every $m > n$ and $1 \leq k < m - n$, we have

$$
\nu_m(\sigma^{-k}[w]) = \frac{\Lambda_m(H_m(w, k), \varphi)}{\Lambda_m(L, \varphi)}
$$

The result follows using Lemma 4.1, Proposition 4.2, averaging over $k$, and sending $m \to \infty$. □

The fact that $\mu$ is the limiting distribution of $\varphi$-weighted periodic orbits is a consequence of Theorem 7.4 below (this comes from [CT13]). This also gives an alternate proof of uniqueness of $\mu$ and the Gibbs property.

5.10. **Relating $\Sigma^+$ to a tower.**

5.10.1. **Young’s definitions.** We recall the definition of the towers used in [You99], which consist of the following elements.

1. A set $\Delta = \{(z, n) \in \Delta_0 \times \{0, 1, 2, \ldots\} \mid n < R(z)\}$, where $\Delta_0$ is the base of the tower and $R: \Delta_0 \to \mathbb{N}$.
2. A (countable) partition $\Delta_0 = \bigsqcup \Delta_{0,i}$, where $R$ is constant on each $\Delta_{0,i}$, with value $R_i$.
3. A map $F: \Delta \to \Delta$ such that $F(z, \ell) = (z, \ell + 1)$ if $\ell + 1 < R(z)$, and with the property that $F^{R_i}: \Delta_{0,i} \to \Delta$ is a bijection for each $i$.
4. A measure $\bar{m}$ on $\Delta$ such that
   - $F$ carries $\bar{m}|_{\Delta_{\ell,i}}$ to $\bar{m}|_{\Delta_{\ell+1,i}}$ for $\ell < R_i - 1$;
   - the maps $F^{R_i}: \Delta_{0,i} \to \Delta_0$ and their inverses are nonsingular with respect to $\bar{m}$;
   - there are $C > 0$ and $\lambda \in (0, 1)$ such that the Jacobian $J F^{R}$ w.r.t. $\bar{m}$ satisfies
     $$
     \left| \frac{J F^{R}(x)}{J F^{R}(y)} - 1 \right| \leq C \lambda^{s(F^{R}x,F^{R}y)}
     $$

for every $i$ and every $x, y \in \Delta_{0,i}$, where $s(x, y)$ is the smallest $n \geq 0$ such that $(F^{R})^n x$ and $(F^{R})^n y$ lie in distinct $\Delta_{0,i}$’s.
It is shown in [You99, Theorem 1] that the above conditions guarantee the existence of an \( F \)-invariant measure \( \nu \ll \tilde{m} \) which is exact (and hence ergodic and mixing). In our setting we will actually have that \( \tilde{m} \) is itself \( F \)-invariant, and so \( \nu = \tilde{m} \). Thus for us the real interest is in [You99, Theorems 2–4], which establish the rate of convergence to equilibrium, the rate of decay of correlations, and the central limit theorem for \( \nu \).

These are all established using estimates on the rate of decay of the tail of the tower. That is, one defines a function \( \hat{R}: \Delta \to \mathbb{Z} \) by letting \( \hat{R}(x) \) be the smallest integer \( n \geq 0 \) such that \( F^n x \in \Delta_0 \) (the return time to the base of the tower). We will show that in our setting, \( \tilde{m}\{ \hat{R} > n \} \) decays exponentially in \( n \). Note that this is equivalent to showing that \( \tilde{m}\{ x \in \Delta_0 \mid R(x) > n \} \) decays exponentially in \( n \), since

\[
\tilde{m}\{ (x, \ell) \in \Delta \mid \hat{R} > n \} = \sum_{k>n} (k-n)\tilde{m}\{ x \in \Delta_0 \mid R(x) = k \},
\]

(5.18)

\[
\tilde{m}\{ x \in \Delta_0 \mid R(x) > n \} = \sum_{k>n} \tilde{m}\{ x \in \Delta_0 \mid R(x) = k \}.
\]

5.10.2. Connection to \( \Sigma^+ \). We describe the construction of \( (\Delta, F, \tilde{m}) \) isomorphic to \( (\Sigma^+, T, m) \). Let \( \Delta_0 = \hat{R} \mathbb{N} \subset X^+ \); that is, \( \Delta_0 \) consists of all those sequences \( x_0 x_1 x_2 \cdots \in X^+ \) that can be written as a (one-sided) infinite concatenation of elements of \( I \). Given \( z \in \Delta_0 \), let \( R(z) \) be the first return time to \( \Delta_0 \) under \( \sigma \); that is,

\[
R(z) = \min\{ r \in \mathbb{N} \mid \sigma^r(z) \in \Delta_0 \}.
\]

Enumerate the elements of \( I \) as \( w^1, w^2, \ldots, \) and let \( \Delta_{0,i} = [w^i] \cap \Delta_0 \) be the set of all sequences in \( \Delta_0 \) that begin with the word \( w^i \). Then \( \Delta_0 = \bigsqcup \Delta_{0,i} \) and \( R(z) = R_i := |w^i| \) for every \( z \in \Delta_{0,i} \).

Define \( F: \Delta \to \Delta \) by \( F(z, \ell) = (z, \ell + 1) \) when \( \ell + 1 < R(z) \), and \( F(wz, |w| - 1) = (z, 0) \) for every \( w \in I \) and \( z \in \Delta_0 \). Then for every \( i \) the map \( F_{R_i}: \Delta_{0,i} \to \Delta_0 \) is a bijection.

Consider the map \( \tau: \Delta \to \Sigma^+ \) given as follows: every \( z \in \Delta_0 \) is of the form \( z = v^1 v^2 v^3 \cdots \) for some \( v^i \in I \), and to the point \( (z, r) \in \Delta \) we associate the point \( z = \tau(z, r) \in \Sigma^+ \) determined by the condition that

\[
z_0 = \begin{cases} 
a & \text{if } r = 0, \\ (v^1, r) & \text{if } r > 0,
\end{cases}
\]

and that thereafter \( z \) follows the sequence of loops marked by the words \( v^1, v^2, v^3, \ldots \). Let \( \tilde{m} = \tau_*^{-1} m; \) then \( \tau \) is an isomorphism between \( (\Delta, F, \tilde{m}) \) and \( (\Sigma^+, T, m) \). We must check that \( \tilde{m} \) satisfies condition (4) from the previous section.

First we note that if \( (v, \ell) \in A_I \) has \( \ell < |v| - 1 \), then the graph defining \( \Sigma^+ \) has an edge from \( (v, \ell) \) to \( (v, \ell + 1) \), and no other edge terminating at \( (v, \ell + 1) \). Thus \( T \) carries \( m_1[ (v, \ell)] \) to \( m_1[ (v, \ell + 1)] \). Let \( i \) be such that \( v = w^i \), then applying the isomorphism \( \tau \), we see that \( \tau[ (v, \ell)] = \Delta_{\ell,i} \) and so \( F \) carries \( \tilde{m}|_{\Delta_{\ell,i}} \) to \( \tilde{m}|_{\Delta_{\ell+1,i}} \).
It follows from (5.16) that for $z \in \Delta$ we have
\begin{equation}
\log JF(z) = \Phi^+(\tau z) + \psi(\tau z) - \psi(\tau F z) - P(\varphi),
\end{equation}
where $\Phi^+ \in C_\beta(\Sigma^+)$ and $\psi$ is locally Hölder continuous. Moreover, given $y, z \in \Delta_{0,i}$, we see that $\tau(y), \tau(z) \in \Sigma^+$ agree for at least the first $R_i + s(y, z) - 1$ symbols. In particular, for every $0 \leq \ell < R_i$ we have
\[|\Phi^+(\tau F^\ell y) - \Phi^+(\tau F^\ell z)| \leq |\Phi^+|_\beta e^{-\beta (R_i - \ell + s(y, z) - 1)},\]
and similarly,
\[|\psi(\tau y) - \psi(\tau z)| \leq |\psi|_\beta e^{-\beta (R_i + s(y, z) - 1)},\]
\[|\psi(\tau F^{R_i} y) - \psi(\tau F^{R_i} z)| \leq |\psi|_\beta e^{-\beta s(y, z)}.
\]
Together with (5.19), this gives
\[
\left| \log \left( \frac{JF_{R_i}(y)}{JF_{R_i}(z)} \right) \right| \leq \sum_{\ell=0}^{R_i-1} \log JF(\tau F^\ell y) - \log JF(\tau F^\ell z) \\
\leq \left( \sum_{\ell=0}^{R_i-1} |\Phi^+|_\beta e^{-\beta (R_i - \ell + s(y, z) - 1)} \right) \\
\quad + |\psi|_\beta \left( e^{-\beta (R_i + s(y, z) - 1)} + e^{-\beta s(y, z)} \right) \\
\leq e^{-\beta s(y, z)} \left( 2 |\psi|_\beta + |\Phi^+|_\beta \sum_{j \geq 0} e^{-\beta j} \right).
\]
To prove (5.17) from this, it suffices to observe the following fact from basic calculus: for every $K > 1$ there is $C > 0$ such that every $r \in (0, K]$ satisfies the inequality $|r - 1| \leq C |\log r|$. Applying this with $r = \frac{JF_{R_i}(y)}{JF_{R_i}(z)}$ gives (5.17).

5.10.3. The tail of the tower. In order to apply [You99, Theorems 2–4], it remains to show that $\tilde{m}\{ \tilde{R} > n \}$ decays exponentially in $n$. In light of (5.18), it suffices to show that $\tilde{m}\{ x \in \Delta_0 \mid R(x) > n \}$. Using the isomorphism $\tau$ between $\Delta$ and $\Sigma^+$ and the Gibbs property for $m$, we have
\[
\tilde{m}\{ x \in \Delta_0 \mid R(x) > n \} = m\{ z \in \Sigma^+ \mid z_0 = a, z_1 = (w, 1), |w| > n \}
\]
\[
= \sum_{k > n} \sum_{z_0 = a; z_1, \ldots, z_{k-1} \neq a} m[z_0 \cdots z_{k-1}]
\]
\[
\leq Q_1 \sum_{k > n} \sum_{z_0 = a; z_1, \ldots, z_{k-1} \neq a} e^{-kP(\varphi) + S_k \Phi^+(z)}
\]
\[
= Q_1 \sum_{k > n} e^{-kP(\varphi)} Z_k^\alpha(\Phi^+, a).
\]
where we recall the definition of $Z_k^*$ in (5.9). By Lemma 5.15, the quantity in the sum decays exponentially in $k$, and hence $	ilde{m}\{x \in \Delta_0 \mid R(x) > n\}$ decays exponentially in $n$.

5.11. **Exponential decay for $X^+$ and $X$.** It now follows from [You99, Theorem 3] that there is $\theta \in (0, 1)$ such that for every $\Psi_1 \in L^\infty(\Sigma^+, m)$ and $\Psi_2 \in C_\beta(\Sigma^+)$, there is $K = K(\Psi_1, \Psi_2) > 0$ such that for every $n \in \mathbb{N}$ we have

\[
\text{Cor}_n^m(\Psi_1, \Psi_2) = \left| \int (\Psi_1 \circ T^n) \Psi_2 \, dm - \int \Psi_1 \, dm \int \Psi_2 \, dm \right| \leq K \theta^n.
\]

(Note that the Hölder exponent of the observables for which we can get exponential decay is the same as that of the log Jacobian from (5.16), (5.17).)

We use this to prove exponential decay of correlations on $X^+$, then obtain the result for $X$ via a standard approximation argument. Recall that $p: X \to X^+$ is the map that takes a bi-infinite sequence to its forward infinite half, and let $\tilde{\mu} = p_* \mu$. By commutativity of the diagrams in (5.5), we have $\tilde{\mu} = \pi^+_1 m$. Thus for any $\psi_1 \in L^\infty(X^+, \tilde{\mu})$ and $\psi_2 \in C_\beta(X^+)$, we can put $\Psi_i = \psi_i \circ \pi^+$ and use (5.20) to get

\[
\text{Cor}_{n}^{\tilde{\mu}}(\psi_1, \psi_2) = \left| \int (\psi_1 \circ \sigma^n) \psi_2 \, d\tilde{\mu} - \int \psi_1 \, d\tilde{\mu} \int \psi_2 \, d\tilde{\mu} \right| \leq K \theta^n.
\]

This proves exponential decay of correlations for $(X^+, \sigma, \tilde{\mu})$.

Exponential decay of correlations for the two-sided shift $(X, \sigma, \mu)$ follows from the one-sided result via a standard approximation argument, which can be found in [PP90, Proposition 2.4] or in [You98, §4]. Roughly speaking, the idea is to approximate $\psi_1, \psi_2 \in C_\beta(X)$ with $\psi_k^1, \psi_k^2$ that depend only on coordinates $-k, \ldots, k$ (for example, one can obtain $\psi_k^i$ as a conditional average of $\psi_i$ over $[-k, k]$-cylinders). In [PP90] this is done for functions on $X$, while in [You98] it is done at the level of the tower; although the notation is different there, the idea is that one takes a function $\psi: X \to \mathbb{R}$, considers $\Psi = \psi \circ \pi: \Sigma \to \mathbb{R}$, and then approximates $\Psi$ with $\Psi_k$. Ultimately one is able to reduce to the one-sided case; we omit the details here as there is nothing new in our setting.

Note that for the two-sided result one must assume that both test functions are Hölder continuous.

5.12. **Central limit theorem for $X^+$ and $X$.** It follows from [You99, Theorem 4] that for every $\Psi \in C_\beta(\Sigma^+)$ with $\int \Psi \, dm = 0$ such that $\Psi$ is not

\[\text{In either case, one must take care to ensure that a single constant $K$ can be chosen to work for all $k$ in (5.20) and (5.21).}\]
cohomologous to a constant, there is \( \sigma > 0 \) such that
\[
(5.22) \quad \lim_{n \to \infty} \mu \left\{ z \in \Sigma^+ \mid \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi(T^k x) \leq \tau \right\} = \frac{1}{\sigma \psi} \int_{-\infty}^{\tau} e^{-\frac{t^2}{2}} dt
\]
for every \( \tau \in \mathbb{R} \).

Given \( \psi \in C_\beta(X^+) \), we can also consider \( \psi \) as a function \( X \to \mathbb{R} \), and so it suffices to prove the central limit theorem for \( (X, \sigma, \mu) \).

To this end, consider \( \psi \in C_\beta(X) \) with \( \int \psi \, d\mu = 0 \). Let \( \Psi = \psi \circ \pi \in C_\beta(\Sigma) \), and let \( \Psi^+, \psi \) be as in Lemma 5.10. If \( \Psi^+ \) is cohomologous to a constant, then \( \Psi \) is as well, so there is \( \hat{f} \in L^2(\Sigma, \pi_*^{-1} \mu) \) such that \( \Psi = f - f \circ T \).

Define \( g \in L^2(\Sigma, \mu) \) by \( g(x) = f(\pi^{-1} x) \) on \( \pi(\Sigma) \), and \( g = 0 \) elsewhere; then \( \psi = g - g \circ \sigma \) on \( \pi(\Sigma) \), so \( \psi \) is cohomologous to a constant.

Thus if \( \psi \) is not cohomologous to a constant, then \( \Psi \) is not either, so \( \Psi - \Psi^+ = 0 \) holds for some \( \sigma > 0 \). By (5.7) we have
\[
(5.23) \quad \left| \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi^+(T^k \hat{p}(z)) - \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi(T^k z) \right| \leq 2 \| u \| \frac{\sqrt{n}}{\sqrt{n}}.
\]

Write \( G_m^\mu(\tau) = \mu \{ x \in X \mid \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi(\sigma^k x) \leq \tau \} \), and similarly for \( G_m^\psi(\tau) \) (summing over \( T \)-orbits on \( \Sigma^+ \)). Then (5.23) gives
\[
G_m^\mu \left( \tau - \frac{2 \| u \| \sqrt{n}}{\sqrt{n}} \right) \leq G_m^\mu(\tau) \leq G_m^\mu \left( \tau + \frac{2 \| u \| \sqrt{n}}{\sqrt{n}} \right),
\]
and it follows that \( G_m^\mu(\tau) \) converges to the right-hand side of (5.22) (this uses continuity of that expression).

### 6. Proof of Theorem 3.2

#### 6.1. Plan of the proof

Proposition 3.7 shows how to produce \( F \) satisfying \( [I_0] \) and is proved in \( \S 6.2 \). Condition \( [II_0] \) comes from the following result.

**Proposition 6.1.** Let \( X \) be a shift space on a finite alphabet and \( \varphi \in C_\beta(X) \) for some \( \beta > 0 \). Suppose \( \mathcal{G} \subset \mathcal{L}(X) \) satisfies \( [I] \), \( [II] \), and \( [III] \), and let \( (r, c, s) \) be any synchronising triple for \( \mathcal{G} \). Then \( \mathcal{F} = \mathcal{F}^{r,c,s} \) satisfies \( [II] \); there are \( \mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L} \) such that \( \mathcal{E} := \mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s) \) has \( P(\mathcal{E}, \varphi) < P(\varphi) \).

The proof of Proposition 6.1 constitutes the bulk of the proof of Theorem 3.2, and we will defer it to the end of this section, in \( \S 6.4 \). First we prove Proposition 3.7 (in \( \S 6.2 \)), and then in \( \S 6.3 \) we establish Condition \( [III^*] \) which requires us to consider two cases.

**Definition 6.2.** Say that \( \mathcal{G} \subset \mathcal{L} \) is periodic if there is a periodic sequence \( x \in X \) such that every \( w \in \mathcal{G} \) appears somewhere in \( x \).

**Proposition 6.3.** If \( (X, \varphi) \) is such that there is a periodic \( \mathcal{G} \subset \mathcal{L} \) satisfying \( [I] \) and \( [II] \), then there are \( \mathcal{F}, \mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L} \) satisfying \( [I_0] \), \( [II_0] \), and \( [III^*] \).

Theorem 3.2 follows directly from Propositions 3.7, 6.1, 6.3 and the following result.
Proposition 6.4. Suppose that $G \subseteq L$ satisfies [I] and is not periodic. Then $G$ has a synchronising triple $(r, c, s)$ such that

$$[rcs] \cap \sigma^{-k}[rcs] = \emptyset \text{ for every } 1 \leq k \leq \max\{|rc|, |cs|\}.$$  

Moreover, if $G$ satisfies [III] and $(r, c, s)$ is any synchronising triple satisfying (6.1), then $F^{r,c,s}$ satisfies [III*].

6.2. Producing a collection of words that can be freely concatenated. In this section we prove Proposition 3.7. The following lemma mimics the proof from [Ber88] that specification implies synchronised, and establishes the existence of a synchronising triple for $G$.

Lemma 6.5. Suppose $G$ satisfies [I]. Then given $v, w \in G$, there are $q \in vL \cap G$, $p \in Lw \cap G$, and $c \in L_{\leq \tau}$ such that $pcq \in G$, and given any $u, u' \in L$ with $up, qu' \in G$, we have $upcu' \in G$.

Proof of Lemma 6.5. Let $C(w, v)$ be the set of connecting words $c \in L_{\leq \tau}$ such that $wcv \in G$. This is non-empty by [I]. Let $v_0 = v$ and $w_0 = w$, then define $v_n, w_n$ recursively as follows (see Figure 6.1):

- $v_{n+1}^{n+1} \in G \cap v^nL$
- $w_{n+1}^{n+1} \in G \cap Lw^n$
- $C(w_{n+1}^{n+1}, v_{n+1}^{n+1}) \neq C(w^n, v^n)$

Each $C(w^n, v^n)$ is finite, non-empty, and contained in $C(w^{n-1}, v^{n-1})$. Thus the process terminates for some finite value of $n$. Let $q = v^n$, $p = w^n$, and pick any $c \in C(p, q)$. Then $pcq \in G$. Moreover, by the construction of $v^n, w^n$ we see that for any $u, u' \in L$ with $up, qu' \in G$, we have $C(up, qu') = C(p, q) \ni c$, hence $upcu' \in G$. □

![Figure 6.1. Producing a synchronising triple.](image)

The triple $(p, c, q)$ produced in Lemma 6.5 is a synchronising triple for $G$. Now we must show that writing $B = Lp \cap qL \cap G$, the collection $cB$ satisfies [I0]. This is a consequence of the following lemma.

Lemma 6.6. If $u^1, \ldots, u^n \in B$, then $u^1cu^2c\cdots cu^n \in B$.  


Proof. The case $n = 1$ is immediate. If the statement holds for $n$, then for any $u^1, \ldots, u^{n+1} \in B$ we have $u^1 \cdots cu^n = v \in G$. By the definition of $B$ we have $u^1 \cdots cu^n = vp \in G$ for some $v \in \mathcal{L}$, and $w^1 = qw \in G$ for some $w \in \mathcal{L}$, so

$$u^1 \cdots cu^n w^{n+1} = vpcqw.$$  

By Lemma 6.5 this is contained in $G$. Moreover, this word begins with the word $q$ (since $w^1$ does) and ends with the word $p$ (since $w^{n+1}$ does), so it is an element of $B$. This establishes the claim for $n + 1$, and the result follows by induction.

Writing $F = cB$, we see that for any $v, w \in cB$ there are $v', w' \in B$ such that $v = cv', w = cw'$; Lemma 6.6 gives $v'cw' \in B$, hence $vw = cv'cw' \in cB = F$. Thus $F$ satisfies $[I]$. It remains to show that a measure $\mu$ has the Gibbs property for $\varphi$ on $F$ if and only if it has the Gibbs property on $G$. Note that the upper bound in (4.12) is required to hold for all $w$, so it suffices to check the lower bound.

Suppose $\mu$ is Gibbs on $G$ with constant $Q_1$. Then as in Remark 3.8 we have $rw \in G$ for each $w \in F^{p,c,q}$, and in particular

$$\mu[w] \geq \mu[rw] \geq Q_1^{-1} e^{-|rw|P(\varphi) + \hat{\varphi}(rw)} \geq Q_1^{-1} e^{-|w|P(\varphi) + \hat{\varphi}(w)} e^{-|r|P(\varphi) + \|\varphi\|},$$

so $\mu$ is Gibbs on $F^{p,c,q}$. Conversely, if $\mu$ is Gibbs on $F^{p,c,q}$ then for each $w \in G$ there are $u, v \in \mathcal{L}_{\leq \tau}$ such that $cquwp \in c(q\mathcal{L} \cap \mathcal{L}p \cap G) = F^{p,c,q}$, and hence

$$\mu[w] \geq \mu[cquwp] \geq Q_1^{-1} e^{-|cquwp|P(\varphi) + \hat{\varphi}(cquwp)} \geq Q_1^{-1} e^{-|w|P(\varphi) + \hat{\varphi}(w)} e^{-|r|P(\varphi) + \|\varphi\|},$$

so $\mu$ is Gibbs on $G$. This completes the proof of Proposition 3.7.

6.3. Establishing Condition $[III^+]$.

6.3.1. The periodic case. We prove Proposition 6.3. Suppose $G$ is periodic and satisfies $[I]$ and $[II]$. Let $x \in X$ be periodic such that every $w \in G$ appears somewhere in $x$. Let $d \in \mathbb{N}$ be the least period of $x$ and let $F = \{x_{[1,k]} \mid k \in \mathbb{N}\}$. Then $F$ has $[I_0]$ and $[III^+]$ (the second assertion uses the fact that $d$ is minimal). Let

$$E^p = C^p \mathcal{L}_{\leq d} \cap \mathcal{L} = \{uvv \in \mathcal{L} \mid w \in C^p, |v| \leq d\},$$

$$E^s = \mathcal{L}_{\leq d} C^s \cap \mathcal{L} = \{vw \in \mathcal{L} \mid |v| \leq d, w \in C^s\}.$$  

Then given any $u^p \in C^p, v \in G, u^s \in C^s$, we note that there are $i \in [1, p]$ and $j \in ([|v| - p, |v|])$ such that $v_{[i,j]} \in F$, and hence

$$uw^pu^s = (w^pu_{[1,i],v_{[i,j]},v_{[j+1,j+p-1]}^s})u_{[j+1,j+p-1]}^s \in E^p F E^s.$$  

Together with the observation that $P(E^p, \varphi) = P(C^p, \varphi)$ and $P(E^s, \varphi) = P(C^s, \varphi)$, this establishes $[III]$ for $E^p, F, E^s$.  

SPECIFICATION AND TOWERS IN SHIFT SPACES 41
6.3.2. A synchronising triple with no long overlaps. Now we prove Proposition 6.4. We assume that \( \mathcal{G} \subset \mathcal{L} \) satisfies (I) and is not periodic. First we prove the existence of a synchronising triple satisfying (6.1); then we prove that such a triple yields a collection \( \mathcal{F} \) satisfying (III).

Condition (6.1) can be thought of as forbidding ‘long overlaps’ of \( rcs \) with itself (see Figure 6.2 below). To produce a synchronising triple with no long overlaps, we start by letting \((p, c, q)\) be any synchronising triple for \( \mathcal{G} \). Then we will take \( r = vup \) and \( s = quw \), where \( v, w \in \mathcal{G} \) will be chosen to satisfy certain conditions given below, and \( u, u' \in \mathcal{L} \leq \tau \) come from (I). Note that then \((r, c, s)\) is once again a synchronising triple for \( \mathcal{G} \).

Before choosing \( v, w \), we suppose \( r, s \) have the form just stated, and that \( k > 0 \) is such that \( [rcs] \sigma^{-k} [rcs] \not= \emptyset \). Let \( x \) be an element of the intersection, then we have

\[
(6.2) \quad x_{[0, |rcs|]} = x_{[k, k+|rcs|]} = rcs = vupcqu'w.
\]

We want to choose \( v, w \) such that (6.2) forces \( k \geq \max(|vuler|, |cqu'w|) \). We will choose \( v \) to be (much) longer than \( p, q \), and \( w \) to be (much) longer than \( v \). Figure 6.2 illustrates the three possible ranges of \( k \) that we must deal with:

1. \( 1 \leq k \leq |vupcqu'| \);
2. \( |vupcqu'| < k < |upcqu'w| = |rcs| - |v| \);
3. \( k \geq |upcqu'w| \geq \max(|vuler|, |cqu'w|) \).

![Figure 6.2](image)

**Figure 6.2.** Cases 1 and 2 will be forbidden by our choice of \( v, w \). Case 3 is permissible.

Note that if (6.2) holds for some \( k < |w| \), then \( w \) has \( w_{[1, |w| - k]} = w_{[k+1, |w|]} \), so that

\[
(6.3) \quad w_{i+k} = w_i \text{ for every } 1 \leq i \leq |w| - k.
\]

Say that a word \( w \) satisfying (6.3) is \( k \)-periodic. Roughly speaking, the idea is that when \( k \ll n \), most words in \( \mathcal{L}_n \) are not \( k \)-periodic, and in particular there is \( \alpha > 0 \) such that there are arbitrarily long words \( w \in \mathcal{G} \) that are not \( k \)-periodic for any \( k \leq \alpha |w| \). Choosing such a \( w \) will force \( k \geq \alpha |w| \) whenever (6.2) holds, and choosing the lengths of \( v, w \) appropriately will guarantee that this rules out Case 1.

Then it will remain to choose \( v \in \mathcal{G} \) to rule out Case 2. We will choose \( v \) to be short enough that \( |vupcqu'| \leq \alpha |w| \) (this is necessary for the argument...
in the previous paragraph for Case 1). Then any $k$ satisfying (6.2) has $i := k - |vupcqu'| > 0$, and it follows from (6.2) that we have either $w_{[i,i+|v|]} = v$ or $i > |w| - |v|$. We will choose $v$ to be a word that does not appear as a subword of $w$, which will force the inequality $i > |w| - |v|$, so that $k = i + |v| + |upcqu'w| ≥ |upcqu'w|$, which will suffice, since $|equw| ≥ |vupc|$ by our choice of $v$.

With the above plan in mind, we now carry out the details. We need to guarantee that $G_n$ contains enough words that we can throw out the ‘bad’ ones and still have something left. This requires the non-periodicity condition.

**Lemma 6.7.** Suppose $G$ has $[I]$ and is not periodic, and let $c$ be the connecting word from some synchronising triple for $G$. Then there is $ℓ ∈ \mathbb{N}$ such that $\#G_{ℓm-|c|} ≥ 2^m$ for all $m ∈ \mathbb{N}$.

**Remark 6.8.** A very similar result is proved in [CT12, Proposition 2.4] (see §6.3 there). Our hypotheses here are weaker (the gluing time in specification is allowed to vary) and the results in the previous section allow us to give a simpler proof. Note that the conclusion is strictly stronger than the inequality $h(G) > 0$.

**Proof of Lemma 6.7.** First note that if $G$ is not periodic, then neither is $cB$.

Indeed, if $x ∈ A^\mathbb{N}$ was periodic and contained every word in $cB$ as a subword, then it would contain every word in $G$ as well, by Remark 3.8.

Now we claim that there are $v, w ∈ cB$ such that $|w| ≥ |v|$ and $w \notin v\mathcal{L}$. Suppose this was false; we will show that in this case $cB$ must be periodic. Indeed, define $x ∈ A^\mathbb{N}$ by $x_i = w_i$ for some $w ∈ cB$ with $|w| ≥ i$. By the assumption this is well-defined since $v_i = w_i$ whenever $v, w ∈ cB$ and $i ≤ \min(|v|, |w|)$. Moreover, given $v ∈ cB$ we have $vv \cdots v ∈ cB$ for arbitrarily long concatenations of $v$ with itself, so $x_{i+k|v|} = v_i = x_i$ for any $k ≥ 0$ and $1 ≤ i ≤ |v|$. It follows that $x$ is periodic, and any $w ∈ cB$ appears as a prefix of $vv \cdots v$, hence as a prefix of $x$, by the assumption.

The previous paragraph shows that non-periodicity of $cB$ implies existence of $v, w ∈ cB$ such that $v_i ≠ w_i$ for some $i ≤ \min(|v|, |w|)$. Let $ℓ = |v| \cdot |w|$. Let $u^1 = vv \cdots v$ and $u^2 = uw \cdots w$, where we concatenate $|w|$ copies of $v$ and $|v|$ copies of $w$ so that $|u^1| = |u^2| = ℓ$. By construction of $v, w$ we have $u^1 ≠ u^2$. Now for every $m ∈ \mathbb{N}$ and $y ∈ \{1,2\}^m$, that is, every finite sequence of 1s and 2s, we have $u^{y_1} \cdots u^{y_m} ∈ (cB)_{mℓ}$. Moreover, different choices of $y$ yield different words in $(cB)_{mℓ}$, so that $\#(cB)_{mℓ} ≥ 2^m$. Since $B ⊂ G$ we get $\#G_{mℓ-|c|} ≥ 2^m$.

Now we choose the lengths of the words $v, w$ described before Lemma 6.7. Recall that $A$ is the alphabet of $X$, and let $ℓ$ be as in Lemma 6.7. Choose $α > 0$ such that $αℓ \log(\#A) < \log 2$. Note that we can choose arbitrarily large $m, n ∈ \mathbb{N}$ such that

\[
\frac{ℓ}{α} m + \frac{2\tau + |pcq|}{α} ≤ ℓn - |c| < 2^m.
\]

(6.4)
We claim that for sufficiently large \( m, n \) satisfying (6.4), there exist \( v \in G_{\ell m - |c|} \) and \( w \in G_{\ell n - |c|} \) such that \( v \) is not a subword of \( w \), and \( w \) is not \( k \)-periodic for any \( 1 \leq k \leq \alpha |w| \).

To this end, we consider the collection
\[
\mathcal{P}^a = \{w \in \mathcal{L} \mid w \text{ is } k\text{-periodic for some } 1 \leq k \leq \alpha |w|\}.
\]

If \( w \) is \( k \)-periodic then it is determined by its first \( k \) entries, so we can estimate the cardinality of \( \mathcal{P}^a_N \) by
\[
(6.5) \quad \#\mathcal{P}^a_N \leq \sum_{k=1}^{\lfloor \alpha N \rfloor} (#A)^k \leq (#A)^{\alpha N} \sum_{j=0}^{\infty} (#A)^{-j} = e^{\alpha \log (#A) N \left( \frac{#A}{#A - 1} \right)}.
\]

Write \( \gamma := (\log 2)/\ell - \alpha \log #A \) and note that \( \gamma > 0 \) by the choice of \( \alpha \).

When \( N = \ell n - |c| \) for some \( n \in \mathbb{N} \), Lemma 6.7 gives \( \#G_N \geq 2^n \geq 2^{N/\ell} \), and so
\[
(6.6) \quad \frac{\#G_N}{\#\mathcal{P}^a_N} \geq \left( \frac{#A - 1}{#A} \right) e^{(\log 2) \frac{N}{\ell} - \alpha \log (#A) N} \geq \left( \frac{#A - 1}{#A} \right) e^{\gamma N}.
\]

For \( n \) sufficiently large this gives \( \#G_N > \#\mathcal{P}^a_N \), so there is \( w \in G_N = G_{|c|} \) that is not \( k \)-periodic for any \( 1 \leq k \leq \alpha |w| \). To put it another way: for every sufficiently large \( n \) there is \( w \in G_{|c|} \) such that
\[
(6.7) \quad \text{for every } 1 \leq k \leq \alpha |w| \text{ there is } 1 \leq j \leq |w| - k \text{ with } w_{k+j} \neq w_k.
\]

Now let \( m, n \) be such that (6.4) is satisfied and (6.7) holds for some \( w \in G_{|c|} \). Note that \( w \) contains at most \( |w| \) subwords of length \( \ell m - |c| \), while \( \#G_{\ell m - |c|} \geq 2^m > |w| \) by (6.4). Thus there is \( v \in G_{\ell m - |c|} \) such that \( w[i, i+|v|] \neq v \) for any \( 1 \leq i \leq |w| - |v| \); that is, \( v \) does not appear as a subword of \( w \).

By (I) there are \( u, u' \in \mathcal{L}_{\leq \tau} \) such that \( vup, qu'w \in G \). Note that by the first inequality in (6.4) we have
\[
|vupcu'| \leq \ell m + 2\tau + |pcq| \leq \alpha |w|.
\]

It follows that \( w \) is not \( k \)-periodic for any \( 1 \leq k \leq |vupcu'| \). Now as in the discussion prior to Lemma 6.7 we see that any \( k \) such that (6.2) holds must fall into one of the three classes described there. The first case described there cannot occur because of the aperiodicity of \( w \). The second case cannot occur because \( v \) is not a subword of \( w \). Thus only the third case can occur, which proves (6.1).

6.3.3. Absence of short overlaps implies \( \textbf{[III]} \). Let \( (r, c, s) \) be a synchronising triple for \( G \) satisfying (6.1), and suppose that \( G \) satisfies \( \textbf{[III]} \) we show that
\[
\mathcal{F} := \mathcal{F}^{r, c, s} = cB^{r, s} = c(sL \cap Lr \cap G)
\]
satisfies \( \textbf{[III]}^* \). Note that \( \mathcal{F} \) satisfies \( \textbf{[I]} \) by Proposition 3.7.
Suppose \( x \in X \) and \( i \leq j < k \leq \ell \in \mathbb{Z} \) are such that \( x_{(i,k)}, x_{(j,\ell)} \in \mathcal{F} \) and there are \( a < j \) and \( b > k \) such that \( x_{(a,j)}, x_{(k,b)} \in \mathcal{F} \) (see Figure 6.3). We must show that \( x_{(j,k)} \in \mathcal{F} \). Let \( j' = j - |r| \) and \( k' = k - |r| \); then we have

\[
x_{[j', j'+ |rcs|]} = x_{[k', k'+ |rcs|]} = rcs,
\]

so by (6.1) we have \( k - j = k' - j' \geq \max(|rc|, |cs|) \). Thus \( x_{(j,j'+|c|)} = c \) and \( x_{(j'+|c|, k)} \in s \mathcal{L} \cap \mathcal{L}r \).

It remains only to show that \( x_{(j,j'+|c|, k)} \in \mathcal{G} \). For this we observe that \( x_{(i,k)} \in \mathcal{F} \) implies \( x_{(i+|c|, k)} \in \mathcal{G} \), and \( x_{(j,\ell)} \in \mathcal{F} \) implies \( x_{(j+|c|, \ell)} \in \mathcal{G} \). Note that \( i + |c| \leq j + |c| < k \leq \ell \), so by (III) we have \( x_{(j,j'+|c|, k)} \in \mathcal{G} \). It follows that \( x_{(j,k)} \in \mathcal{F} \), which establishes (III*). This completes the proof of Proposition 6.1.

### 6.4. Construction of \( \mathcal{E}^p \) and \( \mathcal{E}^s \) – uniform specification

In this section and the next (§6.5) we prove Proposition 6.1. We first give the proof in the case when (I) is satisfied with \( \mathcal{G} = \mathcal{L} \) (that is, \((X, \sigma)\) satisfies the classical specification property, and (II) (III) are automatic with \( \mathcal{C}^p = \mathcal{C}^s = \emptyset \)); then in §6.5 we deal with the more general case when there is \( \mathcal{G} \subset \mathcal{L} \) satisfying (I) (II) (III) Formally speaking the arguments in this section are a subset of the arguments in the next section, but presenting them first under the stronger assumption that \( \mathcal{L} \) has specification helps to clarify the presentation and motivate the extra steps that must be taken in the general case.

#### 6.4.1. Definition of \( \mathcal{E}^p, \mathcal{E}^s \)

Let \((r, c, s)\) be any synchronising triple for \( \mathcal{L} \), and let \( \mathcal{F} = \mathcal{F}^r,c,s = cB^r,c,s = c(s \mathcal{L} \cap \mathcal{L}r) \). Before defining \( \mathcal{E}^p, \mathcal{E}^s \), we consider how a word \( w \in \mathcal{L} \) can be decomposed as \( w = u^p v u^s \) with \( v \in \mathcal{F} \); to get such a decomposition, we need \( v = w_{(i,j)} \) where \( 1 \leq i < j \leq |w| \) are such that

\[
\begin{align*}
(6.8) & \quad w_{[i, |w|]} \in cs \mathcal{L}, \\
(6.9) & \quad w_{[1, j]} \in \mathcal{L}r.
\end{align*}
\]

Later, when \( \mathcal{G} \neq \mathcal{L} \), we will also need to require that \( w_{(i,j)} \in \mathcal{G} \); this will complicate matters somewhat.

Heuristically, since the words \( u^p, u^s \) should come from ‘small’ collections, it is reasonable to try to choose \( v \) to be as long as possible. Thus we should let \( i \) be minimal such that (6.8) is satisfied, and \( j \) maximal such that (6.9) is satisfied. Then \( u^p \) has the property that it does not contain the word \( cs \); indeed, if \( u^p_{(a,b)} = cs \), then \( a < i \) and \( a \) satisfies (6.8), contradicting our
choice of $i$. Similarly, $u^s$ does not contain the word $r$. This suggests the following natural choice for $\mathcal{E}^p$ and $\mathcal{E}^s$:

\begin{align*}
\mathcal{E}^p &= \{ u \in \mathcal{L} \mid u_{[i,u]} \notin cs\mathcal{L} \text{ for every } 1 \leq i \leq |u| \}, \\
\mathcal{E}^s &= \{ u \in \mathcal{L} \mid u_{[1,j]} \notin \mathcal{L} r \text{ for every } 1 \leq j \leq |u| \}.
\end{align*}

(6.10)

Now given $w \in \mathcal{L}_n$, we write $m = \lfloor \frac{n}{2} \rfloor$, and one of the following three things happens.

- There are $i \leq m - |cs|$ satisfying (6.8) and $j > m + |r|$ satisfying (6.9).
- There is no $i < m - |cs|$ satisfying (6.8), so $w_{[1,m]} \in \mathcal{E}^p$.
- There is no $j > m + |r|$ satisfying (6.9), so $w_{[m,w]} \in \mathcal{E}^s$.

Consider the collection

\begin{equation}
\hat{\mathcal{E}} = \{ w \in \mathcal{L} \mid w_{[1,m]} \in \mathcal{E}^p \text{ or } w_{[m,w]} \in \mathcal{E}^s \text{ for } m = \lfloor \frac{n}{2} \rfloor \}.
\end{equation}

(6.11)

The discussion above shows that $\mathcal{L} \setminus (\mathcal{E}^p \mathcal{F} \mathcal{E}^s) \subset \hat{\mathcal{E}}$. Moreover, (4.3) gives

\begin{equation}
\Lambda_n(\hat{\mathcal{E}}, \varphi) \leq \Lambda_{n/2}(\mathcal{E}^p, \varphi)\Lambda_{n/2}(\mathcal{E}^s, \varphi) + \Lambda_{n/2}(\mathcal{L}, \varphi)\Lambda_{n/2}(\mathcal{E}^s, \varphi)
\end{equation}

and hence

\begin{equation}
P(\hat{\mathcal{E}}, \varphi) \leq \frac{1}{2}(P(\mathcal{E}^p \cup \mathcal{E}^s, \varphi) + P(\varphi)).
\end{equation}

(6.12)

Thus to establish \(\Pi\) for $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s$, it suffices to prove that $P(\mathcal{E}^p, \varphi) < P(\varphi)$ and $P(\mathcal{E}^s, \varphi) < P(\varphi)$.

6.4.2. $P(\mathcal{E}^p \cup \mathcal{E}^s, \varphi) < P(\varphi)$. We give the proof for $\mathcal{E}^s$; the proof for $\mathcal{E}^p$ is similar. Given $1 \leq k \leq n \in \mathbb{N}$, let $\mathcal{A}^k_n = \{ w \in \mathcal{L}_n \mid w_{[1,j]} \notin \mathcal{L} r \text{ for every } 1 \leq j \leq k \}$ be the set of $k$-avoiding words of length $n$; that is, words that avoid the word $r$ for the first $k$ times it could occur\(^{19}\). Note that $\mathcal{A}^0_n = \mathcal{L}_n$ and $\mathcal{A}^n_n = \mathcal{E}^s_n$; our goal is to control $\Lambda_n(\mathcal{E}^s, \varphi)$ by estimating $\Lambda_n(\mathcal{A}^k_n \setminus \mathcal{A}^{k+\hat{\tau}}_n, \varphi)$ for a suitable $\hat{\tau} \in \mathbb{N}$.

To this end, we consider the set

\begin{equation}
\mathcal{Z}^k = \{ u \in \mathcal{L}_k \mid u_{[1,j]} \notin \mathcal{L} r \text{ for every } 1 \leq j \leq k \}.
\end{equation}

(6.13)

(Of course this is just $\mathcal{E}^s_k$, but it is helpful to introduce the new notation in order to set up the structure for the general proof in the next section.) Recall from (4.11) that $\mathcal{H}_n(u, i) = \{ w \in \mathcal{L}_n \mid w_{[i,i+|u|]} = u \}$. Then given $1 \leq k \leq n \in \mathbb{N}$, we have

\begin{equation}
\mathcal{A}^k_n = \bigsqcup_{u \in \mathcal{Z}^k} \mathcal{H}_n(u, 1).
\end{equation}

---

\(^{19}\)Technically there are only $k - |r|$ chances, but we avoid unnecessary complications in notation.
Given any $u \in Z^k$, by specification there is $q \in L_{<r}$ such that $v := v(u) = uqr \in L$. Note that $|v| \leq |u| + \tau + |r| = j + \tau + |r|$, and write $\hat{\tau} = \tau + |r|$; then for every $n > |v|$ we have

$$
\mathcal{H}_n(v, 1) \subset A^k(u) \setminus A^{k+\hat{\tau}}.
$$

Using the Gibbs bounds from Proposition 4.2 and the fact that the union in (6.14) is disjoint, we have

$$
\Lambda_n(A^k_n \setminus A^{k+\hat{\tau}}_n, \varphi) = \sum_{u \in Z^k} \Lambda_n(\mathcal{H}_n(u, 1), \varphi) \geq \sum_{u \in Z^k} Q_4^{-1} e^{(n-k-\hat{\tau})P(\varphi) + \hat{\varphi}(v)}
$$

$$
\geq \sum_{u \in Z^k} Q_4^{-1} e^{(n-k-\hat{\tau})P(\varphi) e^{-\hat{\varphi}(u)} - \hat{\varphi}(u) \| \varphi \|}
\geq \sum_{u \in Z^k} Q_4^{-2} \Lambda_n(\mathcal{H}_n(u, 1)) e^{-\hat{\varphi}(P(\varphi) + \| \varphi \|)}
= Q_4^{-2} e^{-\hat{\varphi}(P(\varphi) + \| \varphi \|)} \Lambda_n(A^k_n, \varphi).
$$

Taking $\gamma = Q_4^{-2} e^{-\hat{\varphi}(P(\varphi) + \| \varphi \|)}$, this can be written as

$$
\Lambda_n(A^k_n \setminus A^{k+\hat{\tau}}_n, \varphi) \geq \gamma \Lambda_n(A^k_n, \varphi).
$$

In particular, we get

$$
\Lambda_n(A^{k+\hat{\tau}}_n, \varphi) \leq (1 - \gamma) \Lambda_n(A^k_n, \varphi).
$$

Using the fact that $A^0_n = L_n$ and $A^k_n = E^s_n$, we can iterate (6.18) to get

$$
\Lambda_n(E^s_n, \varphi) = \Lambda_n(A^0_n, \varphi) \leq (1 - \gamma)^{\lfloor n/\hat{\tau} \rfloor} \Lambda_n(L_n, \varphi),
$$

$$
\frac{1}{n} \log \Lambda_n(E^s_n, \varphi) \leq \frac{1}{n} \left\lfloor \frac{n}{\hat{\tau}} \right\rfloor \log(1 - \gamma) + \frac{1}{n} \log \Lambda_n(L_n, \varphi).
$$

Sending $n \to \infty$ gives

$$
P(E^s, \varphi) \leq P(\varphi) + (\hat{\tau})^{-1} \log(1 - \gamma) < P(\varphi).
$$

A similar argument works for $E^p$, replacing $A^k_n$ with

$$
A^k_n = \{ w \in L_n \mid w[j-|c|, j] \neq cs \text{ for every } |w| - k < j \leq |w| \}
$$

and $Z^k$ with

$$
Z^k = \{ u \in L_k \mid u[i,i+|cs|] \neq cs \text{ for every } 1 \leq i \leq |u| \}.
$$

Given $u \in Z^k$ one puts $v = v(u) = csqu$ for some $q \in L_{\leq r}$, obtaining $\mathcal{H}(v, n-|v|) \subset \mathcal{H}_n(u, n - k + 1) \setminus A^{k+\hat{\tau}}$, similarly to (6.15), where this time we must take $\hat{\tau} = \tau + |cs|$. Then the same computation as in (6.16) gives $\Lambda_n(A^{k+\hat{\tau}}_n, \varphi) \leq (1 - \gamma) \Lambda_n(A^k_n, \varphi)$, and we can iterate to get $P(E^p, \varphi) < P(\varphi)$.

6.5. Construction of $E^p$ and $E^s$ – non-uniform specification.
6.5.1. General plan. Now we prove Proposition 6.1 in its full generality. Suppose \( G \) satisfies [I], [II], and [III] and \((r, c, s)\) is any synchronising triple for \( G \). Let \( F = F^{r,c,s} \); we must produce \( E^p, E^s \subset \mathcal{L} \) such that

\[
P(\mathcal{L} \setminus E^p \mathcal{F} \mathcal{E}^s, \varphi) < P(\varphi),
\]
\[
P(E^p \cup E^s, \varphi) < P(\varphi).
\]

We follow the same idea as in §6.4, but because we only have the specification property on \( G \), we are forced to do some extra work. In the previous section, the general plan for the pressure estimate on \( E^s \) was as follows.

- \( \mathcal{A}_n^k \) is the set of words of length \( n \) that have had at least \( k \) opportunities for \( r \) from the synchronising triple to appear, but have always avoided it; in particular, when \( k \) is large enough we have \( \mathcal{A}_n^k \supset \mathcal{E}_n^s \).
- Each \( \mathcal{A}_n^k \) can be decomposed as a disjoint union of collections \( \mathcal{H}_n(u, i) \) according to the shortest word \( u \) during which those \( k \) opportunities occur.
- Writing \( \mathcal{Z}_n^k(n) \) for the set of such \( u \), we can use specification to extend each \( u \) to a word \( v \) in which the synchronising triple does appear, obtaining \( \mathcal{H}_n(v, i) \subset \mathcal{H}_n(u, i) \setminus \mathcal{A}_n^{k+1} \) as in (6.15). The partition sum of \( \mathcal{H}_n(v, i) \) can be estimated using Proposition 4.2.
- Summing over all \( u \in \mathcal{Z}_n^k(n) \) gives the uniform decay bound in (6.18), which can be iterated to give the desired pressure estimate.

In the general setting, the third step described above can only be carried out when \( u \in G \). Thus it becomes important to estimate how many subwords of \( w \) are in \( G \). We accomplish this by using a weaker version of the restricting sets from Definition 5.16. We now describe these sets, and then explain how they are used to produce the desired collections \( E^p, E^s \). The definition of \( E^p, E^s \) will be deferred for some time while we build up some terminology and preliminary results.

6.5.2. Restricting sets of indices.

Definition 6.9. Given \( w \in \mathcal{L} \), say that \([j, k] \subset [1, |w|]\) is a wall in \( w \) if for every \( 1 \leq i \leq j \) and \( k \leq \ell \leq |w| \) we have \( w_{[i, \ell]} \notin \mathcal{C} \). (Note that this is the same as asking the two-element set \( \{j, k\} \) to be \( \mathcal{C} \)-restricting in the sense of Definition 5.16.)

We say that \([j, k] \subset [1, |w|]\) is right \( \mathcal{C} \)-restricting in \( w \) if \( w_{[j, \ell]} \notin \mathcal{C} \) for every \( \ell \geq k \), and left \( \mathcal{C} \)-restricting if \( w_{[i, k]} \notin \mathcal{C} \) for every \( i \leq j \).

- (a) A wall
- (b) Right \( \mathcal{C} \)-restricting
- (c) Left \( \mathcal{C} \)-restricting

Figure 6.4. Walls and half-restricting intervals.
Note that walls are both right and left $\mathcal{C}$-restricting (the converse is not true). The different restricting conditions are illustrated in Figure 6.4. The key property of these intervals is the following.

**Lemma 6.10.** If $1 \leq i < j \leq k < \ell \leq n$ are such that $[i, j]$ is left $\mathcal{C}$-restricting and $(k, \ell]$ is right $\mathcal{C}$-restricting for $w \in \mathcal{L}_n$, then there are $a \in [i, j)$ and $b \in (k, \ell]$ such that $w_{[a, b)} \in \mathcal{G}$.

**Proof.** Since $\mathcal{C} \supset \mathcal{L} \setminus \mathcal{C} \mathcal{G}^\ast$, we have $w_{[i, j]} \in \mathcal{C} \mathcal{G}^\ast$, so there are $i \leq a < b \leq \ell$ such that $w_{[i, a)} \in \mathcal{C}^\ast$, $w_{[a, b)} \in \mathcal{G}$, and $w_{[b, \ell]} \in \mathcal{C}^\ast$. Since $[i, j]$ is left $\mathcal{C}$-restricting we have $a < j$, and since $(k, \ell]$ is right $\mathcal{C}$-restricting we have $k < b$.

Given $0 \leq \alpha < \beta \leq 1$, we will consider the collection of ‘walled words’
\[ \mathcal{W}(\alpha, \beta) = \{ w \in \mathcal{L} \mid |\alpha w|, \beta w| \text{ is a wall in } w \}. \]
We will also consider for each $\delta > 0$ the following collection of words with ‘restricted prefixes’:
\[ \mathcal{R}_\delta^\ast(\alpha, \beta) = \{ w \in \mathcal{L} \mid \text{ there are } \alpha |w| \leq k_0 < k_1 < \cdots < k_\ell \leq \beta |w| \text{ with } \ell \geq \delta |w| \text{ such that every } [k_{i-1}, k_i) \text{ is left } \mathcal{C} \text{-restricting} \}. \]
Similarly, we consider words with ‘restricted suffixes’:
\[ \mathcal{R}_\delta^\ast(\alpha, \beta) = \{ w \in \mathcal{L} \mid \text{ there are } \alpha |w| \leq k_0 < k_1 < \cdots < k_\ell \leq \beta |w| \text{ with } \ell \geq \delta |w| \text{ such that every } [k_{i-1}, k_i) \text{ is right } \mathcal{C} \text{-restricting} \}. \]

When $\alpha = 0$ we will adopt the convention that $|\alpha w| = 1$ in each of the above definitions; that is, the interval under consideration starts at the index 1.

**Lemma 6.11.** For every $0 \leq \alpha < \beta \leq 1$ we have $P(\mathcal{L} \setminus \mathcal{W}(\alpha, \beta), \varphi) < P(\varphi)$.
Moreover, for every such $\alpha, \beta$ there is $\delta > 0$ such that
\[ P(\mathcal{L} \setminus \mathcal{R}_\delta^\ast(\alpha, \beta), \varphi) < P(\varphi), \quad P(\mathcal{L} \setminus \mathcal{R}_\delta^\ast(\alpha, \beta), \varphi) < P(\varphi). \]

**Proof.** Given $w \in \mathcal{L} \setminus \mathcal{W}(\alpha, \beta)$ there are $i < \alpha |w|$ and $j > \beta |w|$ such that $w_{[i, j]} \in \mathcal{C} := \mathcal{C}^\ast \cup \mathcal{C} \mathcal{G} \cup (\mathcal{L} \setminus \mathcal{C} \mathcal{G}^\ast)$. By [11] we can choose $\varepsilon > 0$ such that $P(\mathcal{C}, \varphi) + 2\varepsilon < P(\varphi)$, and so there is $K$ such that $\Lambda_n(\mathcal{C}, \varphi) \leq Ke^{P(\varphi) - \varepsilon}$ for every $n \in \mathbb{N}$. Thus
\[
\Lambda_n(\mathcal{L} \setminus \mathcal{W}(\alpha, \beta), \varphi) \leq \sum_{i=1}^{\lfloor \alpha n \rfloor} \sum_{j=\lfloor \beta n \rfloor} \Lambda_i(\mathcal{L}, \varphi) \Lambda_{j-i}(\mathcal{C}, \varphi) \Lambda_{n-j}(\mathcal{L}, \varphi)
\leq K(Q_2)^2 e^{nP(\varphi)} \sum_{k=[(\beta-\alpha)n]}^{n} \sum_{1 \leq i < an} e^{-k\varepsilon}
\leq K(Q_2)^2 e^{nP(\varphi)} \sum_{k \geq (\beta-\alpha)n} \sum_{j-i=k} ne^{-k\varepsilon}
\leq K(Q_2)^2 ne^{nP(\varphi)} e^{-(\beta-\alpha)\varepsilon n} (1 - e^{-\varepsilon})^{-1},
\]
Thus for every $\delta > 0$ let

$$\ell(w) = \begin{cases} \lceil \alpha n \rceil, \\ \lceil \frac{\beta n}{\ell} \rceil \\ \max \{ j \in (j_i, n] \mid w_{[j_i, j]} \in C \} \end{cases};$$

let $\ell$ be the maximum index such that $j_\ell < \beta n$.

Given $\ell, n \in \mathbb{N}$ and $J \subset (\alpha n, \beta n) \cap \mathbb{N}$ with $\# J = \ell$, let

$$\mathcal{X}_n(J) = \{ w \in \mathcal{L}_n \mid \ell(w) = \ell \text{ and } \{ j_i(w) \}_{i=1}^\ell = J \}.$$ 

Fixing $n \in \mathbb{N}$ and $\ell < n$, let

$$\mathbb{I}_\ell = \{ J \subset (\alpha n, \beta n) \cap \mathbb{N} \mid \# J = \ell \}.$$

Thus for every $\delta > 0$ we have

$$\mathcal{L} \setminus \mathcal{R}_\delta^p(\alpha, \beta) \subset \bigcup_{\ell=0}^{\lfloor \delta n \rfloor} \bigcup_{J \in \mathbb{I}_\ell} \mathcal{X}_n(J).$$

Write $\gamma = \delta/(\beta - \alpha)$. For every $\ell < \delta n$, Stirling’s approximation gives

$$\# \mathbb{I}_\ell \leq \left( \left\lfloor \frac{\beta n}{\ell} \right\rfloor \right) \leq Ce^{-\gamma \log \gamma n},$$

where $C$ is a constant independent of $\alpha, \beta, \delta, \ell, n$.

For each $J = \{ j_1 < j_2 < \cdots < j_\ell \} \in \mathbb{I}_\ell$ and $w \in \mathcal{X}_n(J)$ there is $m \in (\beta n, n]$ such that $w_{[j_\ell, m]} \in C$. Thus we have

$$\mathcal{X}_n(J) \subset \bigcup_{m = \lfloor \beta n \rfloor}^{n} \mathcal{L}_{j_1} C_{j_2-j_1} C_{j_3-j_2} \cdots C_{j_\ell-j_{\ell-1}} C_{m-j_\ell} \mathcal{L}_{n-m}.$$

As before we have $\Lambda_j(\mathcal{C}, \varphi) \leq K e^{(P(\varphi)-\epsilon)}$ for all $j$, and so

$$\Lambda_n(\mathcal{X}_n(J), \varphi) \leq \sum_{m = \lfloor \beta n \rfloor}^{n} \Lambda_{j_1}(\mathcal{L}, \varphi) \left( \prod_{i=1}^{\ell-1} \Lambda_{j_{i+1}-j_i}(\mathcal{C}, \varphi) \right) \Lambda_{m-j_\ell}(\mathcal{C}, \varphi) \Lambda_{n-m}(\mathcal{C}, \varphi) \leq \sum_{m = \lfloor \beta n \rfloor}^{n} (Q_2)^2 e^{(j_1+n-m)P(\varphi)} K^\ell e^{(m-j_1)(P(\varphi)-\epsilon)} \leq (Q_2)^2 K^\ell (1-\beta)n e^{nP(\varphi)} e^{-(\beta-\alpha)\gamma n}.$$ 

Together with (6.22) and (6.23) this gives

$$\Lambda_n(\mathcal{L} \setminus \mathcal{R}_\delta^p(\alpha, \beta), \varphi) \leq \delta n C e^{-(\gamma \log \gamma)n} (1-\beta)n (Q_2)^2 K^\delta \epsilon e^{nP(\varphi)} e^{-(\beta-\alpha)\gamma n}$$

and hence

$$P(\mathcal{L}, \mathcal{R}_\delta^p(\alpha, \beta), \varphi) \leq P(\varphi) + \delta \log K - \gamma \log \gamma - (\beta-\alpha)\epsilon.$$
Since $\gamma \to 0$ as $\delta \to 0$, for small enough choices of $\delta$ this is $< P(\varphi)$, which proves the lemma. \hfill \Box

### 6.5.3. Definition of $E^p$ and $E^s$.

Now that we have some control on how many ‘good’ subwords a typical word $w$ has, it is reasonable to pursue the following line of attack, which modifies \cite[6.4] taking into account the fact that $G = L$. We start by giving a rough outline, which will need some more work to make precise.

- Given $w \in G$, let $i$ be minimal and $j$ maximal such that \cite[6.8] and \cite[6.9] hold, and moreover $x_{[i+|c|,j]} \in G$. Thus $x_{[i+|c|,j]} \in F$.
- For this choice of $i, j$, the word $w_{[1,i+|c|]}$ does not contain $cs$ as a subword in a ‘good’ position; similarly $w_{[j,|w|]}$ does not contain $r$ as a subword in a ‘good’ position.
- Take $E^p$ to be the collection of words $v$ which either have fewer than $\delta|v|$ ‘good’ positions, or which have this many ‘good’ positions but avoid the word $cs$ at all of them. Similarly for $E^s$. Then $w_{[1,i+|c|]} \in E^p$ and $w_{[j,|w|]} \in E^s$, and $P(E^p \cup E^s, \varphi) < P(\varphi)$ using Lemma \cite[6.11] and a version of the argument in \cite[6.4.2].

The first problem with making the above ideas into a proof is that we must clarify what is meant by a ‘good’ position. The issue is that our notion of ‘goodness’ applies to a pair of positions, by asking that $w_{[i,j]} \in G$. One way forward is to say that given $w \in G$ and $a \in [1, \frac{1}{2}|w|)$, we consider $a$ to be a ‘good’ position if there is $b > \frac{1}{2}|w|$ such that $w_{[a,b]} \in G$. But then in order to determine how many ‘good’ positions are in $w_{[1,i]}$ for some $i < \frac{1}{2}|w|$, we must look at the part of $w$ that comes after index $i$. This is a problem because membership in $E^p$ is to be determined only by the word $w_{[1,i]}$ itself, and not by the symbols that follow it in $w$.

Similarly, one can try to produce ‘good’ positions in $w_{[1,i]}$ by asking that $w \in \mathcal{R}^p(\alpha, \beta)$ for a suitable $\alpha < \beta$ and then using Lemma \cite[6.10]. The problem once again is that a set may be right $\mathcal{C}$-restricting in $w_{[1,i]}$ without necessarily being right $\mathcal{C}$-restricting in $w$; this happens, for instance, when $w_{[a,b]} \in \mathcal{C}^p$ for some $1 \leq a < i < b$. This is where the notion of a wall will become particularly useful, thanks to the following lemma.

**Lemma 6.12.** Suppose $[j,k]$ is a wall in $w$ and $v = w_{[1,m]}$ for some $m \geq k$. Then any wall $[a,b]$ in $v$ with $a \leq j$ is also a wall in $w$.

Similarly, suppose $[j,k]$ is a wall in $w$ and $v = w_{[m,|w|]}$ for some $m < j$. Suppose $[|v| - a, |v| - b]$ is a wall in $v$ with $|v| - b \geq k$. Then $[|w| - a, |w| - b]$ is a wall in $w$.

**Proof.** Suppose $i, \ell \in \mathbb{N}$ are such that $1 \leq i \leq a$ and $b \leq \ell \leq |w|$. Then one of the following two cases holds.

- $\ell \leq |v|$. In this case $w_{[i,\ell]} = v_{[1,\ell]} \notin \mathcal{C}$ since $[a,b]$ is a wall in $v$.
- $\ell > v$. In this case $\ell \geq k$ and $i \leq a \leq j$, so $w_{[i,\ell]} \notin \mathcal{C}$ since $[j,k]$ is a wall in $w$.
The proof of the second claim is similar. □

To define $E^p, E^s$, we will use the collections $c(sL \cap G)$ and $Lr \cap G$. These are the words for which one end of the word has the correct part of the synchronising triple to be in $F$ (and the correct part of the word is in $G$), but the other end may not.

Fix $δ > 0$ such that Lemma 6.1 holds, and let

$$E^p = \bigcap_{i \in \{1, \frac{1}{3} |v|, \frac{2}{3} |v|\}, \ j \in \{\frac{1}{3} |v|, \frac{2}{3} |v|\}} \{v \in L \mid v_{i,j} \notin c(sL \cap G)\}$$

(6.24)

$$\cup \left( L \setminus \left( W\left(\frac{1}{4}, \frac{1}{2}\right) \cap R^s\left(\frac{1}{2}, \frac{1}{2}\right) \cap W\left(\frac{2}{3}, 1\right) \right) \right) .$$

That is, $E^p$ consists of words $v$ satisfying at least one of the following criteria:

1. $v$ does not contain the subword $cs$ from the synchronising triple at a ‘good’ time between $(\frac{1}{3} |v|, \frac{2}{3} |v|)$;
2. some element of $C$ appears in $v$ in a position that crosses one of the index ranges $(\frac{1}{4} |v|, \frac{1}{2} |v|), (\frac{1}{2} |v|, \frac{3}{2} |v|), (\frac{3}{4} |v|, |v|)$ — that is, one of the index ranges fails to be a wall in $v$;
3. $v$ does not have a large right-restricting set in $(\frac{1}{3} |v|, \frac{2}{3} |v|)$.

Thus, if a word in $E^p$ has walls and many right-restricting intervals, then there are $≥ δ |w|$ indices $i \in (\frac{1}{3} |w|, \frac{1}{2} |w|)$ such that $w_{i,j} \in G$ for some $j \in (\frac{2}{3} |w|, |w|)$ (see Figure 6.5), but none of these sees the word $cs$ in the right position.

![Figure 6.5](image-url)

**Figure 6.5.** If $w \in E^p$ has walls and many right-restricting intervals, then it does not contain $cs$ at a ‘good’ time.

A similar description can be given for

$$E^s = \bigcap_{i \in \{1, \frac{1}{3} |v|\}, \ j \in \{\frac{1}{3} |v|, \frac{2}{3} |v|\}} \{v \in L \mid v_{i,j} \notin Lr \cap G\}$$

(6.25)

$$\cup \left( L \setminus \left( W(0, \frac{1}{3}) \cap W\left(\frac{1}{3}, \frac{1}{2}\right) \cap R^s\left(\frac{1}{2}, \frac{3}{2}\right) \cap W\left(\frac{2}{3}, \frac{3}{2}\right) \right) \right) .$$

In §6.5.4 we show that $E^p, E^s$ have small pressure; first we describe the collection $E^p, F E^s$. The crucial result is the following lemma (and its counterpart for $E^s$).

**Lemma 6.13.** Given any $w \in L \setminus E^p$, there are $i \in [1, \frac{1}{2} |w| - |c|]$ and $j \in (\frac{2}{3} |w|, |w|)$ such that
(1) \( w_{|i,k|} \in \mathcal{E}_p \); and
(2) \( w_{|i,j|} \in c(s \mathcal{L} \cap \mathcal{G}) \).

**Proof.** Taking the complement of \((6.24)\), we get
\[
\mathcal{L} \setminus \mathcal{E}_p = \bigcup_{i \in (\frac{1}{3}|w|, \frac{1}{2}|w| - |c|)} \{ v \in \mathcal{L} \mid v_{|i,j|} \in c(s \mathcal{L} \cap \mathcal{G}) \}
\]
\( \cap \mathcal{W}(\frac{1}{4}, \frac{1}{3}) \cap \mathcal{W}(\frac{1}{3}, \frac{1}{2}) \cap \mathcal{W}(\frac{2}{3}, 1). \)

In particular, given \( w \in \mathcal{L} \setminus \mathcal{E}_p \) there are \( i_0 \in (\frac{1}{3}|w|, \frac{1}{2}|w| - |c|) \) and \( j_0 \in (\frac{2}{3}|w|, |w|) \) such that \( w_{|i_0,j_0|} \in c(s \mathcal{L} \cap \mathcal{G}) \).

If \( v := w_{|i_0,j_0|} \in \mathcal{E}_p \) then we are done. Otherwise we observe that by \((6.26)\), both of the following are true (see Figure 6.6):

- \([\frac{1}{3}i_0, \frac{1}{2}i_0]\) and \([\frac{1}{3}i_0, \frac{2}{3}i_0]\) are both walls in \( v \);
- there is \( i_1 \in (\frac{1}{3}i_0, \frac{1}{2}i_0 - |c|) \) and \( k_1 \in (\frac{2}{3}i_0, i_0) \) such that \( w_{|i_1,k_1|} \in c(s \mathcal{L} \cap \mathcal{G}) \).

Because \([\frac{1}{3}|w|, \frac{1}{2}|w|]\) is a wall in \( w \) and \( i_0 > \frac{1}{3}|w| \), it follows from Lemma 6.12 that \([\frac{1}{3}i_0, \frac{1}{3}i_0]\) and \([\frac{1}{3}i_0, \frac{2}{3}i_0]\) are both walls in \( w \).

\[
\begin{array}{cccccccc}
\frac{1}{3}|w| & \frac{1}{2}|w| - i_0 & \frac{1}{2}|w| & \frac{2}{3}|w| & j_0 & |w| \\
\frac{1}{3}i_0 & \frac{1}{2}i_0 & \frac{2}{3}i_0 & k_1 & i_0 & \in \mathcal{G} \\
i_1 & \in \mathcal{G} & \downarrow & \in \mathcal{G} & \in \mathcal{G} & \downarrow & \in \mathcal{G} \\
\end{array}
\]

**Figure 6.6. Proving Lemma 6.13**

Since \([\frac{1}{3}i_0, \frac{2}{3}i_0]\) and \([\frac{2}{3}|w|, |w|]\) are both walls in \( w \), by Lemma 6.10 there are \( a \in \left[\frac{2}{3}i_0, \frac{2}{3}i_0\right] \) and \( j_1 \in \left[\frac{2}{3}|w|, |w|\right] \) such that \( w_{|a,j_1|} \in \mathcal{G} \). Note that \( i_1 + |c| \leq a < k_1 < j_1 \), and so by (III) we have \( w_{|i_1+|c|,j_1|} \in \mathcal{G} \). In particular, \( w_{|i_1,j_1|} \in c(s \mathcal{L} \cap \mathcal{G}) \).

Proceeding in this manner we construct iteratively \( i_n \in [1, \frac{1}{2}|w|] \) and \( j_n \in (\frac{2}{3}|w|, |w|) \) such that

- \([\frac{1}{3}i_{n-1}, \frac{1}{3}i_{n-1}]\) is a wall in \( w \);
- \( w_{|i_n,j_n|} \in c(s \mathcal{L} \cap \mathcal{G}) \).
The argument above shows that if we have carried out this construction up to some \( n \), then either \( w_{[1,i_n]} \in \mathcal{E}^p \) or we can iterate the construction another time, obtain \( i_{n+1}, j_{n+1} \). Because \( i_n \) is a decreasing sequence of positive integers this process eventually terminates; when it does, we obtain the \( i, j \) required in the statement of the lemma. \( \square \)

A similar lemma for \( \mathcal{E}^s \) is proved in completely analogous manner.

**Lemma 6.14.** Given any \( w \in \mathcal{L} \setminus \mathcal{E}^s \), there are \( k \in [1, \frac{1}{3} |w|] \) and \( \ell \in [\frac{1}{2} |w|, |w|] \) such that

1. \( w_{[\ell,|w|]} \in \mathcal{E}^s \); and
2. \( w_{[k,\ell]} \in \mathcal{L} \cap \mathcal{G} \).

Now we consider the collection

\[
(6.27) \quad \mathcal{D} = \mathcal{W}(\frac{1}{4}, \frac{1}{3}) \cap \mathcal{W}(\frac{2}{3}, \frac{3}{4}).
\]

It follows from Lemma \[6.11\] that \( P(\mathcal{L} \setminus \mathcal{D}, \varphi) < P(\varphi) \). Using Lemmas \[6.13\] and \[6.14\] we show that words in \( \mathcal{D} \) satisfy a similar trichotomy to the one described following (6.10); we think of elements of \( \mathcal{D} \) as words having a ‘decomposition’, either as \( \mathcal{E}^p \mathcal{F} \mathcal{E}^s \), or as \( \mathcal{E}^p \mathcal{L} \cup \mathcal{L} \mathcal{E}^s \), where in the latter case the two subwords in the decomposition have roughly the same length. More precisely, we let \( \hat{\mathcal{E}} \) be as in (6.11) (using the new definition of \( \mathcal{E}^p, \mathcal{E}^s \) from (6.24)–(6.25)), and have the following result.

**Proposition 6.15.** \( \mathcal{D} \subset (\mathcal{E}^p \mathcal{F} \mathcal{E}^s) \cup \hat{\mathcal{E}} \).

**Proof.** Given \( w \in \mathcal{D} \), let \( m = \lfloor \frac{w}{2} \rfloor \) and observe that if \( w_{[1,m]} \in \mathcal{E}^p \) or \( w_{[m,|w|]} \in \mathcal{E}^s \), then \( w \in \hat{\mathcal{E}} \) and we are done. Otherwise by Lemma \[6.13\] there are \( i \in [1, \frac{1}{4} |w| - |c|] \) and \( j \in [\frac{1}{2} |w|, \frac{1}{2} |w|] \) such that

\[
w_{[1,i]} \in \mathcal{E}^p, \quad w_{[i,j]} \in \mathcal{E}^p \mathcal{L} \cap \mathcal{G}.
\]

Similarly, by Lemma \[6.14\] there are \( k \in [\frac{1}{2} |w|, \frac{2}{3} |w|] \) and \( \ell \in [\frac{2}{3} |w|, |w|] \) such that

\[
w_{[\ell,|w|]} \in \mathcal{E}^s, \quad w_{[k,\ell]} \in \mathcal{L} \cap \mathcal{G}.
\]

By the definition of \( \mathcal{D} \) and Lemma \[6.10\] there are \( a \in [\frac{1}{4} |w|, \frac{1}{3} |w|] \) and \( b \in [\frac{2}{3} |w|, \frac{3}{4} |w|] \) such that \( w_{[a,b]} \in \mathcal{G} \) (see Figure \[6.7\]). Observing that \( i + |c| < a < j < k < b < \ell \), we can apply \[III\] twice to get \( w_{[i+|c|,\ell]} \in \mathcal{G} \), so that in particular

\[
w_{[i,\ell]} \in \mathcal{E}^p \mathcal{L} \cap \mathcal{L} \cap \mathcal{G} = \mathcal{F}
\]

and we have \( w = w_{[1,i]} w_{[i,\ell]} w_{[\ell,|w|]} \in \mathcal{E}^p \mathcal{F} \mathcal{E}^s \). \( \square \)
6.5.4. $P(\mathcal{E}^p \cup \mathcal{E}^s, \varphi) < P(\varphi)$. With $\mathcal{E}^p, \mathcal{E}^s, \mathcal{E}, \mathcal{D}$ as in the previous section, it follows from Proposition 6.15 and (6.12) that
\[
P(\mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s, \varphi) \leq P(\tilde{\mathcal{E}} \cup (\mathcal{L} \setminus \mathcal{D}), \varphi)
\leq \max\left(P(\mathcal{L} \setminus \mathcal{D}, \varphi), \frac{1}{2}(P(\mathcal{E}^p \cup \mathcal{E}^s, \varphi) + P(\varphi))\right).
\]

Lemma 6.11 gives $P(\mathcal{L} \setminus \mathcal{D}, \varphi) < P(\varphi)$, and so to show that $P(\mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s, \varphi) < P(\varphi)$ it suffices to show that $P(\mathcal{E}^p \cup \mathcal{E}^s, \varphi) < P(\varphi)$. Thus we have reduced the problem of proving (11) to establishing this inequality. We prove that $P(\tilde{\mathcal{E}}^s, \varphi) < P(\varphi)$. The proof for $\mathcal{E}^p$ is analogous.

Let $\mathcal{V} = \mathcal{W}(0, \frac{1}{3}) \cap \mathcal{W}(\frac{1}{3}, \frac{2}{3}) \cap \mathcal{R}_\delta^s(\frac{1}{2}, \frac{2}{3}) \cap \mathcal{W}(\frac{2}{3}, \frac{3}{4})$ and let $\tilde{\mathcal{E}}^s = \mathcal{E}^s \cap \mathcal{V}$.

From (6.25) (see also Figure 6.5) we see that
\[
(6.28) \quad \tilde{\mathcal{E}}^s \subset \bigcap_{i \in [1, \frac{1}{|v|}], j \in (\frac{1}{2}|v|, \frac{2}{3}|v|)} \{v \in \mathcal{L} \mid v_{[i,j]} \notin \mathcal{L} \cap \mathcal{G}\}.
\]

Note that $\mathcal{E}^s \setminus \tilde{\mathcal{E}}^s \subset \mathcal{L} \setminus \mathcal{V}$, and so by Lemma 6.11 we have $P(\mathcal{E}^s \setminus \tilde{\mathcal{E}}^s, \varphi) < P(\varphi)$. Thus it suffices to show that $P(\tilde{\mathcal{E}}^s, \varphi) < P(\varphi)$.

The idea is to mimic the proof given in (6.4.2) for the uniform case, using Lemma 6.10 and the inclusion

\[
(6.29) \quad \tilde{\mathcal{E}}^s \subset \mathcal{V} \subset \mathcal{W}(0, \frac{1}{3}) \cap \mathcal{R}_\delta^s(\frac{1}{2}, \frac{2}{3})
\]

to guarantee that specification can be invoked for enough subwords.

Given $w \in \mathcal{V}$, consider the set of indices
\[
(6.30) \quad J(w) = \{j \in (\frac{1}{2}|w|, \frac{2}{3}|w|) \mid w_{[i,j]} \in \mathcal{G} \text{ for some } i \in [1, \frac{1}{3}|w|]\}.
\]

Let $\ell = \ell(w) = |J(w)|$ and note that by (6.29) and Lemma 6.10 we have
\[
(6.31) \quad \ell(w) \geq \delta |w| \text{ for every } w \in \tilde{\mathcal{E}}^s.
\]

For a given $w$ we enumerate the elements of $J(w)$ as
\[j_1(w) < j_2(w) < \cdots < j_\ell(w)\]
for each $k \in [1, \ell]$ let
\[
i_k(w) := \max\{i \in [1, \frac{1}{3}|w|] \mid w_{[i,j_k]} \in \mathcal{G}\}.
\]

We can use (11) to show that $i_k$ is non-increasing and that the values of $i_k, j_k$ depend only on the word $w_{[i_k,j_k]}$.
Lemma 6.16. For every $w \in V$ and $1 \leq k < \ell(w)$, we have $i_{k+1}(w) \leq i_k(w)$. Moreover, suppose that some $n \in \mathbb{N}$ and $i \in [1, \frac{1}{3} n]$, $j \in \left(\frac{1}{2} n, \frac{2}{3} n\right)$ there are $v, w \in L_n$ and $k \geq 1$ such that $i_k(v) = i$, $j_k(v) = j$, and $w'_{[i,j]} = v'_{[i,j]}$. Then $i_k(w) = i$ and $j_k(w) = j$.

Proof. Suppose $w \in V$ is such that $i_k(v) > i_k(j)$. By the definition of $i_k, j_k$ we have $i_k < i_{k+1} < j_k < j_{k+1}$ and $w'_{[i_k, j_k]}, w'_{[i_{k+1}, j_{k+1}]} \in G$. Thus \[\text{Lemma 6.16}\] gives $w'_{[i_{k+1}, j_k]} \in G$, contradicting maximality of $i_k$ in the definition (6.32).

For the second part of the lemma, let $v, w$ be as described and first note that for every $m \in [1, k]$ we have $i_m(v), j_m(v) \in [i, j]$ (since $i_m$ is non-increasing and $j_m$ is increasing); thus $w'_{[i_m(v), j_m(v)]} = v'_{[i_m(v), j_m(v)]} \in G$. It follows that $j_k(w) \leq j = j_k(v)$.

Now suppose $i' \in [1, i)$ and $j' \in \left[\frac{1}{2} n, j\right]$ are such that $w'_{[i', j']} \in G$. Then $i' < i < j' \leq j$, and since $w'_{[i', j]} = v'_{[i, j]} \in G$, \[\text{Lemma 6.17}\] implies that $w'_{[i', j]} \in G$. This in turn implies that $v'_{[j', j]} \in G$, so $j' \in J(v)$. We conclude that $J(w)$ agrees with $J(v)$ on $[\frac{1}{2} n, j_k(v)]$, and hence $j_k(w) = j_k(v) = j$.

To see that $i_k(w) = i_k(v) = i$, we observe that $w'_{[i, j_k(v)]} \in G$ so $i_k(v) \geq i$; thus $v'_{[i_k(w), j]} = w'_{[i_k(w), j]} \in G$, and maximality of $i_k(v)$ implies that $i_k(w) = i_k(v) = i$. \\[\square\]

Given $1 \leq k \leq \delta n$, define

$$\Theta_{k,n} : V_n \rightarrow G \times \mathbb{N}$$

$$w \mapsto (w'_{[i_k(w), j_k(w)]}, i_k(w)),$$

and consider the collection

$$\mathcal{Y}_k(n) = \Theta_{k,n}(V_n) \subset G \times \mathbb{N}.$$ 

Recall from (4.11) that we write $\mathcal{H}_n(u, i) = \{ w \in L_n \mid w'_{[i, i+|u|]} = u \}$.

Lemma 6.17. Given $w \in V_n$ and $u \in \mathcal{Y}_k(n)$, we have $w \in \mathcal{H}_n(u, i)$ if and only if $(u, i) = \Theta_{k,n}(w)$.

Proof. The backward implication is immediate from the definition of $\Theta_{k,n}$. The forward implication is a consequence of Lemma 6.16. \\[\square\]

By Lemma 6.17 the following union is disjoint for every $1 \leq k \leq \delta n$:

$$\mathcal{E}_n^s \subset V_n = \bigcup_{(u, i) \in \mathcal{Y}_k(n)} \mathcal{H}_n(u, i).$$

By (6.25), every $w \in \mathcal{E}_n^s$ has the property that $w'_{[i_k, j_k]} \notin L_r$ for every $1 \leq k \leq \ell(w)$. Thus we consider the collections

$$\mathcal{A}_n = \{ w \in V_n \mid w'_{[a, b]} \notin L_r \cap G \text{ for every } a \in [i_k(w), \frac{1}{3} n], b \in [\frac{1}{2} n, j_k(w)] \}$$

and $\mathcal{Z}_k(n) = \Theta_{k,n}(\mathcal{A}_n^k) \subset \mathcal{Y}_k(n)$, so that

$$\mathcal{A}_n^k = \bigcup_{(u, i) \in \mathcal{Z}_k(n)} \mathcal{H}_n(u, i).$$
That is, $A_k^n$ is the collection of words in $V_n$ that avoid the word $r$ from the synchronising triple for the first $k$ times that they contain a good subword crossing $(\frac{1}{3}n, \frac{1}{2}n)$, and $Z^k(n)$ is the collection of elements of $Y^k(n)$ corresponding to these words.

Observe that $A_{k+1}^n \subseteq A_k^n$ and $E_n^k \subseteq A_n^{|\delta_n|}$, so we can estimate $\Lambda_n(A_{k+\hat{\tau}}, \varphi)$ by estimating $\Lambda_n(A_k^{k+\hat{\tau}}, \varphi)$$/\Lambda_n(A_k^n, \varphi)$ for $\hat{\tau} = \tau + |r|$.

**Lemma 6.18.** Given $(u, i) \in Z^k(n)$, let $q \in L_{\leq r}$ be such that $uqr \in G$ (by [I]). Then $H_n(uqr, i) \cap A_k^{k+\hat{\tau}} = \emptyset$.

**Proof.** Given $w \in H_n(uqr, i)$, suppose that $\ell(w) \geq k$. It suffices to show that there are $a \in [i_{k+r}(w), \frac{1}{2}n]$ and $b \in [\frac{1}{2}n, j_k(w)]$ such that $w[a,b) \in Lr \cap G$. Because $w_{[i,i+|uqr|)} = uqr \in Lr \cap G$, it suffices to show that $i_{k+r}(w) \leq i$ and that $j_{k+r}(w) \geq j + |qr|$. The first inequality is immediate because $i_k$ is non-increasing. The second follows because $j_k$ is increasing and $\hat{\tau} = \tau + |r| \geq |qr|$.

It follows from Lemma 6.18 that for every $(u, i) \in Z^k(n)$ there is $v = v(u) \in uL \cap Lr \cap G$ with $|v| \leq |u| + \hat{\tau}$ such that

$$H_n(v, i) \subset H_n(u, i) \setminus A_n^{j+\hat{\tau}}.$$  

Using disjointness of the union in (6.33) together with the Gibbs bound in Proposition 4.2 we get

$$\Lambda_n(A_k^n \setminus A_k^{k+\hat{\tau}}, \varphi) = \sum_{(u, i) \in Z^k(n)} \Lambda_n(H_n(u, i) \setminus A_k^{k+\hat{\tau}}, \varphi)$$

$$\geq \sum_{(u, i) \in Z^k(n)} \Lambda_n(H_n(v(u), i), \varphi)$$

$$\geq \sum_{(u, i) \in Z^k(n)} Q_4^{-1} e^{(n-k-\hat{\tau})P(\varphi) + \hat{\rho}(v)}$$

$$\geq \sum_{(u, i) \in Z^k(n)} Q_4^{-1} e^{(n-k)P(\varphi)} e^{-\hat{\tau}P(\varphi)} e^{-\hat{\rho}(u) - \hat{\tau}||\varphi||}$$

$$\geq \sum_{(u, i) \in Z^k(n)} Q_4^{-2} \Lambda_n(H_n(u, i)) e^{-\hat{\tau}(P(\varphi) + ||\varphi||)}$$

$$= Q_4^{-2} e^{-\hat{\tau}(P(\varphi) + ||\varphi||)} \Lambda_n(A_k^n, \varphi).$$

Taking $\gamma = Q_4^{-2} e^{-\hat{\tau}(P(\varphi) + ||\varphi||)}$, we see from (6.35) that

$$\Lambda_n(A_k^n \setminus A_k^{k+\hat{\tau}}, \varphi) \geq \gamma \Lambda_n(A_k^n, \varphi).$$

In particular, we get

$$\Lambda_n(A_k^{k+\hat{\tau}}, \varphi) \leq (1 - \gamma) \Lambda_n(A_k^n, \varphi).$$
Since $A_n^0 = \mathcal{L}_n$ and $\hat{\mathcal{E}}_n^a \subset A_n^{\lfloor \delta_n \rfloor}$, we can iterate (6.37) to get
\[
\Lambda_n(\hat{\mathcal{E}}^s, \varphi) = \Lambda_n(A_n^{\lfloor \delta_n \rfloor}, \varphi) \leq (1 - \gamma)^{\lfloor \delta_n \rfloor} \Lambda_n(\mathcal{L}, \varphi),
\]
\[
\frac{1}{n} \log \Lambda_n(\hat{\mathcal{E}}^s, \varphi) \leq \frac{1}{n} \left\lfloor \frac{\delta_n}{\hat{\tau}} \right\rfloor \log(1 - \gamma) + \frac{1}{n} \log \Lambda_n(\mathcal{L}, \varphi).
\]
Sending $n \to \infty$ gives
\[
P(\hat{\mathcal{E}}^s, \varphi) \leq P(\varphi) + \delta(\hat{\tau})^{-1} \log(1 - \gamma) < P(\varphi).
\]
A similar argument works to show that $P(\mathcal{E}^p, \varphi) < P(\varphi)$. Apart from the obvious modification that in this case one must work from right to left rather than left to right, the only difference is that when defining $v(u)$ by attaching part of the synchronising tuple to $u \in \mathcal{G}$, one must take $v(u) = \text{rcsqu}$ rather than the natural choice $\text{csqu}$, because we must have $v \in \mathcal{G}$. Taking $\hat{\tau} = \tau + |\text{rcs}|$, the proof completes in the same way.

7. Discussion

7.1. Conditions [III], [III'] and non-symbolic systems. In [CT14, CT15], the uniqueness results from [CT12, CT13] are generalised to the setting where $X$ is a compact metric space, $f: X \to X$ is a continuous map, and $\varphi: X \to \mathbb{R}$ is a continuous potential function. An application of these results to non-uniformly hyperbolic diffeomorphisms with dominated splittings is given in [CFT15].

In this setting the language $\mathcal{L}$ is replaced with the space of finite orbit segments $X \times \mathbb{N}$, where the pair $(x, n)$ is associated to the orbit segment $x, f(x), \ldots, f^{n-1}x$. Then one asks the collection $\mathcal{G} \subset X \times \mathbb{N}$ of ‘good’ orbit segments to satisfy a specification property (among other things). The analogue of condition [III] in this setting is as follows.

[III'] If $x \in X$ and $i \leq j < k \leq \ell$ are such that $(f^i x, k - i), (f^j x, \ell - j) \in \mathcal{G}$, then $(f^i x, \ell - i), (f^j x, k - j) \in \mathcal{G}$.

(Note that when $(X, f)$ is a shift space $(X, \sigma)$, this is equivalent to [III]).

When working with a diffeomorphism $f: M \to M$, the idea behind obtaining a collection $\mathcal{G} \subset X \times \mathbb{N}$ with specification is to take the dominated splitting $TM = E^s \oplus E^u$, fix $\chi > 0$, and to let $\mathcal{G}$ be the set of all $(x, n)$ such that
\[
\|Df^k|_{E^s(x)}\| \leq e^{-\chi k} \text{ for all } 1 \leq k \leq n,
\]
and in addition
\[
\|Df^{-k}|_{E^u(f^n x)}\| \leq e^{-\chi k} \text{ for all } 1 \leq k \leq n.
\]

With this definition of $\mathcal{G}$, suppose $x \in M$ and $i \leq j < k \leq \ell$ are such that $(f^i x, k - i), (f^j x, \ell - j) \in \mathcal{G}$. Then (7.1) gives
\[
\|Df^a|_{E^s(f^j x)}\| \leq e^{-\chi k} \text{ for all } 1 \leq a \leq k - i,
\]
\[
\|Df^a|_{E^s(f^j x)}\| \leq e^{-\chi k} \text{ for all } 1 \leq a \leq \ell - j.
\]
It immediately follows that
\[ \|Df^a\|_{E^s(f^{i-j})} \leq e^{-\chi a} \text{ for all } 1 \leq a \leq k - j; \]
a similar observation for \( E^u \) shows that \((f^ix, k-j) \in \mathcal{G}\). For \((f^ix, \ell-i)\), we observe that for every \(1 \leq a \leq \ell-i\), we either have \(1 \leq a \leq k-i\), or we have \(a > k-i\). In the first case we get the bound on \(Df^a\) immediately from (7.3). In the second case we use both (7.3) and (7.4) to get
\[
\|Df^a\|_{E^s(f^{i-j})} \leq \|Df^i\|_{E^s(f^{i-j})} \cdot \|Df^{a-(j-i)}\|_{E^s(f^{i-j})} \leq e^{-\chi(j-i)}e^{-\chi(a-(j-i))} = e^{-\chi a}.
\]
The bound for \(E^u\) comes similarly, and we conclude that \((f^ix, \ell-i) \in \mathcal{G}\). Thus (III) is automatically satisfied when \(\mathcal{G}\) is defined via (7.1)–(7.2).

One can give a formulation of the mechanism at work here that continues to make sense in the symbolic setting. If we write \(\mathcal{G}^-\) for the set of all \((x,n)\) satisfying (7.1) and \(\mathcal{G}^+\) for the set of all \((x,n)\) satisfying (7.2), then one has \(\mathcal{G} = \mathcal{G}^+ \cap \mathcal{G}^-\), and moreover, for every \(i < j < k\), we have
\[
(f^ix, j-i), (f^ix, k-j) \in \mathcal{G}^- \Rightarrow (f^ix, k-i) \in \mathcal{G}^-;
\]
\[
(f^ix, j-i), (f^ix, k-j) \in \mathcal{G}^+ \Rightarrow (f^ix, k-i) \in \mathcal{G}^+.
\]
Similarly, for a shift space \((X, \sigma)\) one could define \(\mathcal{G}\) satisfying (III) by defining \(\mathcal{G}^-, \mathcal{G}^+ \subseteq \mathcal{L}\) such that for every \(x \in X\) and \(i < j < k\), we have
\[
\begin{align*}
x_{[i,j)}, x_{[j,k)} & \in \mathcal{G}^- \Rightarrow x_{[i,k)} \in \mathcal{G}^-, \\
x_{[i,j)}, x_{[j,k)} & \in \mathcal{G}^+ \Rightarrow x_{[i,k)} \in \mathcal{G}^+,
\end{align*}
\]
and then taking \(\mathcal{G} = \mathcal{G}^- \cap \mathcal{G}^+\). If \(\mathcal{G}^-\) is chosen so that it has ‘specification to the left’ (for every \(u \in \mathcal{G}\) and \(v \in \mathcal{G}^-\) there is \(w \in \mathcal{L}_{\leq}\) such that \(uwv \in \mathcal{G}^-\)) and \(\mathcal{G}^+\) is chosen to have ‘specification to the right’, then it is reasonable to expect \(\mathcal{G}\) to have specification.

**Example 7.1.** For the \(\beta\)-shift one can let \(z\) be the \(\beta\)-expansion of 1, take \(\mathcal{C}^s = \{z_{[0,n]} \mid n = 0, 1, 2, \ldots\}\), and then put
\[
\mathcal{G}^+ = \{w \in A^* \mid w_{[k,|w|]} \notin \mathcal{C}^s \text{ for every } 1 \leq k \leq |w|\}.
\]
(See [CT12] for further details.) Then (7.5) holds automatically for \(\mathcal{G}^+\) and \(\mathcal{G}^- = \mathcal{L}\), so \(\mathcal{G} = \mathcal{G}^+\) satisfies (III) by the above discussion. Conditions (II) and (III) for this example are discussed in [CT12] CT13.

**Remark 7.2.** For the smooth systems considered in [CFT15], it is likely possible to build a Young tower with exponential tails, and thus prove exponential decay of correlations, central limit theorem, etc., for the unique equilibrium state. This has not yet been carried out, and work on other similar examples suggests that it presents non-trivial technical challenges, even though the basic ideas are clear [AP10] AL13 ADLP14. One motivation for the present work is the hope that it may be possible to give a set of conditions that imply the existence of such a tower without the need to establish by hand the necessary liftability results and tail decay rate.
7.2. Comparison of $P(\mathcal{E}, \varphi)$ and $P(\mathcal{C}, \varphi)$. It was claimed in Remark 3.3 that in Theorem 3.2 we may have $P(\mathcal{E}, \varphi) > P(\mathcal{C}, \varphi)$. Here we give an example supporting this claim.

Fix $d \geq 4$ and let $X$ be the SFT on $A = \{1, \ldots, d\}$ such that the allowed transitions are $a \to a + 1$ (mod $d$) and $a \to a + 2$ (mod $d$). Suppose $\mathcal{F} \subseteq \mathcal{L}(X)$ satisfies $[\mathcal{I}_0]$. Let $B = \{w_1 \mid w \in \mathcal{F}\}$ and $C = \{w \mid w \in \mathcal{F}\}$. Then by $[\mathcal{I}_0]$ we have $c \to b$ for every $c \in C$ and $b \in B$.

Since each $a \in A$ has exactly two followers (two choices of $b$ such that $a \to b$) and no two choices of $a$ have the same set of two followers, one of $B, C$ must be a singleton, call it $\{a\}$. Then every word in $\mathcal{F}$ either starts or ends with $a$, and we see that for any choice of $\mathcal{E}^p, \mathcal{E}^s$, we have $\mathcal{E} \supseteq \{w \in \mathcal{L} \mid w_k \neq a \text{ for all } 1 \leq k \leq |w|\} =: \mathcal{D}$.

Because every state has two followers we see that $h(X) = \log 2$ (there are always two choices for the next symbol, so $\#\mathcal{L}_n = d2^{n-1}$). On the other hand, we can estimate $h(\mathcal{D})$ from below as follows: given $n \in \mathbb{N}$ and $u \in \{1, 2\}^n$, define $\pi(u) \in \mathcal{D}_{n+1}$ by $\pi(u)_1 = 1$ and

$$
\pi(u)_{i+1} = \begin{cases} 
  a - 1 & \text{if } \pi(u)_i = a - 2, \\
  a + 1 & \text{if } \pi(u)_i = a - 1, \\
  \pi(u)_i + u_i & \text{otherwise},
\end{cases}
$$

where we work mod $d$. Given $u, v \in \{1, 2\}^n$, we have $\pi(u) = \pi(v)$ if and only if $u_i = v_i$ for all $i$ such that $\pi(u)_i \notin \{a - 2, a - 1\}$, and since every index interval of length $d/2$ contains at most 2 values of $i$ with $\pi(u)_i \in \{a - 2, a - 1\}$, we see that every $w \in \mathcal{D}_{n+1}$ has $\#\pi^{-1}(w) \leq 2^2 \pi n^2$, so $\#\mathcal{D}_{n+1} \geq 2^n 2 - 4n/d$, and we get $h(\mathcal{E}) \geq h(\mathcal{D}) \geq (1 - \frac{4}{d}) \log 2$. This shows that no matter what choice of $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s$ we make, we have $h(\mathcal{E}) > 0$ and that in fact the entropy gap between $h(\mathcal{E})$ and $h(X)$ can be forced to be arbitrarily small by taking $d$ large.

7.3. Proofs of results on factors.

Proof of Proposition 3.13. We recall the proof of [CT12, Proposition 2.2]: given two shifts $X, \hat{X}$ on finite alphabets $A, \hat{A}$ with a factor map $\pi: X \to \hat{X}$, there is some $m \in \mathbb{N}$ and $\theta: \mathcal{L}_{2m+1}(X) \to B$ such that $\pi(x)_n = \theta(x|_{n-m,n+m}]$ for every $x \in X$ and $n \in \mathbb{Z}$. Writing $\Theta: \mathcal{L}_{n+2m} \to \mathcal{L}_n$ for the map induced by $\theta$, we suppose that we are given $\mathcal{G}, \mathcal{C}, \mathcal{C}^s \subseteq \mathcal{L}$, and put

$$
\tilde{\mathcal{G}} = \Theta(\mathcal{G}), \quad \tilde{\mathcal{C}}^p = \Theta(\mathcal{C}^p \cdot i_{2m}^{-}(\mathcal{G})), \quad \tilde{\mathcal{C}}^s = \Theta(i_{2m}^{-}(\mathcal{G}) \cdot \mathcal{C}^s),
$$

where $i_{2m}^{-}(w) = w|_{|k|}$ and $i_{2m}^{-}(w) = w|_{[k,|w|]}$. The proof given in [CT12] that $[\mathcal{I}]$ for $\mathcal{G}$ implies $[\mathcal{I}]$ for $\tilde{\mathcal{G}}$ also shows that $[\mathcal{I}]$ carries through, which proves (1).

For (2), we suppose $\mathcal{G}$ satisfies $[\mathcal{I}]$ and claim that $\tilde{\mathcal{G}}$ does as well. Indeed, if $x \in X$ and $i \leq j < k \leq \ell$ are such that $\theta(x|_{[i,k]}, \theta(x|_{[j,\ell]} \in \tilde{\mathcal{G}}$, then we have $x|_{[i-m,k+m]}, x|_{[j-m,\ell+m]} \in \tilde{\mathcal{G}}$. Moreover, $i - m \leq j - m < k + m \leq \ell + m$,

---

20When $d = 4$ we have $1 - \frac{4}{d} = 0$, but a more careful inspection shows that $h(\mathcal{D}) > 0$. 
and since $\mathcal{G}$ satisfies \textbf{III} we see that $x_{[i-m,\ell+m]}, x_{[j-m,k+m]} \in \mathcal{G}$. By the definition of $\tilde{\mathcal{G}}$ this gives $x_{[i,\ell]}, x_{[j,k]} \in \tilde{\mathcal{G}}$, which verifies \textbf{III}.

For (3), it is shown in [CT12] that $\Theta(C^pG^s) \subseteq \tilde{\mathcal{G}}$, and so $\#\tilde{\mathcal{L}}_n \leq (\#\mathcal{L}_n)(\#\mathcal{L}_m)^2$, which suffices. For statement (4), given any $\tilde{w} \in \tilde{\mathcal{L}}$ we have $\tilde{w} = \Theta(w)$ for some $w \in \mathcal{L}$, and by \textbf{E} there are $u, v \in \mathcal{L}$ such that $uvw \in \mathcal{G}$, hence $\Theta(uvw) \in \tilde{\mathcal{G}}$, and $\Theta(uvw)$ contains $\tilde{w}$ as a subword, which proves $\textbf{E}^*$ for $\tilde{\mathcal{G}}, \tilde{\mathcal{L}}$.

\textbf{Proof of Proposition 3.14.} This is essentially a corollary of Lemma 6.7. If $\mathcal{G}$ is periodic (in the sense of Definition 6.2) then by \textbf{E} $\mathcal{L}$ is periodic as well, so $X$ is a single periodic orbit. Thus if $X$ is not a single periodic orbit, Lemma 6.7 applies to give $h(\mathcal{G}) > 0$, hence $h(X) > 0$. □

7.4. Coded systems. In light of the fact that every system with the uniform specification property is synchronised, and hence coded, it is natural to ask whether systems with the non-uniform specification property given by \textbf{I} \textbf{III} lie in these classes.

One quickly sees that such systems need not be synchronised; indeed, every $\beta$-shift satisfies \textbf{I} \textbf{III}, but not all $\beta$-shifts are synchronised [Sch97].

On the other hand, if $\mathcal{L}(X)$ contains a collection of words $\mathcal{F}$ satisfying \textbf{I} \textbf{III}, then one can consider the coded shift $X'$ with generating set $\mathcal{F}$, and observe that $X' \subseteq X$. If $\mathcal{F}$ is ‘large enough’ we can expect $X'$ to be large. The following example shows that we may have $X' \neq X$ even if $X$ satisfies \textbf{I} \textbf{III}.

\textbf{Example 7.3.} Let $X \subseteq \{0, 1, 2\}^\mathbb{Z}$ be the SFT defined by forbidding the words 20 and 21. Then $X$ satisfies \textbf{I} \textbf{III} for $\varphi = 0$ by taking $\mathcal{F} = \{0, 1\}^*$, $C^p = \emptyset$, and $C^s = \{2\}^*$, but the corresponding coded shift is $X' = \{0, 1\}^\mathbb{Z} \neq X$.

On the other hand, if $X$ satisfies \textbf{I} \textbf{III}, then Lemma 5.19 shows that every equilibrium state $\mu$ for $(X, \sigma, \varphi)$ has $\mu(X') = 1$. Thus from the thermodynamic point of view, every question about a system satisfying \textbf{I} \textbf{III} is a question about a coded shift.

Now suppose $X$ is a coded system; we make some observations comparing the approach of Theorem 3.1 to previous work on coded systems, and expanding on the questions raised in Remark 3.4. Given $\mathcal{F}$ satisfying \textbf{I} \textbf{III} we produced in (5.1) the set $I$ of ‘irreducible’ words in $\mathcal{F}$; the set $I$ is a generating set for the coded shift $X'$ described above. Consider

\begin{equation}
\mathcal{E}^p = \{w[i,\ell] \mid w \in I, 1 \leq i \leq |w|\},
\end{equation}

\begin{equation}
\mathcal{E}^s = \{w[1,j] \mid w \in I, 1 \leq j \leq |w|\},
\end{equation}

and observe that $\mathcal{E} = \mathcal{E}(\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s) \subseteq \mathcal{D}(I)$, where

\begin{equation}
\mathcal{D}(I) := \{w[i,j] \mid w \in I, 1 \leq i \leq j \leq |w|\}
\end{equation}

is the collection of words which occur as a subword of a single generator. This is the case because every word $v \in \mathcal{L}(X)$ is a subword of $w^1 \cdots w^n$ for some
generators \(w^1, \ldots, w^n \in I\), and if \(n > 1\) then we have
\[
(w) v = v^1_{[i, |w^1|]} w^2 \cdots w^{n-1} v^n_{[1, |j|]} \in \mathcal{E}^p \mathcal{F} \mathcal{E}^s
\]
for some \(i, j\), and hence \(\mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s \subset \mathcal{D}(I)\).

It is shown in [CT12 §4] that with this choice of \(\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s\), condition \([\mathbf{E}]\) holds, and in particular, every \(\mathcal{G}^M\) satisfies \([\mathbf{I}]\). Thus by Theorem 3.11 we have proved the following result.

**Theorem 7.4.** Let \(X\) be a coded shift and \(\varphi: X \to \mathbb{R}\) Hölder continuous. If there is a generating set \(I\) for \(X\) such that \(P(\mathcal{D}(I), \varphi) < P(\varphi)\), then

1. \((X, \varphi)\) has a unique equilibrium state \(\mu\);
2. \(\mu\) has the Gibbs property (2.5) with respect to \(\mathcal{F} = I^*\);
3. \(\mu\) is the limiting distribution of \(\varphi\)-weighted periodic orbits.

If in addition we know that \(\mathcal{F} = I^*\) satisfies \([\mathbf{III}^*]\), then we can apply Theorem 3.1 and deduce that \(\mu\) satisfies the stronger statistical properties (4)–(6) as well; these do not follow from Theorem 3.11.

Recall that the key construction in the proof of Theorem 3.1 was the description of \(X'\) in (5.3) and (5.4) as the closure of the uniformly continuous image of a countable-state irreducible topological Markov chain \(\Sigma\) on which the potential is strongly positive recurrent. Condition \([\mathbf{III}^*]\) was used (via Lemma 5.8) to guarantee that the map \(\pi: \Sigma \to X\) is 1-1, and also in the proof that every equilibrium state gives full weight to \(\pi(\Sigma)\) (not just to its closure). Without this condition, it is possible that our construction yields a map \(\pi\) that decreases entropy.

**Example 7.5.** Let \(X \subset \{0, 1\}^\mathbb{Z}\) be the SFT defined by forbidding the word 111. Let \(\mathcal{F} \subset \mathcal{L}(X)\) be the set of all words that neither start nor end with the word 11. Then \(\mathcal{F}\) satisfies condition \([\mathbf{I}_0]\) and we get \(I = I(\mathcal{F}) = \{0, 01, 10\}\) as the irreducible elements of \(\mathcal{F}\). But then \(010 = (01)(0) = (0)(10) \in \mathcal{F}\) has two different ‘factorisations’, which can also be used to show that \([\mathbf{III}^*]\) fails and that \(\pi^{-1}((010)\infty)\) is uncountable. In fact one can show that \(h(\Sigma) = \log 2 > h(X)\).

The ‘unique coding’ property guaranteed by Condition \([\mathbf{III}^*]\) plays a similar role to other conditions sometimes imposed on generating sets for coded systems, such as being uniquely decodable [LM95 §8.1 and §13.5] or prefix-free [BH86 §2.1]. It is shown in [BH86 Proposition 2.1] that every coded shift admits a uniquely decodable generating set, which can be used to build a good cover \(\Sigma\); similarly, [FP92 Theorem 1.7] shows that it is always possible to build a ‘bi-resolving’ cover, which in particular gives a 1-1 map \(\pi\). However, in both cases one must abandon the original generating set \(I\) and pass to a new generating set \(I'\), for which the set of obstructions \(\mathcal{D}(I')\) may be quite large, and in particular there is no a priori reason why \([\mathbf{II}]\) should hold. Since \([\mathbf{II}]\) was required to prove positive recurrence, we do not get a proof of statistical properties in this way. Thus we have the following open question.
Question 7.6. Let $X$ be a coded shift with a generating set $I \subset \mathcal{L}(X)$ such that $P(D(I)), \varphi) < P(\varphi)$ for some Hölder continuous $\varphi$. Let $\mu$ be the unique equilibrium state for $(X, \varphi)$ guaranteed by Theorem 7.4. Do conclusions (4)–(6) of Theorem 3.1 still hold? That is, is some iterate of $(X, \sigma, \mu)$ Bernoulli with exponential decay of correlations; does $(X, \sigma, \mu)$ satisfy the central limit theorem; and is the pressure function real analytic at $\varphi$?

It seems appropriate to conclude by pointing out an error in [CT12, Theorem B] that is corrected by Theorem 7.4. Given a coded shift $X$ and a generating set $I \subset \mathcal{L}(X)$, let $d_n = \# D(I)_n$ be the number of words of length $n$ that are contained in a single element of $I$. Recall that a system is said to be \textit{intrinsically ergodic} if it has a unique measure of maximal entropy. Combining Theorem 7.4 with Propositions 3.13 and 3.14, we obtain the following result.

**Theorem 7.7.** Let $X$ be a coded shift on a finite alphabet and let $d_n$ be as above for some generating set $I \subset \mathcal{L}(X)$.

1. If $\lim \frac{1}{n} \log d_n < h(X)$, then $(X, \sigma)$ is intrinsically ergodic.
2. If $\lim \frac{1}{n} \log d_n = 0$, then every subshift factor of $(X, \sigma)$ is intrinsically ergodic.

Moreover, under these conditions, the unique measure of maximal entropy is the limiting distribution of (unweighted) periodic orbits.

In [CT12, Theorem B], it is claimed that the above result holds with $d_n$ replaced by $c_n = \# (\mathcal{E}^p(I) \cup \mathcal{E}^s(I))_n$, the number of words of length $n$ that appear either at the beginning or the end of some generator. This claim is incorrect, as the following example shows.

**Example 7.8.** Let $X \subset \{0, 1, 2, 3, 4\}^\mathbb{Z}$ be the coded shift generated by

$$I = \{0^n w 0^n \mid w \in \{1, 2\}^* \cup \{3, 4\}^*, |w| \leq n\}.$$ 

Then we have

$$\# \mathcal{E}_n^p = \sum_{k=1}^{n} \# \{0^k w 0^j \mid w \in \{1, 2\}^* \cup \{3, 4\}^*, |w| \leq k, k + |w| + j = n\}$$

$$\leq \sum_{k=\lceil n/3 \rceil}^{n} k 2^{n-k+1} \leq 2^{2n/3} \sum_{j=0}^{\infty} j 2^{-j} = C e^{(\frac{2}{3} \log 2)n},$$

and similarly for $\mathcal{E}^s$. A straightforward computation shows that $h(X) = \log 2$ and so there are two distinct ergodic equilibrium states $((\frac{1}{2}, \frac{1}{2})$-Bernoulli on the subshifts $\{1, 2\}^\mathbb{Z}$ and $\{3, 4\}^\mathbb{Z}$), but we have

$$\lim_{n \to \infty} \frac{1}{n} \log c_n \leq \frac{2}{3} \log 2 < h(X).$$

\[21\] Note that if $\mathcal{L}$ is the language of a coded shift and $I \subset \mathcal{L}$ is a generating set, then $I^*, \mathcal{L}$ satisfy [E*] by definition.
The problem with \cite[Theorem B]{CT12} is that words appearing in the interior of a single generator do not have a decomposition in terms of $E^{p}, F = I^{*}$, and $E^{s}$, so $E^{p}F^{*}E^{s} \neq L$ and \cite[Theorem C]{CT12} cannot be applied. The solution is to replace \cite[(II)']{II} in \cite[Theorem C]{CT12} with the weaker condition \cite[(II)]{II}; the result remains true as indicated in Theorem 3.11. To apply it here, we replace $c_{n}$ with $d_{n}$ and obtain Theorem 7.7 as the corrected version of \cite[Theorem B]{CT12}.

**References**

[ADLP14] José F. Alves, C. L. Dias, S. Luzzatto, and V. Pinheiro, *SRB measures for partially hyperbolic systems whose central direction is weakly expanding*, Preprint, arXiv:1403.2937, 2014.

[AHS14] Artur Avila, Pascal Hubert, and Alexandra Skripchenko, *Diffusion for chaotic plane sections of 3-periodic surfaces*, Preprint. arXiv:1412.7913, 2014.

[AL13] José F. Alves and Xin Li, *Gibbs-Markov-Young structures with (stretched) exponential tail for partially hyperbolic attractors*, Preprint, arXiv:1301.7287, 2013.

[AP10] José F. Alves and Vilton Pinheiro, *Gibbs-Markov structures and limit laws for partially hyperbolic attractors with mostly expanding central direction*, Adv. Math. 223 (2010), no. 5, 1706–1730. MR 2592507 (2011e:37068)

[BCLS13] Martin Bridgeman, Richard Canary, François Labourie, and Andres Sambarino, *The pressure metric for convex representations*, Preprint. arXiv:1301.7459, 2013.

[Ber88] Anne Bertrand, *Specification, synchronisation, average length*, Coding theory and applications (Cachan, 1986), Lecture Notes in Comput. Sci., vol. 311, Springer, Berlin, 1988, pp. 86–95. MR 960710 (89i:94022)

[BG11] A. I. Bufetov and B. M. Gurevich, *Existence and uniqueness of a measure with maximal entropy for the Teichmüller flow on the moduli space of abelian differentials*, Mat. Sb. 202 (2011), no. 7, 3–42. MR 2857792 (2012h:37068)

[BH86] F. Blanchard and G. Hansel, *Systèmes codés*, Theoret. Comput. Sci. 44 (1986), no. 1, 17–49. MR 858689 (88m:68029)

[BL13] Henk Bruin and Renaud Leplaideur, *Renormalization, thermodynamic formalism and quasi-crystals in subshifts*, Comm. Math. Phys. 321 (2013), no. 1, 209–247. MR 3089670

[Bow72] Rufus Bowen, *The equidistribution of closed geodesics*, Amer. J. Math. 94 (1972), 413–423. MR 0315742 (47 #4291)

[Bow74] _____, *Some systems with unique equilibrium states*, Math. Systems Theory 8 (1974), no. 3, 193–202. MR 0399413 (53 #3257)

[Bow75] _____, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics, Vol. 470, Springer-Verlag, Berlin-New York, 1975. MR 0442989 (56 #1364)

[BS03] Jérôme Buzzi and Omri Sarig, *Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps*, Ergodic Theory Dynam. Systems 23 (2003), no. 5, 1383–1400. MR 2018604 (2004k:37053)

[BSS02] Luis Barreira, Benoît Saussol, and Jörn Schmeling, *Higher-dimensional multifractal analysis*, J. Math. Pures Appl. (9) 81 (2002), no. 1, 67–91. MR 1994883 (2004g:37038)

\footnote{We stress that there is no problem with \cite[Theorem C]{CT12}, which is the main result of that paper, or with any of the other results therein; the only problem is in the application to \cite[Theorem B]{CT12}.}
Henk Bruin and Mike Todd, *Equilibrium states for interval maps: the potential \(-t \log |Df|\)*, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 4, 559–600. MR 2568876 (2011c:37079)

Vaughn Climenhaga, Todd Fisher, and Daniel J. Thompson, *Unique equilibrium states for the robustly transitive diffeomorphisms of Mañé and Bonatti–Viana*, Preprint, 2015.

Vaughn Climenhaga and Daniel J. Thompson, *Intrinsic ergodicity beyond specification: \(\beta\)-shifts, \(S\)-gap shifts, and their factors*, Israel J. Math. **192** (2012), no. 2, 785–817. MR 3009742

---

Doris Fiebig and Ulf-Rainer Fiebig, *Covers for coded systems*, Symbolic dynamics and its applications (New Haven, CT, 1991), Contemp. Math., vol. 135, Amer. Math. Soc., Providence, RI, 1992, pp. 139–179. MR 1185086 (93m:54068)

Franz Hofbauer, *\(\beta\)-shifts have unique maximal measure*, Monatsh. Math. **85** (1978), no. 3, 189–198. MR 0492180 (58 #11326)

Franz Hofbauer, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy*, Israel J. Math. **34** (1979), no. 3, 213–237 (1980). MR 570882 (82c:28039a)

Franz Hofbauer, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. II*, Israel J. Math. **38** (1981), no. 1-2, 107–115. MR 599481 (82c:28039b)

Irene Inoquio-Renteria and Juan Rivera-Letelier, *A characterization of hyperbolic potentials of rational maps*, Bull. Braz. Math. Soc. (N.S.) **43** (2012), no. 1, 99–127. MR 2909925

Godofredo Iommi and Mike Todd, *Natural equilibrium states for multimodal maps*, Comm. Math. Phys. **300** (2010), no. 1, 65–94. MR 2725183 (2011m:37048)

A. Katok, *Entropy and closed geodesics*, Ergodic Theory Dynam. Systems **2** (1982), no. 3-4, 339–365 (1983). MR 721728 (85b:53047)

Gerhard Keller, *Lifting measures to Markov extensions*, Monatsh. Math. **108** (1989), no. 2-3, 183–200. MR 1026617 (91b:28011)

Yuri Kifer, *Large deviations in dynamical systems and stochastic processes*, Trans. Amer. Math. Soc. **321** (1990), no. 2, 505–524. MR 1025756 (91e:60091)

Gerhard Knieper, *The uniqueness of the measure of maximal entropy for geodesic flows on rank 1 manifolds*, Ann. of Math. (2) **148** (1998), no. 1, 291–314. MR 1652924 (2000b:37016)

Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995. MR 1369092 (97a:58050)

Huaibin Li and Juan Rivera-Letelier, *Equilibrium states of weakly hyperbolic one-dimensional maps for Hölder potentials*, Comm. Math. Phys. **328** (2014), no. 1, 397–419. MR 3196990

Curtis T. McMullen, *Thermodynamics, dimension and the Weil-Petersson metric*, Invent. Math. **173** (2008), no. 2, 365–425. MR 2415311 (2010c:37101)
[Pet89] Karl Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989, Corrected reprint of the 1983 original. MR 1073173 (92c:28010)

[PP90] William Parry and Mark Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque (1990), no. 187-188, 268. MR 1085356 (92f:58141)

[PS08] Yakov Pesin and Samuel Senti, *Equilibrium measures for maps with inducing schemes*, J. Mod. Dyn. 2 (2008), no. 3, 397–430. MR 2417478 (2009k:37072)

[PSZ08] Ya. B. Pesin, S. Senti, and K. Zhang, *Lifting measures to inducing schemes*, Ergodic Theory Dynam. Systems 28 (2008), no. 2, 553–574. MR 2408392 (2009f:37004)

[PZ07] Yakov Pesin and Ke Zhang, *Thermodynamics of inducing schemes and liftability of measures*, Partially hyperbolic dynamics, laminations, and Teichmüller flow, Fields Inst. Commun., vol. 51, Amer. Math. Soc., Providence, RI, 2007, pp. 289–305. MR 2388701 (2009a:37062)

[Rue76] David Ruelle, *A measure associated with axiom-A attractors*, Amer. J. Math. 98 (1976), no. 3, 619–654. MR 0415683 (54 #3763)

[Rue92] ———, *Thermodynamic formalism for maps satisfying positive expansiveness and specification*, Nonlinearity 5 (1992), no. 6, 1223–1236. MR 1192516 (94a:58115)

[Sar99] Omri M. Sarig, *Thermodynamic formalism for countable Markov shifts*, Ergodic Theory Dynam. Systems 19 (1999), no. 6, 1565–1593. MR 1738951 (2000m:37009)

[Sar01] ———, *Phase transitions for countable Markov shifts*, Comm. Math. Phys. 217 (2001), no. 3, 555–577. MR 1822107 (2002b:37040)

[Sar11] ———, *Bernoulli equilibrium states for surface diffeomorphisms*, J. Mod. Dyn. 5 (2011), no. 3, 593–608. MR 2854097 (2012k:37071)

[Sch97] Jörg Schmeling, *Symbolic dynamics for β-shifts and self-normal numbers*, Ergodic Theory Dynam. Systems 17 (1997), no. 3, 675–694. MR 1452189 (98c:11080)

[Wal78] Peter Walters, *Equilibrium states for β-transformations and related transformations*, Math. Z. 159 (1978), no. 1, 65–88. MR 0466492 (57 #6370)

[Wal82] ———, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR 648108 (84e:28017)

[You98] Lai-Sang Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) 147 (1998), no. 3, 585–650. MR 1637655 (99h:58140)

[You99] ———, *Recurrence times and rates of mixing*, Israel J. Math. 110 (1999), 153–188. MR 1750438 (2001j:37002)

Department of Mathematics, University of Houston, Houston, Texas 77204

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